

Restricted Cubic Spline Plots of Hazard Ratios, Odds/Risk/Rate Ratios

By David R. Gagnon^{1,2}, Serkalem Demisse^{1,2}, Yuk Lam Ho¹, Katherine Kurgansky¹, Petra Shubert¹ Rachel Ward¹ and Gheorghe Doros^{1,2}

¹MAVERIC [Massachusetts Veterans Epidemiology Research and Information Center], VA Boston Healthcare System, Boston Massachusetts

²Boston University School of Public Health, Department of Biostatistics, Boston Massachusetts

Contact Information: David R. Gagnon, MD MPH PhD
Department of Biostatistics, Boston University School of Public Health
801 Massachusetts Avenue, Rm 328
Boston, MA 02118
gagnon@bu.edu

In examining possible non-linear relationships between exposures and outcomes, it is usually necessary to do a graphical presentation of results in order to evaluate this relationship. Previous work by Hertzmark et al. presented an approach that produced “publication quality” graphics in their GLMCURV9 macro, that can be found at: <https://www.hsph.harvard.edu/donna-spiegelman/software/glmcurv9/>. The introduction of ODS graphics and the PLM procedure have greatly simplified the production of these graphics, resulting in the creation of a relatively simple set of macros with fewer than 300 lines of code each that can be easily modified by the user. Additionally, the use of ODS graphics and PROC PLM have eliminated the need for either SAS GRAPH or SAS IML software, increasing usability to systems that do not have these components installed.

These macros create spline plots for both generalized linear models using **PROC GLIMMIX** [the GLIMSP and GLIMSPM versions] and Cox proportional hazards models using **PROC PHREG** [the PHREGSP and PHREGSPM versions]. The **GLIMMIX** versions have been tested with linear, logistic, Poisson and negative binomial models. For the models using effect modifiers/ stratifying variables [PHREGSPM and GLIMSPM], a single categorical EM variable to interact with the main spline variable is allowed and **requires** it be coded 1 to N. Effect modifiers not involving the main spline variable are allowed in all models by including them in the covariate list.

In addition, the latest versions of ODS graphics allow for exquisite control of the format and appearance of produced graphics. We have implemented this using three steps: 1) by producing a SAS style sheet through the use of **PROC TEMPLATE**; 2) by creating a GRAPHSET macro, which allows selection of the output destination, file format, resolution, orientation and style; and 3) by creating a “discrete attribute map” data set to control graphics attributes [data set GRAPHATTR].

The spline macros

These four macros are designed to produce restricted cubic spline plots for a continuous predictor variable.

- **glimsp** performs generalized linear models splines **without** effect modification/stratification.
- **glimspm** performs generalized linear model splines **with** effect modification/stratification.
- **phregsp** performs Cox regression models splines **without** effect modification/stratification.

- **phregspem** performs Cox regression models splines *with* effect modification/stratification.

PHREG is used for survival analyses and **GLIMMIX** is used for generalized linear models [linear, logistic, Poisson and negative binomial models]. In some instances, we have found that negative binomial models do not converge. One option is to add a “**random _residual_**” statement to the **GLIMMIX** procedures while running a Poisson model to account for overdispersion [not currently included in the macros].

These macros are written and tested in SAS version 9.4, with SAS/STAT version 14.2. Users should verify backward compatibility if using an older version of the SAS system. It is required that the SAS system support the **EFFECT** statement, which allows for splines to be created in the macro and the **STORE** statement, which produces data that can be used by the **PLM procedure**. **PHREG** and **GLIMMIX** support these statements. Proc **SGPLOT** & Proc **SGPANEL** are used for graphics.

If you desire to substitute other procedures here, the following procedures support the **STORE** and **EFFECT** statements as of SAS/STAT 14.2.; **GLIMMIX**, **GLMSELECT**, **LOGISTIC**, **ORTHOREG**, **PHREG**, **SURVEYLOGISTIC**, and **SURVEYREG**.

