The Stata Journal

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press, and Stata is a registered trademark of StataCorp LP.

The Stata Journal (2006)
6, Number 3, pp. 348–363

Graphical representation of interactions

Friederike Maria-Sophie Barthel
friederikeb@ctu.mrc.ac.uk

Patrick Royston
patrick.royston@ctu.mrc.ac.uk

MRC Clinical Trials Unit
222 Euston Road
London NW1 2DA, UK

Abstract. We provide a program to illustrate interactions between treatment and covariates or between two covariates by using forest plots under either the Cox proportional hazards or the logistic regression model. The program is flexible in both the possibility of illustrating more than one interaction at a time and variable specifications of scale.

Keywords: gr0024, fintplot, interaction, forest plot, randomized controlled trial, survival analysis, logistic regression

1 Introduction

When examining a particular treatment in a trial setting, we are often interested in the consistency of an observed relationship across covariates. We might suspect that a treatment works better in older patients than younger ones or that because of the genetic makeup of men and women the treatment works differently on the two sexes. Examining the relationship can be helpful later when developing guidance on how to use that particular treatment. One ongoing MRC Clinical Trials Unit study seeks to identify in colorectal cancer patients an interaction between the prevalence of the mutated gene *p53* and the results of chemotherapy.

As outlined by Shuster and van Eys (1983), tests for such interactions can have two uses. First, by retrospective analysis of possible interaction effects, one can formulate interesting hypotheses for future trials. Second, in planning a prospective trial, one may incorporate a test of an interaction effect if one suspects that the therapies manipulate important factors differently. Hence the analysis of interactions in a trial or study can be exploratory or consist of a test for interactions as defined in the protocol.

Gail and Simon (1985) discussed quantitative and qualitative interactions. In a quantitative interaction, the magnitude of the treatment effect will vary with a patient's characteristics while the direction of the overall treatment effect stays the same. In a qualitative interaction, a change in the direction of the treatment effect is involved. Both are illustrated in figure 1.

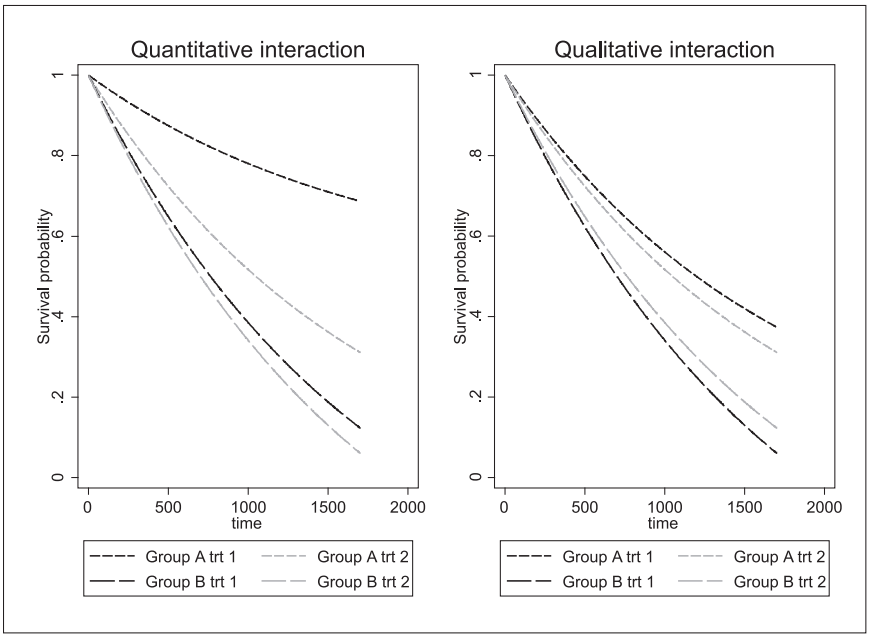

Figure 1: Quantitative and qualitative interactions illustrated using Kaplan–Meier survival curves

Thall and Lachin (1986) proposed a test based on proportional hazards regression models. Uesaka (1993) used logarithmic generalized odds ratios. Simon (2002) used Bayesian subset analysis. Xiang, Sather, and Azen (1994), whose test statistic is based on a weighted residual sum of squares, examined $2 \times k$ factorial experiments. To estimate the parameters of the test statistic, they used the Mantel–Haenszel, maximum likelihood estimation, and a method based on the ratio of observed to expected events.

