* [Analyzing Interval-Censored Survival Data (Generalized Linear Models)](http://127.0.0.1:54857/help/topic/com.ibm.spss.modeler.tutorial/spss/tutorials/genlin_ulcer_intro.htm)

# Analyzing Interval-Censored Survival Data (Generalized Linear Models)

When analyzing survival data with interval censoring—that is, when the exact time of the event of interest is not known but is known only to have occurred within a given interval—then applying the Cox model to the hazards of events in intervals results in a complementary log-log regression model.

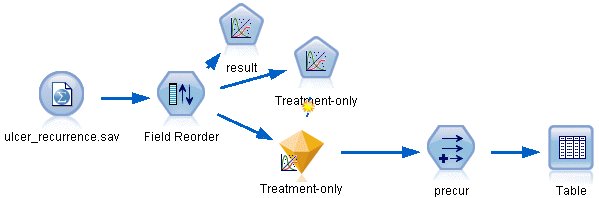
Partial information from a study designed to compare the efficacy of two therapies for preventing the recurrence of ulcers is collected in ulcer\_recurrence.sav. This dataset has been presented and analyzed elsewhere [1](http://127.0.0.1:54857/help/topic/com.ibm.spss.modeler.tutorial/spss/tutorials/genlin_ulcer_intro.htm" \l "fntarg_1). Using generalized linear models, you can replicate the results for the complementary log-log regression models.

This example uses the stream named ulcer\_genlin.str, which references the data file ulcer\_recurrence.sav. The data file is in the Demos folder and the stream file is in the streams subfolder.

# Creating the Stream

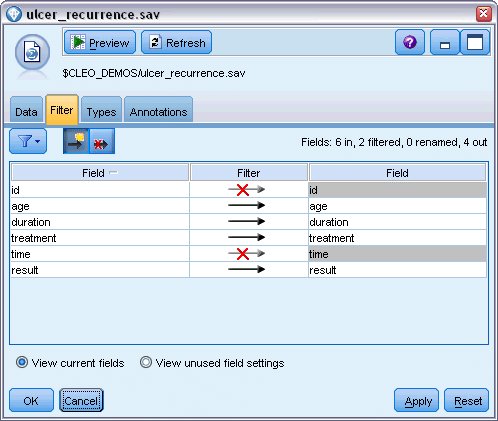
1. Add a Statistics File source node pointing to ulcer\_recurrence.sav in the Demos folder.

*Figure 1. Sample stream to predict ulcer recurrence*



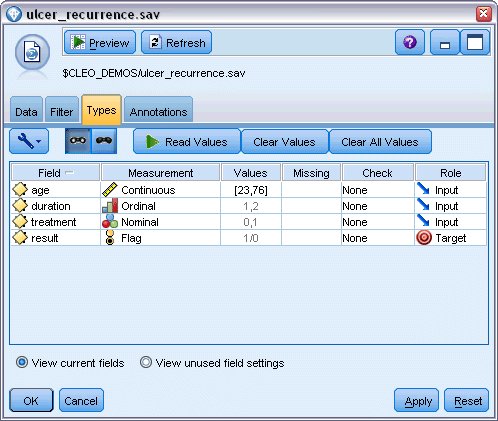
1. On the Filter tab of the source node, filter out id and time.

*Figure 2. Filter unwanted fields*



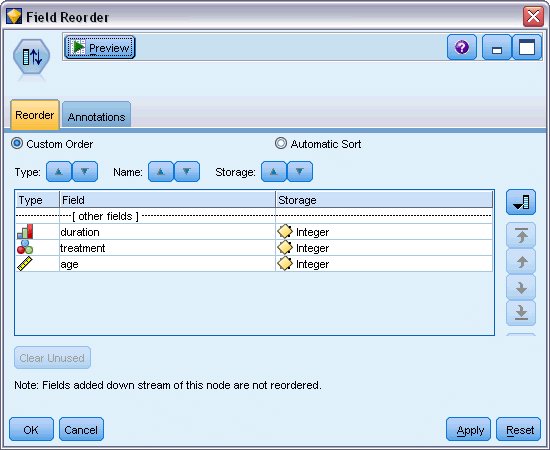
1. On the Types tab of the source node, set the role for the result field to **Target** and set its measurement level to **Flag**. A result of 1 indicates that the ulcer has recurred. All other fields should have their role set to **Input**.
2. Click **Read Values** to instantiate the data.