Splines and Knot Placement

“Natural” or “Restricted Cubic” splines are invoked in all of the macros by an **EFFECT** statement. The default **METHOD** parameter chooses among three different approaches for knot placement:

1. places knots at equal distances along the range of the splined variable;
2. places knots equally according to percentiles of the splined variable; and
3. places knots according to Frank Harrell’s recommended approach [see Regression Modeling Strategies], with knots fairly close to the extreme values and then evenly distributed internal knots. We have extrapolated extreme value knot positions for more than seven knots.

There is a good discussion of alternatives at: <https://blogs.sas.com/content/iml/2017/04/19/restricted-cubic-splines-sas.html>.

Macro Output: Spline Plot, Cumulative Incidence Plot, and Histogram

An initial crude model is executed to produce either a **cumulative incidence plot for PHREG** or a **predicted risk or outcome plot [GLIMMIX versions]**. For **GLIMMIX**, **PROC PLM** is used to display the crude predicted plots. For all macros, PLM displays knot locations in the output. For **PHREG**, the predicted crude survival and cumulative incidence are displayed for the requested survival time [REFTIME parameter] in the output. For models with effect modification, multiple curves are produced.

For multiple curves, the **GRAPHATTR** data set provides color order and other options. Either color or black and white output can be requested [COLOR parameter].

For covariate-adjusted models, **PROC PLM** is used to generate predicted hazard ratio [**PHREG**] or odds/risk ratios [**GLIMMIX**] curves. For Gaussian data [IDENTITY link], these curves are not produced. All plotted values are comparisons with the reference value of the splined variable [REFVAL parameter]. In effect modifier models, a reference value for the splined variable and the reference stratum [EMREF parameter] is selected, making all comparisons against this reference value and stratum. For these predicted comparison plots, an additional plot with a log-transformed X-axis is presented. Axes for

these plots are controlled by the MIN_X,MAX_X,BY_X, and MIN_Y,MAX_Y,BY_Y which control the minimum, maximum and interval for the X and Y axis, respectively. If you wish to use other graphics code, the data sets that feed into SGPLOT and SGPANEL are easily found.

In the **GLIMMIX** versions, when specifying an IDENTITY link, different plots are produced. Two different predicted value plots are produced: 1) displaying the predicted mean value and confidence intervals for the mean and 2) predicted mean value and confidence intervals for observations. Obviously, the confidence intervals for the mean are substantially narrower than those for observations. Additionally, scatterplots of the data are also produced.

A final plot is produced with a histogram of the distribution of the splined variable. For effect modifier models, histograms are presented stratified by the effect modifier variable. The “bins” parameter determines the number of bars in the histogram.

In the **PHREG** versions, time varying covariates are not currently supported, though this should be a simple modification. Time varying coefficients would require some extensive work to allow. Both approaches are currently restricted to independent observations, though it would take minor modifications to allow for correlated data.

There are two forms of output from these macros. Graphics are presented either as parts of a PDF file or in separate files stored on disk. The “Graphics Setup.sas” program sets up the choice of graphics as well as the destination for stored output. You must execute the %graphset macro to set this up. Three examples [lines 96-98] are available to be modified. Other functions of this program include the creation of the SPLINECURVE style template that set some graphics defaults and the creation of the GRAPHATTR data set, which has default color and line schemes for color and black-and-white graphics output. In the **GLIMMIX** versions, you will get a plot of the spline effects, either odds ratios [logistic] or risk ratios [Poisson, negative binomial]; a similar plot with a log scale for the spline variable; a predicted value plot, where predicted rates are for periods of one time unit when there is an offset [OFFSET parameter]; and the density plot for the spline variable, stratified by the effect modifier in EM models. The **PHREG** version will first produce a cumulative incidence plot [several lines for EM models]; a hazard ratio plot; a hazard ratio plot with a log scale for the spline variable; and a density plot for the spline variable, stratified by the effect modifier in EM models.