Our estimate of the interaction effect is based on a ratio of hazard ratios (RHR) or a ratio of odds ratios (ROR) derived from a 2×2 table as described in section 3. The definition is similar to that of Peterson and George (1993). This RHR describes quantitative interactions. We provide a Stata program to illustrate interactions more easily during the analysis of a clinical trial or study. It provides both numerical and graphical output in the form of a forest plot while giving a choice of using either the Cox proportional hazards model or logistic regression.

The following section describes a conventional interaction analysis using a cancer trial dataset. This analysis is then rerun using the forest plot methods in section 5.1.

2 Analysis of a cancer trial with suspected treatment–covariate interactions

We ran the following analysis of a trial with potential treatment–covariate interaction effects to understand the magnitude of interaction effects and the best way to represent these. We conducted analyses by using the Cox proportional hazards model, as well as Kaplan–Meier survival curves. To run the formal interaction analysis, we created an interaction variable of treatment and a covariate.

The dataset `glioma2` is a multicenter German–Austrian randomized trial that tested two different chemotherapy regimens for brain tumors in adults. There were 447 patients randomized between February 1983 and June 1988. During the trial, 274 of 411 patients died. The overall hazard ratio of the trial was 0.89 in favor of chemotherapy, with a confidence interval ranging from 0.71 to 1.14 and a significance level of 0.38. Hence there was no evidence of a significant improvement in survival on the basis of treatment.

We can identify the time from first symptom, grade of malignancy, Karnofsky index, and aphasia as possible interaction candidates. An investigation of the influence of these variables was initially carried out by Ulm et al. (1989). Each of these variables was split into two levels, and the Karnofsky index itself has two different level definitions. The grade of malignancy and the second definition of the Karnofsky index show large discrepancies in the numbers of patients present in each group. Therefore, power for the comparison is relatively low.

Kaplan–Meier survival curves indicate that there may be an interaction, especially for grade of malignancy and the second specification of the Karnofsky index; see figures 2 and 3, respectively.

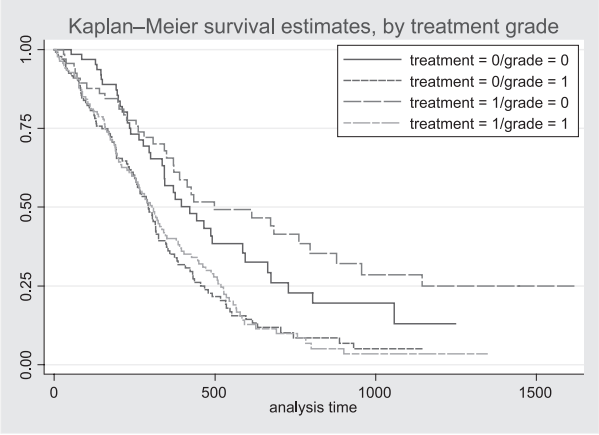

Figure 2: Kaplan–Meier survival estimates, by treatment (`treatment`) and grade of malignancy (`grade`) in the `glioma2` dataset

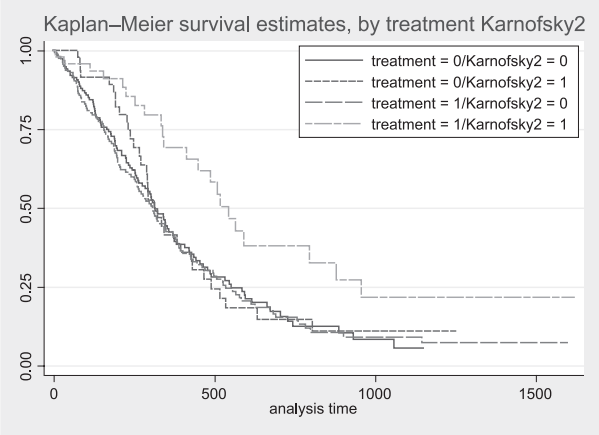

Figure 3: Kaplan–Meier survival estimates, by treatment (`treatment`) and Karnofsky index (type 2) (`Karnofsky2`) in the `glioma2` dataset

When running a log-rank test for each of the covariates alone as prognostic factors, the differences between the categories for survival were found to be significant at the 5% level apart from grade of malignancy and aphasia.