*Figure 3. Setting field role*



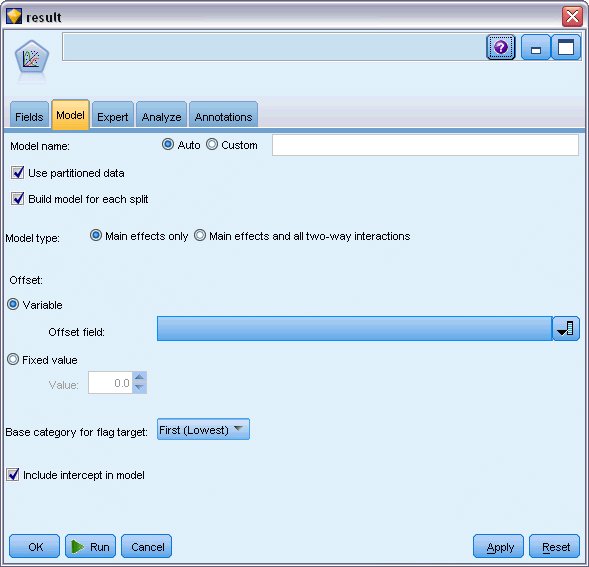
1. Add a Field Reorder node and specify duration, treatment, and age as the order of inputs. This determines the order in which fields are entered in the model and will help you try to replicate Collett's results.

*Figure 4. Reordering fields so they are entered into the model as desired*



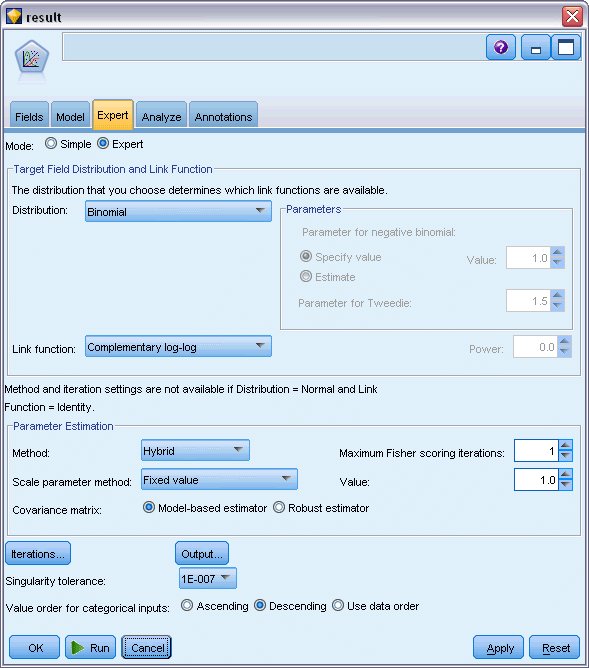
1. Attach a GenLin node to the source node; on the GenLin node, click the **Model** tab.
2. Select **First (Lowest)** as the reference category for the target. This indicates that the second category is the event of interest, and its effect on the model is in the interpretation of parameter estimates. A continuous predictor with a positive coefficient indicates increased probability of recurrence with increasing values of the predictor; categories of a nominal predictor with larger coefficients indicate increased probability of recurrence with respect to other categories of the set.

*Figure 5. Choosing model options*



1. Click the **Expert** tab and select **Expert** to activate the expert modeling options.
2. Select **Binomial** as the distribution and **Complementary log-log** as the link function.
3. Select **Fixed value** as the method for estimating the scale parameter and leave the default value of 1.0.
4. Select **Descending** as the category order for factors. This indicates that the first category of each factor will be its reference category; the effect of this selection on the model is in the interpretation of parameter estimates.

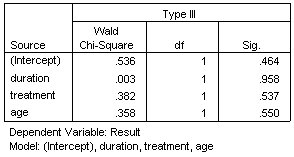
*Figure 6. Choosing expert options*



1. Run the stream to create the model nugget, which is added to the stream canvas, and also to the Models palette in the upper right corner. To view the model details, right-click the nugget and choose **Edit** or **Browse**.