Also produced are results in the SAS output destination. Output from all regression models are displayed. Calculated knot percentiles for Harrell’s method are printed. Note that having more than seven knots results in calculated knot locations not present in Harrell’s writings – we’ve extrapolated the pattern.

In the Cox model output, crude predicted survival probabilities are presented for the reference value of the spline variable at the time specified by the REFTIME macro variable. In the **PLM** output, you can find the actual locations of the specified knots. In the **GLIMMIX** versions you can also see the **PLM** knot locations.

Note also that each macro produces multiple output data sets that can be further manipulated for additional images or output.

This accompanying ZIP file contains a single SAS program with all of the graphics setup code as well as the SAS macros for the production of publication quality plots of restricted cubic splines (also referred to as natural cubic splines in SAS) of Cox regression models and generalized linear models. . All macros have sample execution code which will work with the sample data set.

The contents of the ZIP file include:

README Restricted Cubic Spline Plots.docx	This document, describing the macros
Demo[DATE].sas	Program which sets up the demonstration data set and runs the examples
fram.sas7bdat	Framingham teaching data set, used for demonstrations
Combined Spline Macros [DATE]	Contains the following code
Graphics Setup	Code which sets up the graphics environment and default options, creating graphattr.sas7bdat [below]
glimspem	GLIMMIX based macro with effect modifiers/stratifying variables
glimsp	GLIMMIX based macro without effect modifiers
phregspem	PHREG based macro with effect modifiers/stratifying variables
phregsp	PHREG based macro without effect modifiers/stratifying variables

Graphics Setup: Setting up the graphics environment.

PROC TEMPLATE

This code creates a style sheet that is used for these graphics. This alters the default graphic options. It uses **PROC TEMPLATE** to do this. On a system where you have write permission, you should only have to run this **PROC TEMPLATE** procedure once to save this style [Styles.splinecurve] into your style library. On other systems, you may have to run this procedure once per session.

This execution of **PROC TEMPLATE** turns off full framing of the graphic [line 18 “frameborder=off”], thickens the axis lines to 3 pixels [line 24, “linethickness= 3px”], defines the default fonts as sans-serif of

various sizes [lines 29-43] and sets the default background and foreground colors to white and black, respectively [lines 44-51].

The GRAPHSET macro

This macro sets up the ODS destination for the graphics produced. This macro creates global macro variables that open the graphics destination [&ods1, &ods2, &ods3] and close the destination [&ods4, which is only needed for the PDF destination]. It also sets a landscape orientation with a 6.8" x 10" frame. This macro should only need to be run once per session. Two example executions of this macro are provided at the end of the macro, for JPEG and PDF output. The default dots-per-inch [DPI] for pixel graphic formats [JPEG, PNG, GIF, etc.] is 300, and going beyond this can cause memory issues. Note that the GPATH parameter is a directory for all but PDF output and must be a directory path and filename for PDF output. PDF output will usually pop up after execution. To execute the spline macros again with a PDF destination, **the popped up PDF must be closed and will be overwritten** unless the destination is changed or the file saved is renamed.

Because there are several instances where the ODS destination is either opened or closed, you may find it necessary to issue an "ODS listing;" statement or similar HTML statement after running one of the spline macros to return to normal operation. It'll be obvious when your next procedure has no errors and produces no output.

The GRAPHATTR data set

This data set creates the default color schemes and other options used in the plots for the effect modification/stratified versions of the macros. This can be modified by changing the color order [the "value" variable] or by changing the colors themselves, using SAS color names or hexadecimal colors. Other variables [markersymbol, markersize, filltransparency, linepattern and linethickness] can also be modified. As presented, this data set has two different ID values – ID='scatter' is for color presentations, while ID='grays' is for gray-scale plots. Both the code to create the data set and a separate copy of the produced data set are provided – use either one.