The interaction of time from first symptom and treatment was found to be significant at the 5% level (p-value 0.03) with a hazard ratio for the interaction term of 0.58 and a confidence interval from 0.35 to 0.96, which is wide. Similarly, both specifications of the Karnofsky index were found to have a significant interaction with treatment (p-values of 0.002 and 0.031) and similar interaction hazard ratios of 0.64 and 0.66 (confidence interval 1: 0.49–0.82; confidence interval 2: 0.49–0.89). A multiplicative interaction term was created between treatment and the covariates. The Kaplan–Meier survival curves for both levels of the interaction term between treatment and the Karnofsky index (type 1) are shown in figure 4. These data also suggest a significant interaction between the Karnofsky index (type 1) and treatment.

(*Continued on next page*)

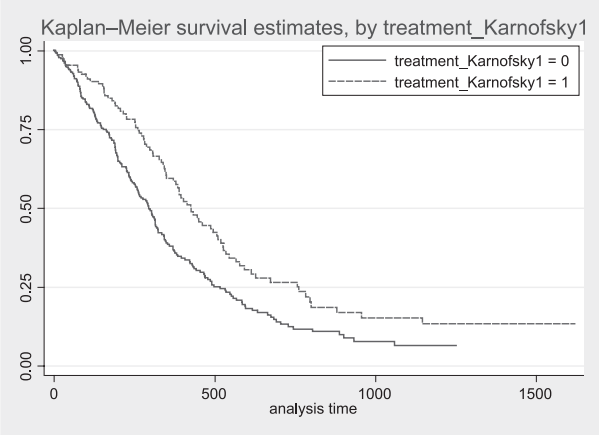

Figure 4: Kaplan–Meier survival estimates for interaction between treatment and Karnofsky index (type 1) (`treatment_Karnofsky1`) in the `glioma2` dataset

The interactions of grade of malignancy and aphasia with treatment illustrated were, however, not found to be significant (p-values 0.37 and 0.39).

3 Model and computation

We wrote a Stata 8 program and dialog to more readily show treatment–covariate interaction effects. The program produces tables and graphics of the interaction effects. This and the next two sections describe first the mathematical background for the calculations and then the program setup. Two trial examples are given at the end.

The model underlying the calculations is based on a 2 × 2 table for interactions as illustrated in table 1. For the Cox proportional hazards model, the hazard ratio between treatment = 1 and treatment = 0, while the covariate is equal to 0, is λ. Similarly, we arrive at a hazard ratio of υ between the covariate being equal to 1 and 0, while treatment is equal to 0. We then define the RHR as τ, which illustrates the interaction effect and is derived as

$$\text{RHR} = \frac{\left(\frac{\lambda\upsilon\tau}{\upsilon}\right)}{\left(\frac{\lambda}{1}\right)} = \tau$$

A similar definition arises when looking at the logistic regression model, as the parameters remain the same, but we are dealing with odds ratios instead of hazard ratios. So again we can use table 1 for illustration and define the ROR as τ.

Table 1: A 2×2 table of hazards in a model with interaction effects

	Treatment = 0	Treatment = 1
Covariate = 0	1	λ
Covariate = 1	υ	$\lambda\upsilon\tau$

Let A denote the treatment and Z a covariate of interest. The overall hazard is calculated using

$$h(t|A) = h_0(t)\exp(\alpha_1 A)$$

where α_1 is the coefficient for the treatment variable, whereas the hazards in the two groups as well as the hazard for the RHR are based on the model

$$h(t|A,Z) = h_0(t)\exp(\beta_1 A + \beta_2 Z + \beta_{12} AZ)$$

We can estimate λ by β_1 and υ by β_2. The interaction term is given by β_{12}.

The logistic option uses logistic regression. The overall treatment odds ratio is estimated using

$$\pi(A) = \frac{\exp(\alpha_0 + \alpha_1 A)}{1 + \exp(\alpha_0 + \alpha_1 A)}$$

The odds ratio in both levels of the covariate and the ROR are based on the following model

$$\pi(A,Z) = \frac{\exp\{g(A,Z)\}}{1 + \exp\{g(A,Z)\}}$$

for

$$g(A,Z) = \beta_0 + \beta_1 A + \beta_2 Z + \beta_{12} AZ$$

where β_0 is the coefficient on the constant term; β_i, $i = 1, 2$, are the coefficients on the independent variables; and β_{12} denotes the coefficient for the interaction term.

The graphical output of this program is based on forest plots—pictorial presentations of the hazard or odds ratio with corresponding confidence intervals. A more detailed description of forest plots and their history may be found in Lewis and Clarke (2001).