# Tests of Model Effects

*Figure 1. Tests of model effects for main-effects model*

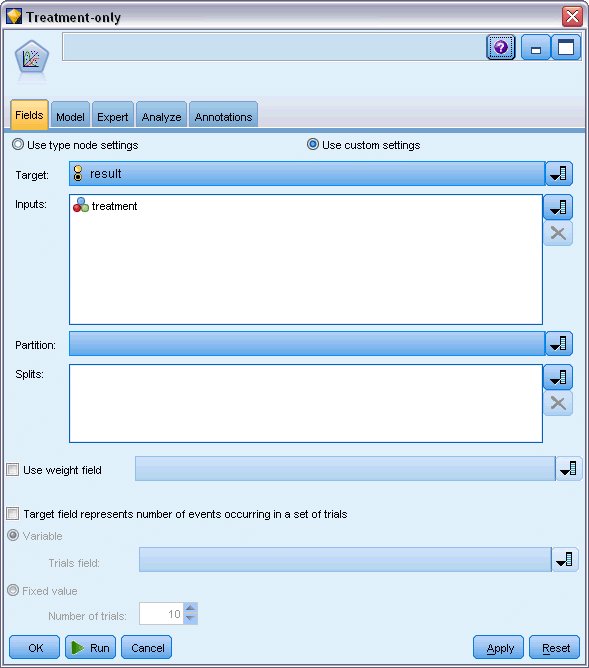


None of the model effects is statistically significant; however, any observable differences in the treatment effects are of clinical interest, so we will fit a reduced model with just the treatment as a model term.

# Fitting the Treatment-Only Model

1. On the Fields tab of the GenLin node, click **Use custom settings**.
2. Select result as the target.
3. Select treatment as the sole input.

*Figure 1. Choosing field options*

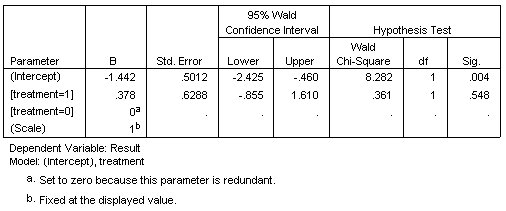


1. Run the stream and open the resulting model nugget.

On the model nugget, select the **Advanced** tab and scroll to the bottom.

# Parameter Estimates

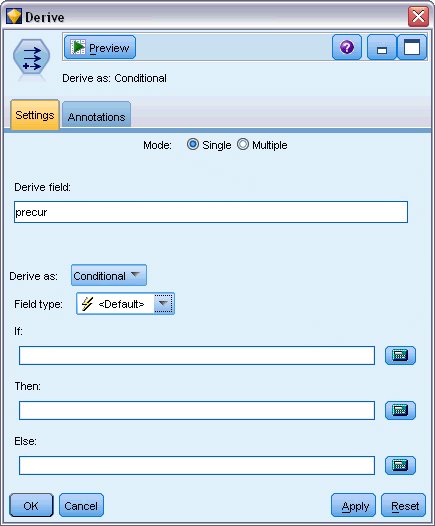
*Figure 1. Parameter estimates for treatment-only model*



The treatment effect (the difference of the linear predictor between the two treatment levels; that is, the coefficient for [treatment=1]) is still not statistically significant, but only suggestive that treatment A [treatment=0] may be better than B [treatment=1] because the parameter estimate for treatment B is larger than that for A, and is thus associated with an increased probability of recurrence in the first 12 months. The linear predictor, (intercept + treatment effect) is an estimate of log(−log(1−P(recur12,t)), where P(recur12, t) is the probability of recurrence at 12 months for treatment t(=A or B). These predicted probabilities are generated for each observation in the dataset.

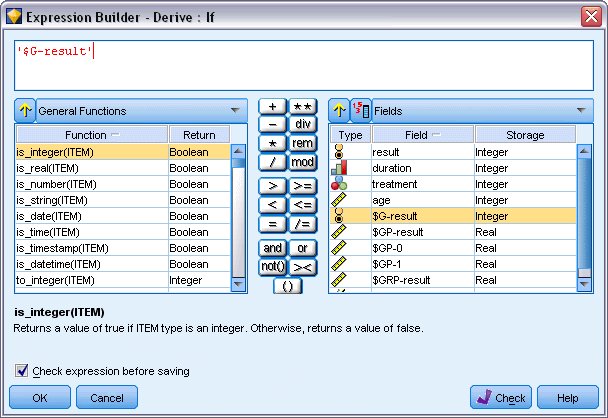
**Predicted Recurrence and Survival Probabilities**

*Figure 1. Derive node settings options*



1. For each patient, the model scores the predicted result and the probability of that predicted result. In order to see the predicted recurrence probabilities, copy the generated model to the palette and attach a Derive node.
2. In the Settings tab, type precur as the derive field.
3. Choose to derive it as **Conditional**.
4. Click the calculator button to open the Expression Builder for the **If** condition.