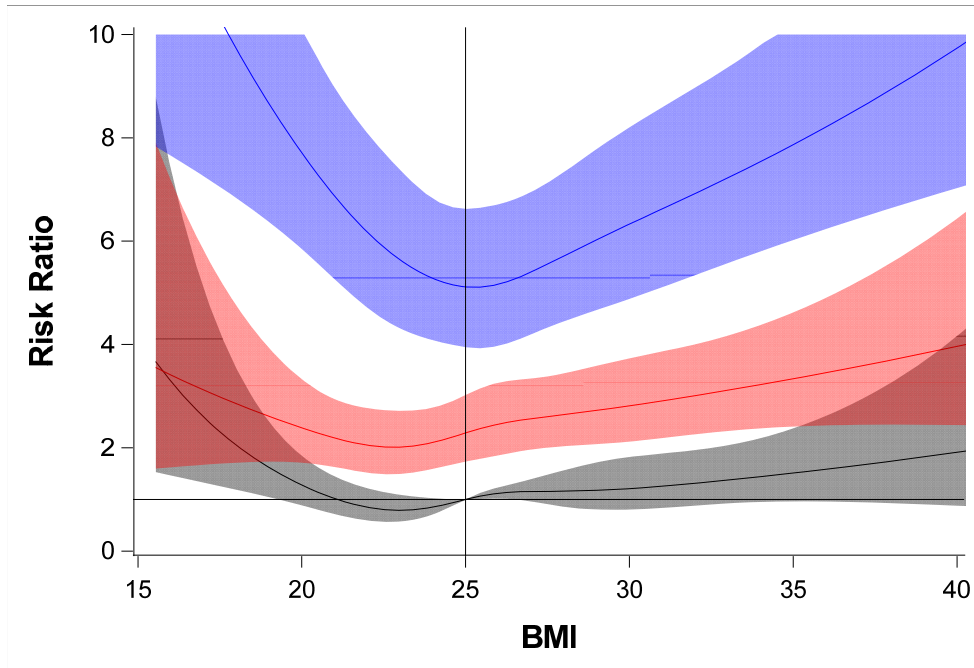
Demo[DATE].sas

This program sets up a data set for demonstration runs. Note that the release date is in the filename. We use the Framingham teaching data set that is publicly available and included in this ZIP file. This can be used with all versions of the macros.

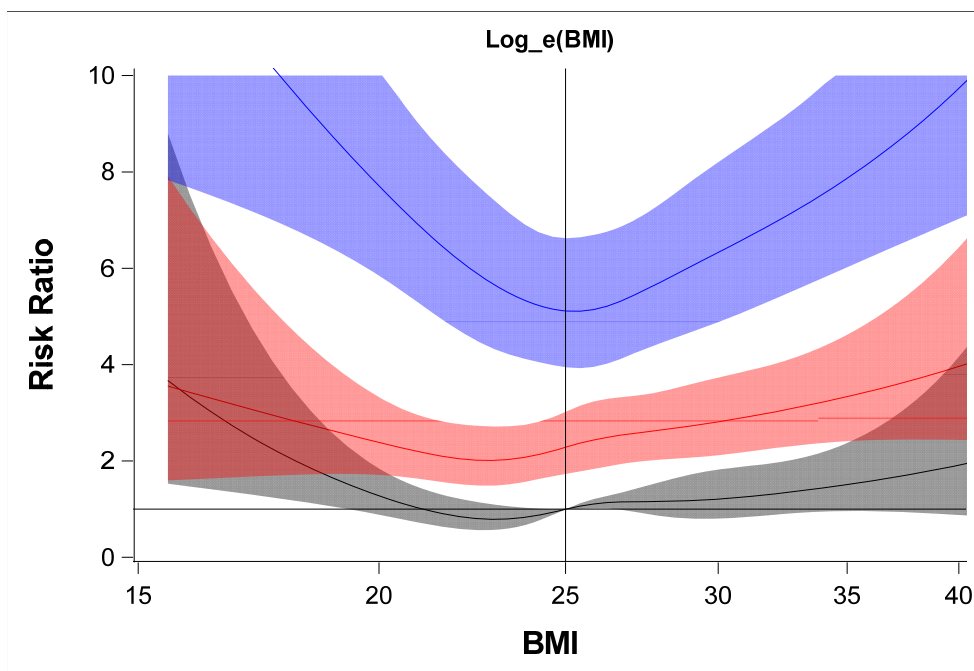
Illustration: Plots using the sample data