4 Design of the dialog

The `fintplot` command may be invoked by using the **fintplot** menu and its associated dialog box. A table of output contains the overall treatment hazard ratio, the hazard ratio in both groups of the prognostic factor chosen, and an estimate of the RHR or ROR for interaction. A forest plot is also displayed. The program has an `overview` dialog box to provide a forest plot of the overall treatment hazard or odds ratio and RHRs or RORs for up to five covariates with treatment. Calculations are performed in the ado-files `fintplot` and `fintplotk`. The default method of analysis is the Cox proportional hazards model.

`fintmenu` can be executed by typing `fintmenu on` and a new item, **fintplot**, will appear on the Stata menu bar under **User**. This menu can be turned off again by typing `fintmenu off`.

4.1 Forest plot and table for interaction

Selecting **User** > **fintplot** > **fintplot - overview** or **User** > **fintplot** > **fintplot - detail** will open a new window titled *fintplot - overview* or *fintplot - detail*, respectively. The following description will concentrate on the *fintplot - detail* dialog box; however, the *fintplot - overview* dialog box may be used the same way. The dataset used in the analysis must be `stset` before using this dialog box if the Cox proportional hazards model is to be used and the covariate levels need to be binary. The user may decide on sensible binary levels for the covariates that are of further interest by first using the *fintplot - overview* dialog. The program also allows logistic regression by checking the *Logistic regression* box in the **Main** tab.

The *fintplot - detail* dialog allows both `by()` and `if` to be executed separately or at the same time. Variables used for the `by()` option of the program must be discrete and can be entered in the **by** tab under *Separate by observations*. If the Cox proportional hazards model is chosen, the program also allows stratification. The variable to be used for stratification must be entered in the **by** tab under *Stratify by observations*. Under the **if/in** tab, the *Create...* button allows easier construction of the logical argument. The confidence level may be set before running the program in the usual way—typing `set level #`. Finally, if the log scale is preferred for the forest plot, one needs to check the box for *Log scale* in the **Main** tab. This option will not change the table.

5 Illustration using two cancer trials

The examples given below illustrate the program, using the Glioma and Low Infant Birth Weight studies. Because tests for interactions were not predefined in the protocol, interpret the results with caution.

5.1 Forest plot for an interaction of two different covariates with treatment

The first example was run using the `glioma2` dataset described above. More information on this study is available in an article by Ulm et al. (1989).

Overview

We `stset` the data before running the main analysis. We use the *fintplot - overview* dialog to corroborate our answers from section 2 and run an interaction analysis on grade of malignancy (`grade`) and the two categories of the Karnofsky index (`Karnofsky1` and `Karnofsky2`). On the **Main** tab of the dialog box, select *Treatment variable* from

the *Variables for test* box and enter `treatment` in the *Treatment variable* box that appears; select *Covariate 1* from the *Variables for test* box and enter `grade` in the *Covariate 1* box that appears; select *Covariate 2* from the *Variables for test* box and enter `Karnofsky1` in the *Covariate 2* box that appears; and select *Covariate 3* from the *Variables for test* box and enter `Karnofsky2` in the *Covariate 3* box that appears. The *Cox proportional hazards* box should already be checked for you. Clicking either **OK** or **Submit** produces the following table and the graph in figure 5:

```
. fintplotk treatment grade Karnofsky1 Karnofsky2, logistic(0) logscale(0)
OVERALL HAZARD RATIO
```

Factor	lnHR	HR	[95% Conf.	Interval]
overall HR	-.10629226	.89916182	.70907197	1.1402114

```
INTERACTIONS WITH treatment
```

Factor	lnRHR	RHR	[95% Conf.	Interval]
grade	.24755585	1.2808909	.74894992	2.1906424
Karnofsky1	-.80227556	.44830765	.27112886	.74127023
Karnofsky2	-.65781756	.51798057	.28376617	.9455104

```
Analysed using Cox proportional hazards model
```

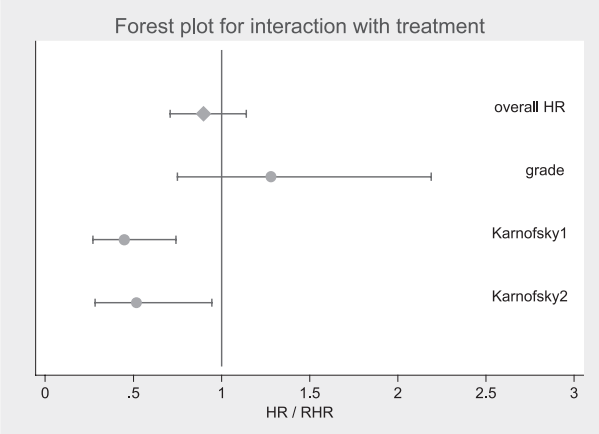

Figure 5: Forest plot output for interaction of treatment with three different covariates from the `glioma2` dataset. The interaction terms given are RHR.