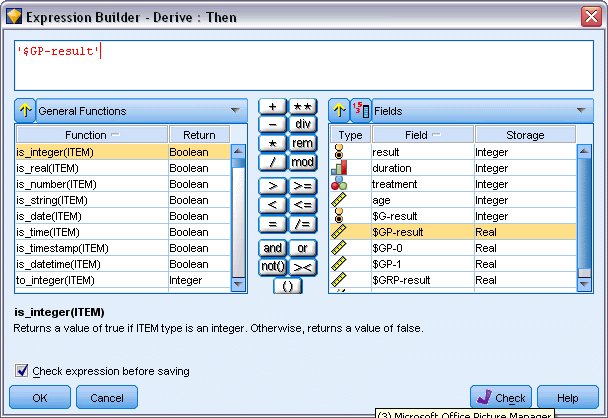
*Figure 2. Derive node: Expression Builder for If condition*



1. Insert the *$G-result* field into the expression.
2. Click **OK**.

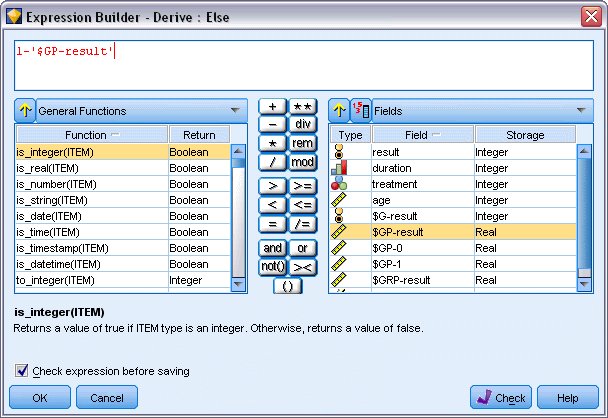
The derive field *precur* will take the value of the **Then** expression when *$G-result* equals 1 and the value of the **Else** expression when it is 0.

*Figure 3. Derive node: Expression Builder for Then expression*



1. Click the calculator button to open the Expression Builder for the **Then** expression.
2. Insert the *$GP-result* field into the expression.
3. Click **OK**.

*Figure 4. Derive node: Expression Builder for Else expression*



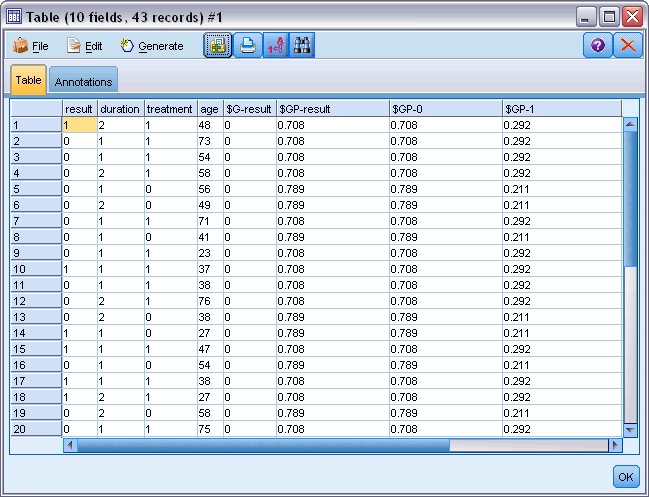
1. Click the calculator button to open the Expression Builder for the **Else** expression.
2. Type 1- in the expression and then insert the *$GP-result* field into the expression.
3. Click **OK**.

*Figure 5. Derive node settings options*



1. Attach a table node to the Derive node and execute it.

*Figure 6. Predicted probabilities*



There is an estimated 0.211 probability that patients assigned to treatment *A* will experience a recurrence in the first 12 months; 0.292 for treatment *B*. Note that 1−P(recur12, t) is the survivor probability at 12 months, which may be of more interest to survival analysts

# Modeling the Recurrence Probability by Period

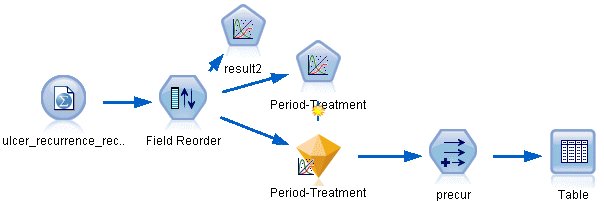
A problem with the model as it stands is that it ignores the information gathered at the first examination; that is, that many patients did not experience a recurrence in the first six months. A "better" model would model a binary response that records whether or not the event occurred during each interval. Fitting this model requires a reconstruction of the original dataset, which can be found in ulcer\_recurrence\_recoded.sav. This file contains two additional variables:

* Period, which records whether the case corresponds to the first examination period or the second.
* Result by period, which records whether there was a recurrence for the given patient during the given period.

Each original case (patient) contributes one case per interval in which it remains in the risk set. Thus, for example, patient 1 contributes two cases; one for the first examination period in which no recurrence occurred, and one for the second examination period, in which a recurrence was recorded. Patient 10, on the other hand, contributes a single case because a recurrence was recorded in the first period. Patients 16, 28, and 34 dropped out of the study after six months, and thus contribute only a single case to the new dataset.

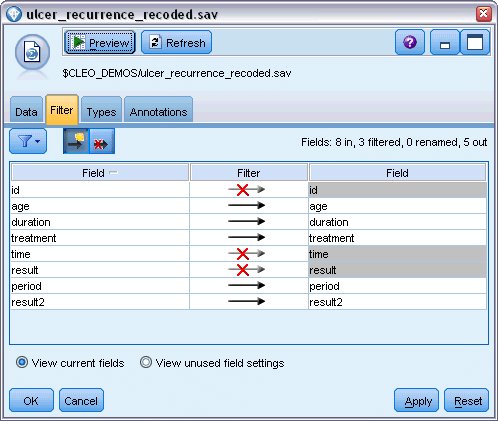
1. Add a Statistics File source node pointing to ulcer\_recurrence\_recoded.sav in the Demos folder.

*Figure 1. Sample stream to predict ulcer recurrence*



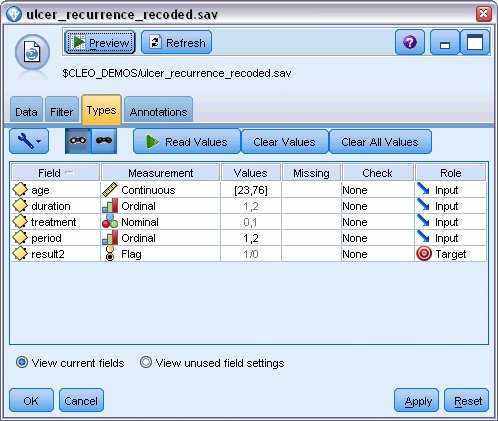
1. On the Filter tab of the source node, filter out id, time, and result.

*Figure 2. Filter unwanted fields*



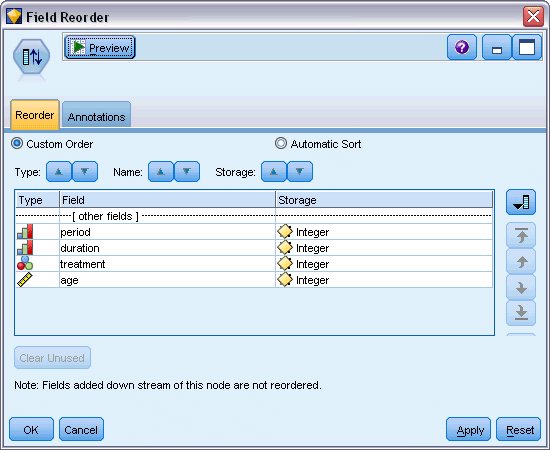
1. On the Types tab of the source node, set the role for the result2 field to **Target** and set its measurement level to **Flag**. All other fields should have their role set to **Input**.

*Figure 3. Setting field role*



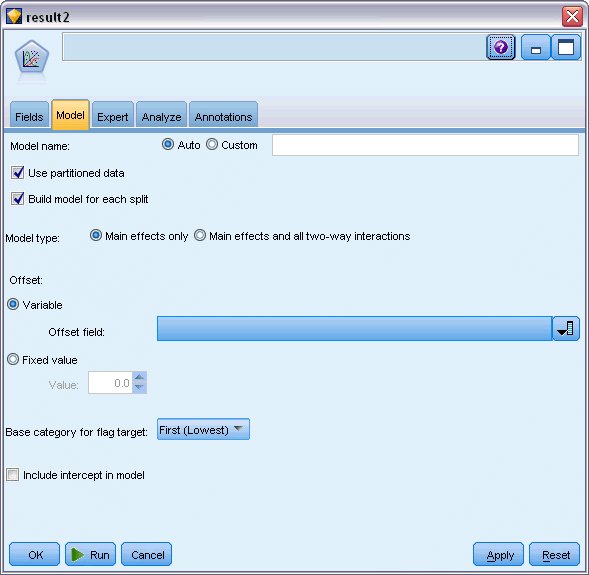
1. Add a Field Reorder node and specify period, duration, treatment, and age as the order of inputs. Making period the first input (and not including the intercept term in the model) will allow you to fit a full set of dummy variables to capture the period effects.