From both the graph and table, we arrive at the same results as in section 2. However, we suggest that the graph using the forest plots is easier to interpret than the Kaplan–Meier plots because the forest plots provide point estimates as well as corresponding confidence intervals. Thus the user can discern whether an interaction effect is significant from looking only at the graph, which is not the case for Kaplan–Meier plots.

Detail

For this second run, we decided to look at the possible interaction between treatment (`treatment`) and two different binary categories of the Karnofsky index (`Karnofsky1` and `Karnofsky2`), as these had been identified as having a significant interaction effect with `treatment`. Figure 6 illustrates how we enter the information into the dialog box.

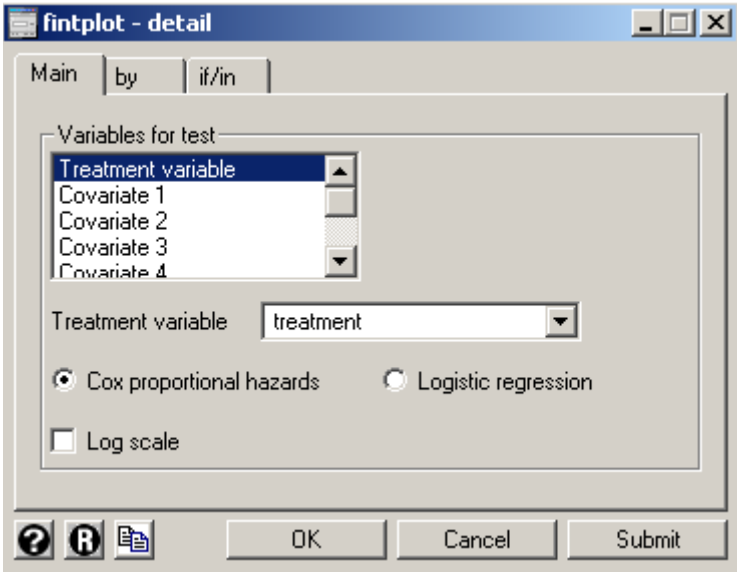

Figure 6: Analysis of two interactions under the Cox model

The `treatment` variable should always be entered first—select *Treatment variable* from the *Variables for test* box and enter `treatment` in the *Treatment variable* box that appears on the **Main** tab. Next select *Covariate 1* from the *Variables for test* box and enter `Karnofsky1` in the *Covariate 1* box that appears; and select *Covariate 2* from the *Variables for test* box and enter `Karnofsky2` in the *Covariate 2* box that appears. Again the *Cox proportional hazards* box should already be checked for you. Upon pressing **OK** or **Submit**, we obtain the following output:

```
. fintplot treatment Karnofsky1 Karnofsky2, logistic(0) logscale(0)
-> interaction with Karnofsky1
```

Factor	lnHR	HR	[95% Conf.	Interval]
overall HR	-.10629226	.89916182	.70907197	1.1402114
Karnofsky1=0	.46974751	1.5995903	1.0625393	2.4080888
Karnofsky1=1	-.44747156	.63924239	.32618748	1.2527484

Factor	lnRHR	RHR	[95% Conf.	Interval]
interaction	-.80227556	.44830765	.27112886	.74127023

```
-> interaction with Karnofsky2
```

Factor	lnHR	HR	[95% Conf.	Interval]
overall HR	-.10629226	.89916182	.70907197	1.1402114
Karnofsky2=0	.0340827	1.0346702	.79335045	1.349394
Karnofsky2=1	-.82794355	.43694692	.22492848	.84881476

Factor	lnRHR	RHR	[95% Conf.	Interval]
interaction	-.65781756	.51798057	.28376617	.9455104

```
Analysed using Cox proportional hazards model
```

The log-hazard ratios and hazard ratios in both levels of the factor and the overall hazard ratio calculated without adjustment for covariates are given, as well as confidence intervals. This output is split into both categories of the Karnofsky index (`Karnofsky1` and `Karnofsky2`). Most importantly, the second table for each categorization gives the log RHR and RHR for the interaction between treatment and the Karnofsky index. All coefficients were obtained using table 1.