*Figure 4. Reordering fields so they are entered into the model as desired*



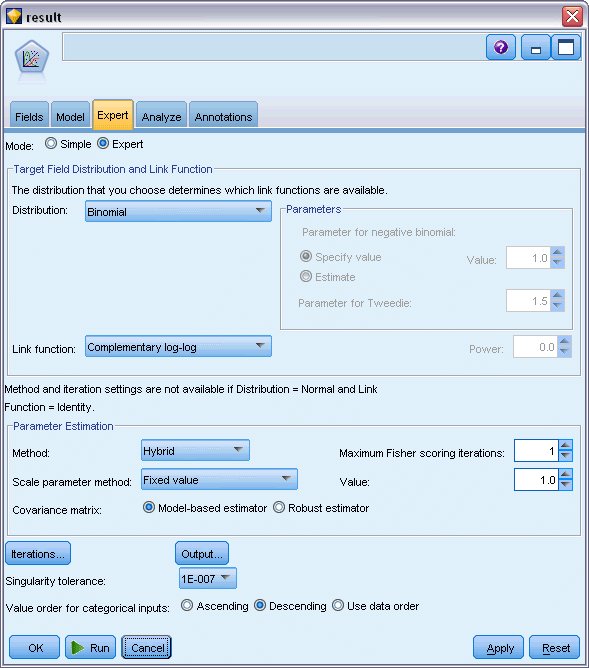
1. On the GenLin node, click the **Model** tab.

*Figure 5. Choosing model options*



1. Select **First (Lowest)** as the reference category for the target. This indicates that the second category is the event of interest, and its effect on the model is in the interpretation of parameter estimates.
2. Deselect **Include intercept in model**.
3. Click the **Expert** tab and select **Expert** to activate the expert modeling options.

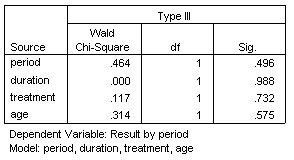
*Figure 6. Choosing expert options*



1. Select **Binomial** as the distribution and **Complementary log-log** as the link function.
2. Select **Fixed value** as the method for estimating the scale parameter and leave the default value of 1.0.
3. Select **Descending** as the category order for factors. This indicates that the first category of each factor will be its reference category; the effect of this selection on the model is in the interpretation of parameter estimates.
4. Run the stream to create the model nugget, which is added to the stream canvas, and also to the Models palette in the upper right corner. To view the model details, right-click the nugget and choose **Edit** or **Browse**.

# Tests of Model Effects

*Figure 1. Tests of model effects for main-effects model*

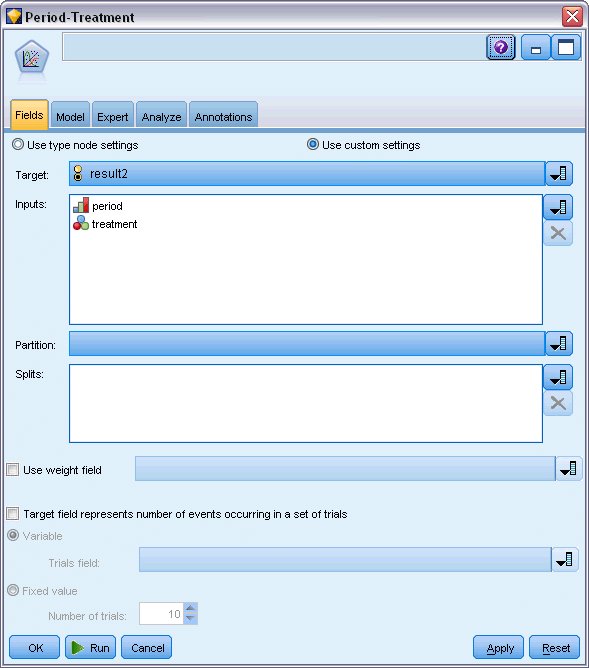


None of the model effects is statistically significant; however, any observable differences in the period and treatment effects are of clinical interest, so we will fit a reduced model with just those model terms.

# Fitting the Reduced Model

1. On the Fields tab of the GenLin node, click **Use custom settings**.
2. Select result2 as the target.
3. Select period and treatment as the inputs.

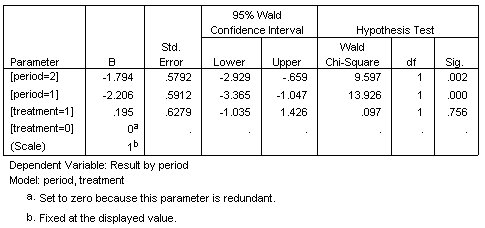
*Figure 1. Choosing field options*



1. Execute the node and browse the generated model, and then copy the generated model to the palette, attach a table node, and execute it.

# Parameter Estimates

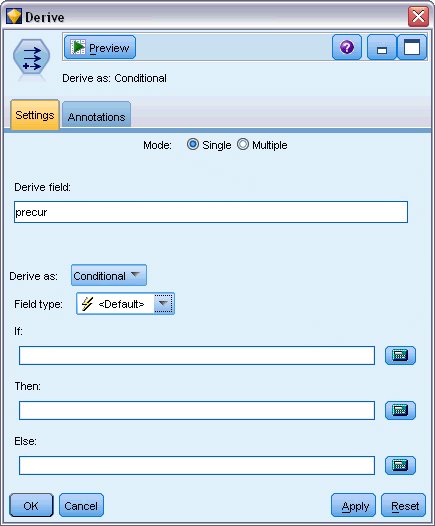
*Figure 1. Parameter estimates for treatment-only model*



The treatment effect is still not statistically significant but only suggestive that treatment A may be better than B because the parameter estimate for treatment B is associated with an increased probability of recurrence in the first 12 months. The period values are statistically significantly different from 0, but this is because of the fact that an intercept term is not fit. The period effect (the difference between the values of the linear predictor for [period=1] and [period=2]) is not statistically significant, as can be seen in the tests of model effects. The linear predictor (period effect + treatment effect) is an estimate of log(−log(1−P(recurp, t)), where P(recurp, t) is the probability of recurrence at the period p(=1 or 2, representing six months or 12 months) given treatment t(=A or B). These predicted probabilities are generated for each observation in the dataset.

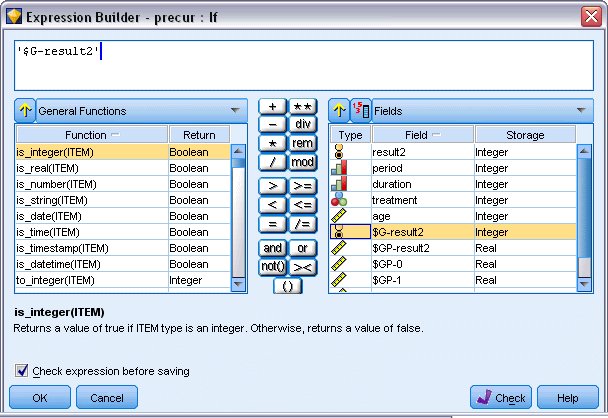
**Predicted Recurrence and Survival Probabilities**

*Figure 1. Derive node settings options*



1. For each patient, the model scores the predicted result and the probability of that predicted result. In order to see the predicted recurrence probabilities, copy the generated model to the palette and attach a Derive node.
2. In the Settings tab, type precur as the derive field.
3. Choose to derive it as **Conditional**.
4. Click the calculator button to open the Expression Builder for the **If** condition.

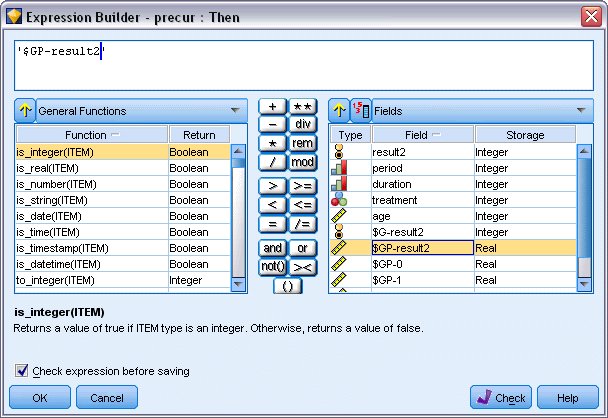
*Figure 2. Derive node: Expression Builder for If condition*