Figure 7 illustrates the forest plot output by the program for these interactions. For the plot of `treatment` and `Karnofsky1`, the confidence interval for the first level of `Karnofsky1` is too wide for the table. It has been truncated at a value of 2.5. Both the tables and forest plots show evidence of an interaction between treatment and the Karnofsky index with an RHR of 0.45 or 0.52 depending on the specification.

(*Continued on next page*)

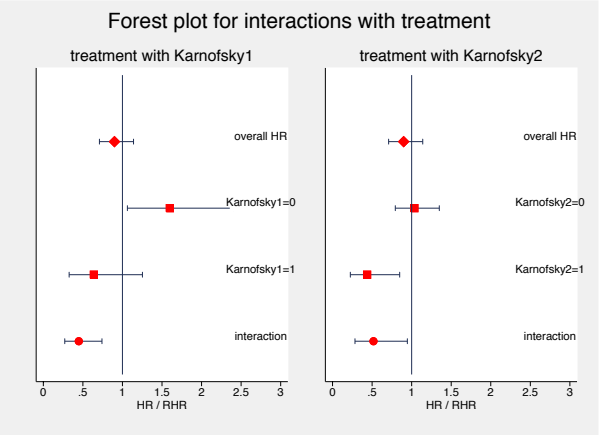

Figure 7: Forest plot for interaction of treatment with two categories of the Karnofsky index. The interaction term given is the lnRHR. ⧫, overall hazard ratio for treatment without differentiating by factor; ■, estimated hazard ratio in the two groups; •, RHR for the interaction.

5.2 Forest plot of an interaction of one covariate with treatment using both by() and log scale options

The data used here (`birth.dta`) originate from a study of the risk factors associated with low infant birth weight. Data collection took place at Baystate Medical Center in Springfield, Massachusetts, during 1986. Information was gathered on the age of the mother (`age`), smoking status during pregnancy (`smoke`), and the mother's weight in pounds at the last menstrual period (`lwt`). Birth weight in grams was also gathered; however, we retained only the low birth weight (`low`) category where 1 = birth weight $<$ 2,500 g. More information on the analysis of this dataset is given in Hosmer and Lemeshow (2000).

Hosmer and Lemeshow (2000) suggest splitting `lwt` into two categories (`lwd`) where 1 denotes a weight of less than 110 pounds. Furthermore, they have investigated a possible interaction between `smoke` and `lwd` split by `age`. Hence we create a new variable, `age5`, that takes on the value 2 for `age` $>$ 25 and 1 otherwise.

We will be using logistic regression here. The dialog box is invoked as before. We need to enter an outcome variable for the events; in the **Main** tab, select *Outcome variable (logistic)* from the *Variables for test* scroll box and enter `low` in the *Outcome variable* box that appears. Next, select *Covariate 1* from the *Variables for test* scroll box and enter `lwd` in the *Covariate 1* box that appears; repeat this step to enter the second covariate, `smoke`. Furthermore, we check the box for *Logistic regression* and *Log scale*. To split the data by `age5`, switch to the **by** tab and enter `age5` as a variable under *Separate by observations*. Figures 8 and 9 illustrate these steps.