1. Insert the *$G-result2* field into the expression.
2. Click **OK**.

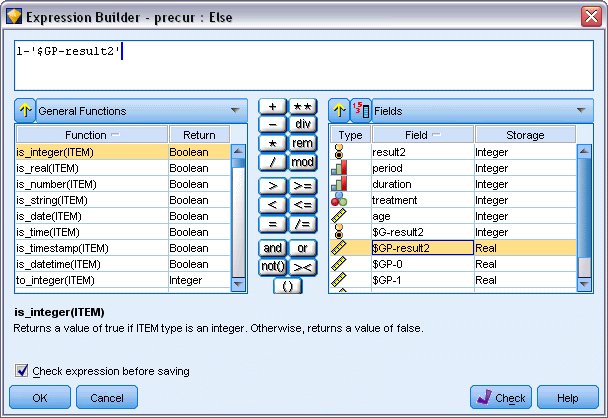
The derive field *precur* will take the value of the **Then** expression when *$G-result2* equals 1 and the value of the **Else** expression when it is 0.

*Figure 3. Derive node: Expression Builder for Then expression*



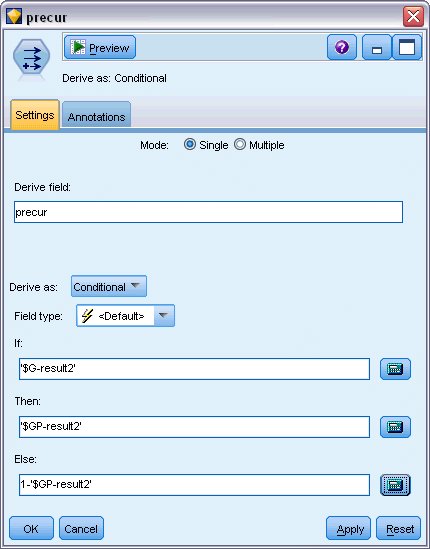
1. Click the calculator button to open the Expression Builder for the **Then** expression.
2. Insert the *$GP-result2* field into the expression.
3. Click **OK**.

*Figure 4. Derive node: Expression Builder for Else expression*



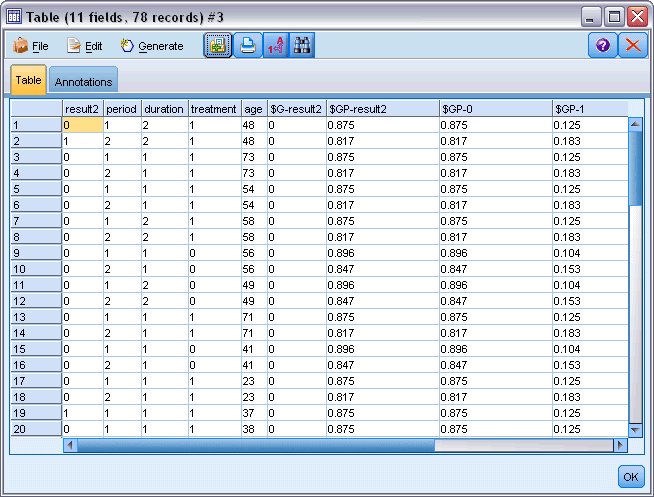
1. Click the calculator button to open the Expression Builder for the **Else** expression.
2. Type 1- in the expression and then insert the *$GP-result2* field into the expression.
3. Click **OK**.

*Figure 5. Derive node settings options*



1. Attach a table node to the Derive node and execute it.

*Figure 6. Predicted probabilities*



| *Table 1. Estimated recurrence probabilities* | | |
| --- | --- | --- |
| **Treatment** | **6 months** | **12 months** |
| A | 0.104 | 0.153 |
| B | 0.125 | 0.183 |

From the estimated recurrence probabilities, the survival probability through 12 months can be estimated as 1−(P(recur1, t) + P(recur2, t)×(1−P(recur1, t))); thus, for each treatment:

*A*: 1 − (0.104 + 0.153\*0.896) = 0.759

*B*: 1 − (0.125 + 0.183\*0.875) = 0.715

which again shows nonstatistically significant support for *A* as the better treatment.

# Summary

Using Generalized Linear Models, you have fit a series of complementary log-log regression models for interval-censored survival data. While there is some support for choosing treatment A, achieving a statistically significant result may require a larger study. However, there are some further avenues to explore with the existing data.

* It may be worthwhile to refit the model with interaction effects, particularly between Period and Treatment group.

Explanations of the mathematical foundations of the modeling methods used in IBM® SPSS® Modeler are listed in the *IBM SPSS Modeler* Algorithms Guide.