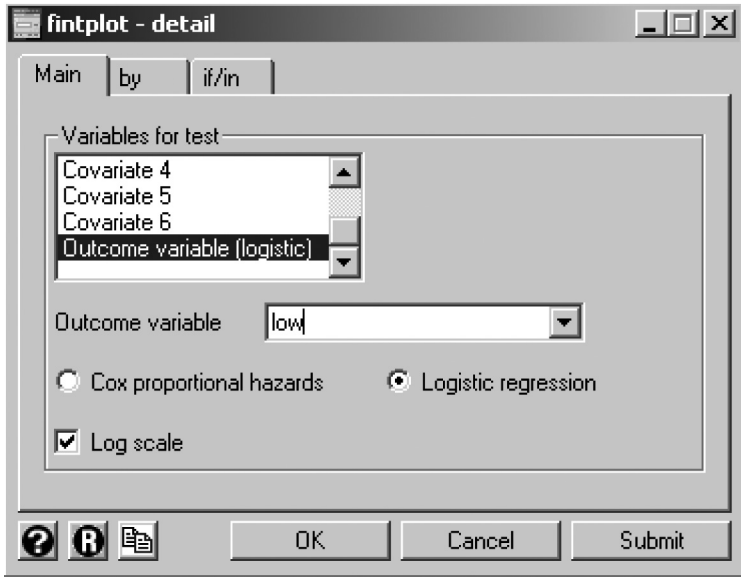

Figure 8: Input of outcome variable for logistic regression

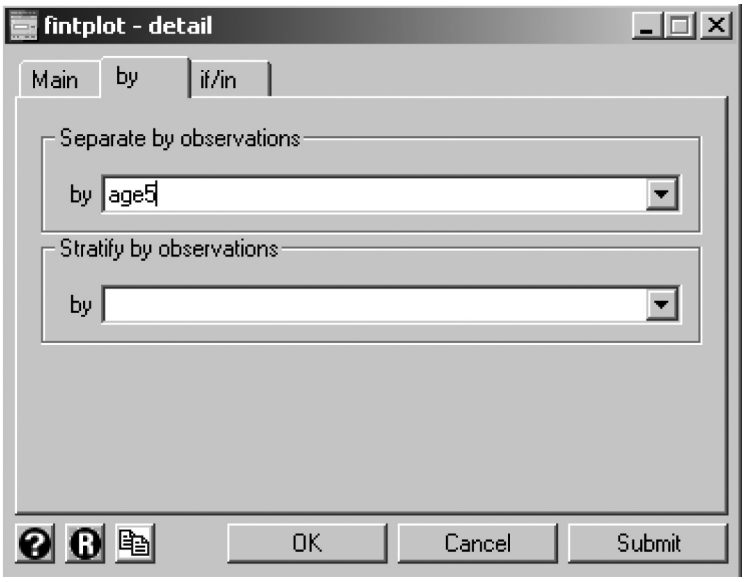

Figure 9: Use of the `by()` option

Once we press the **OK** or **Submit** button, we obtain the following output:

```
. fintplot low lwd smoke, by(age5) logistic(1) logscale(1)
Response variable: low
```

-> for age5==1

Factor	lnOR	OR	[95% Conf.	Interval]
overall OR	.5389965	1.7142857	.71798501	4.0930876
smoke=0	.82198005	2.275	.71135751	7.2757016
smoke=1	5.6333333	279.59254	.00766853	10193868

Factor	lnROR	ROR	[95% Conf.	Interval]
interaction	-.55801451	.57234432	.09691536	3.3800424

Response variable: low

-> for age5==2

Factor	lnOR	OR	[95% Conf.	Interval]
overall OR	2.0918641	8.1	2.2292439	29.431503
smoke=0	2.7725887	16	2.4137899	106.05728
smoke=1	21.005128	1.326e+09	1.709e-15	1.028e+33

Factor	lnROR	ROR	[95% Conf.	Interval]
interaction	-1.5293952	.21666667	.0157211	2.9860787

Analysed using logistic regression

This output can be read in the same way as in the first example. However, here we have a split by `age5`. The forest plot is illustrated in figure 10. We can hence illustrate the potential influence of other variables.

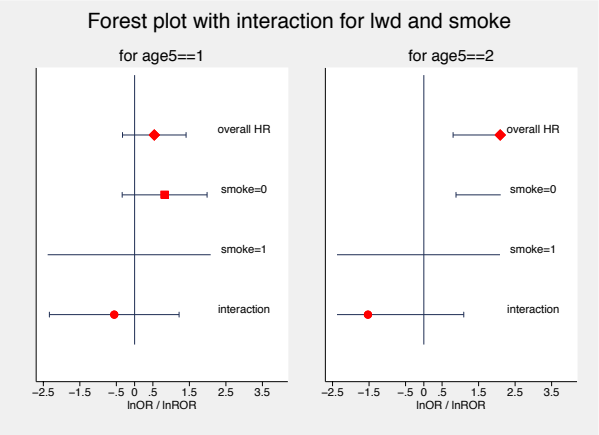

Figure 10: Forest plot using logistic regression, log scale, and `by()` options for the Low Infant Birth Weight dataset. Symbols are defined in figure 7.

The output from both the table and the forest plot suggest no evidence of an interaction between smoking and weight at the last menstrual period when we separate the data by `age5`. However, because of wide confidence intervals due to the few data points available for each group, the analysis is not conclusive. Also the estimate of the log-odds ratio for `smoke=1` is off the scale for the forest plot.

6 Conclusions

Analyzing an intervention's effect in subgroups of patients becomes more and more important to allow for more individual patient care. Hence we provide a Stata tool to express such interactions both quantitatively and visually within a 2×2 table framework. It is flexible in the options it provides and operates under either the Cox proportional hazards or the logistic regression model.

In the presence of a treatment by covariate interaction, one can then determine whether drug efficacy differs for specific high- or low-risk subgroups. Similarly, in epidemiological studies, one may wish to establish whether there is a difference in risk between smokers and nonsmokers in the development of a certain disease. An oftenmentioned example is a study of Danish porcelain painters, which found that the adverse effects of cobalt exposure on lung function were more severe among smokers than nonsmokers (Raffn et al. 1988). Such an analysis should always be planned and based on clinically meaningful subgroups.

The program we presented considers multiplicative interaction effects. However, even without such a multiplicative effect, if two risk factors are individually important, the presence of both in the same patient may lead to a level of risk significantly greater than if either of the risk factors was present alone. These additive effects may be clinically relevant.

Our examples have concentrated on medical applications. However, the forest plot analysis is also applicable in the social sciences. Hout (1984) considers factors influencing occupational mobility. The main impact is made by socioeconomic background but other subgroups include autonomy and the degree of specialization. Other potentially interesting topics include marriage (Mare 1991) and voting behaviors (Bartels 2000).

7 References

Bartels, L. M. 2000. Partisanship and voting behavior, 1952–1996. *American Journal of Political Science* 44: 35–50.

Gail, M., and R. Simon. 1985. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* 41: 361–372.

Hosmer, D. W., Jr., and S. Lemeshow. 2000. *Applied Logistic Regression*. 2nd ed. New York: Wiley.

Hout, M. 1984. Status, autonomy, and training in occupational mobility. *American Journal of Sociology* 89: 1379–1409.

Lewis, S., and M. Clarke. 2001. Forest plots: Trying to see the wood and the trees. *British Medical Journal* 322: 1479–1480.

Mare, R. D. 1991. Five decades of educational assortative mating. *American Sociological Review* 56: 15–32.

Peterson, B., and S. L. George. 1993. Sample size requirements and length of study for testing interaction in a $1 \times k$ factorial design when time-to-failure is the outcome. *Controlled Clinical Trials* 14: 511–522.

Raffn, E., S. Mikkelsen, D. G. Altman, J. M. Christensen, and S. Groth. 1988. Health effects due to occupational exposure to cobalt blue dye among plate painters in a porcelain factory in Denmark. *Scandinavian Journal of Work and Environmental Health* 14: 378–384.

Shuster, J., and J. van Eys. 1983. Interaction between prognostic factors and treatment. *Controlled Clinical Trials* 4: 209–214.

Simon, R. 2002. Bayesian subset analysis: Application to studying treatment-by-gender interactions. *Statistics in Medicine* 21: 2909–2916.

Thall, P. F., and J. M. Lachin. 1986. Assessment of stratum–covariate interactions in Cox's proportional hazards regression model. *Statistics in Medicine* 5: 73–83.

Uesaka, H. 1993. Test for interaction between treatment and stratum with ordinal responses. *Biometrics* 49: 123–129.

Ulm, K., C. Schmoor, W. Sauerbrei, G. Kemmler, U. Aydemir, B. Mueller, and M. Schumacher. 1989. Strategien zur auswertung einer therapiestudie mit der ueberlebenszeit als zielkriterium. *Biometrie und Informatik in Medizin und Biologie* 20: 171–205.

Xiang, A., H. Sather, and S. Azen. 1994. Power considerations for testing an interaction in a $2 \times k$ factorial design with a failure time outcome. *Controlled Clinical Trials* 15: 489–502.

About the authors

Friederike Barthel is a medical statistician at the UK Medical Research Council Clinical Trials Unit. She currently works on sample-size issues, particularly concerning multistage trials, and treatment–covariate interaction designs. Friederike has taught undergraduate courses in statistics at the University of Westminster and Kingston University.

Patrick Royston is a medical statistician with 30 years' experience, with a strong interest in biostatistical methodology and in statistical computing and algorithms. He now works in cancer clinical trials and related issues. Currently, he is focusing on problems of model building and validation with survival data, including prognostic factor studies; on parametric modeling of survival data; and on new trial designs.