The following plots were produced using the sample data set with the glimspemV1 macro using Harrell's knot positions for five knots. Note that these example plots do not touch the Y-axis as the lowest BMI value is greater than fifteen.

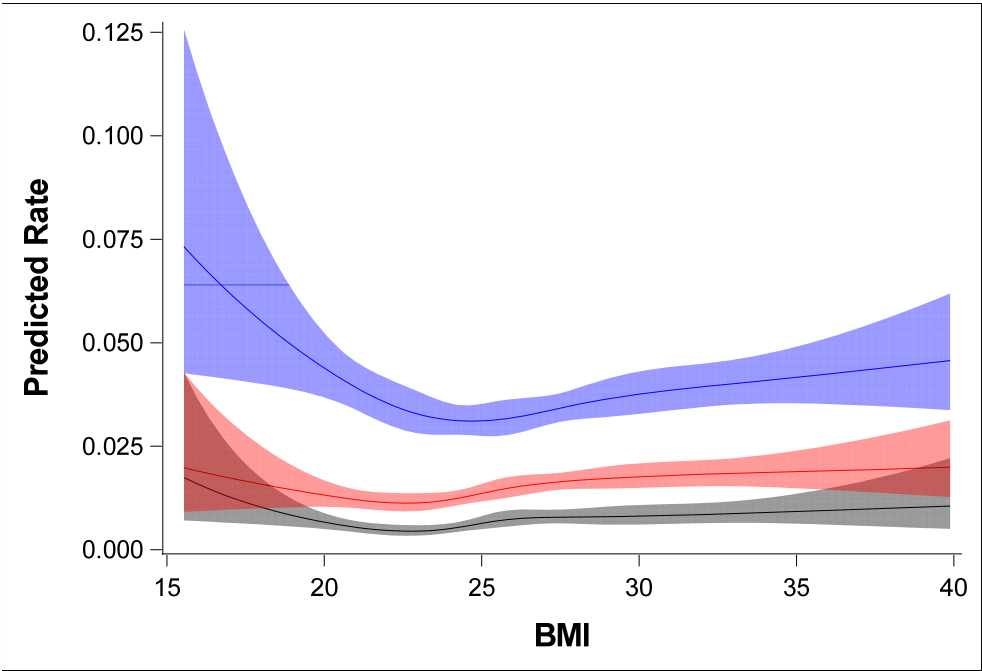
Sample output, risk ratio plot for a Poisson model:



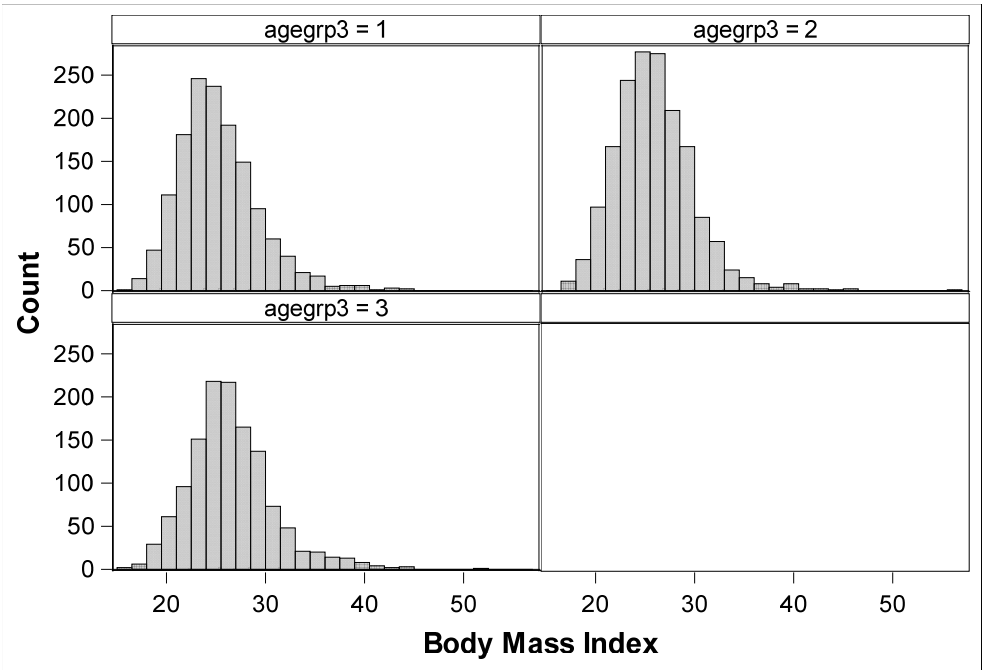
Sample Log Risk Ratio Plot for Poisson Model



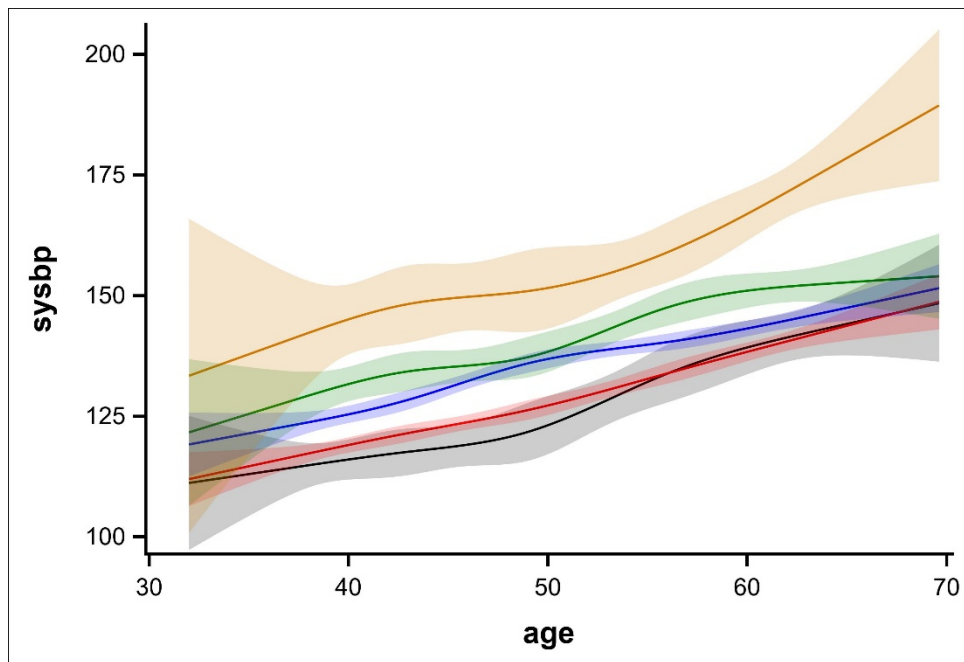
Sample Risk plot for Poisson model:



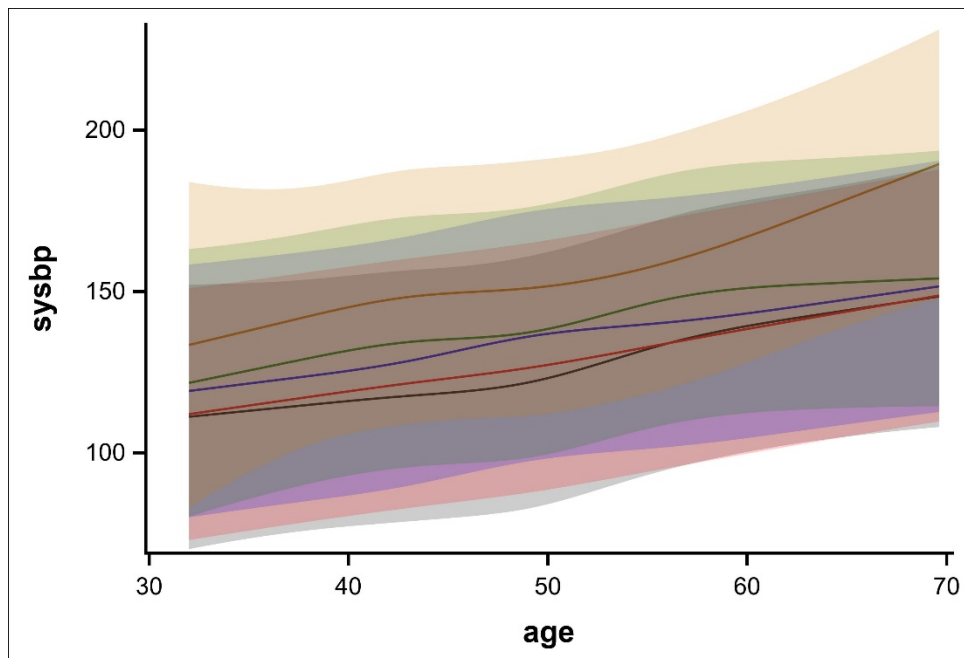
Sample Histogram:



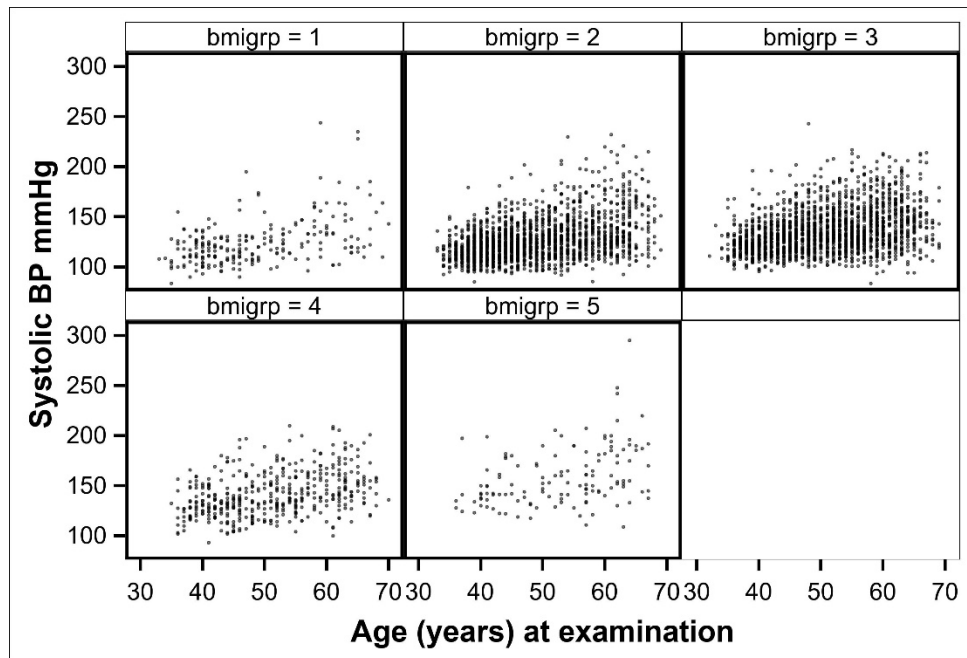
The plots for data with a continuous outcome are a bit different. When using the GLIMSPEM macro with link=IDENTITY, the first plot produced has the predicted splines for each group. Confidence intervals for this plot are based on the confidence intervals for the mean.



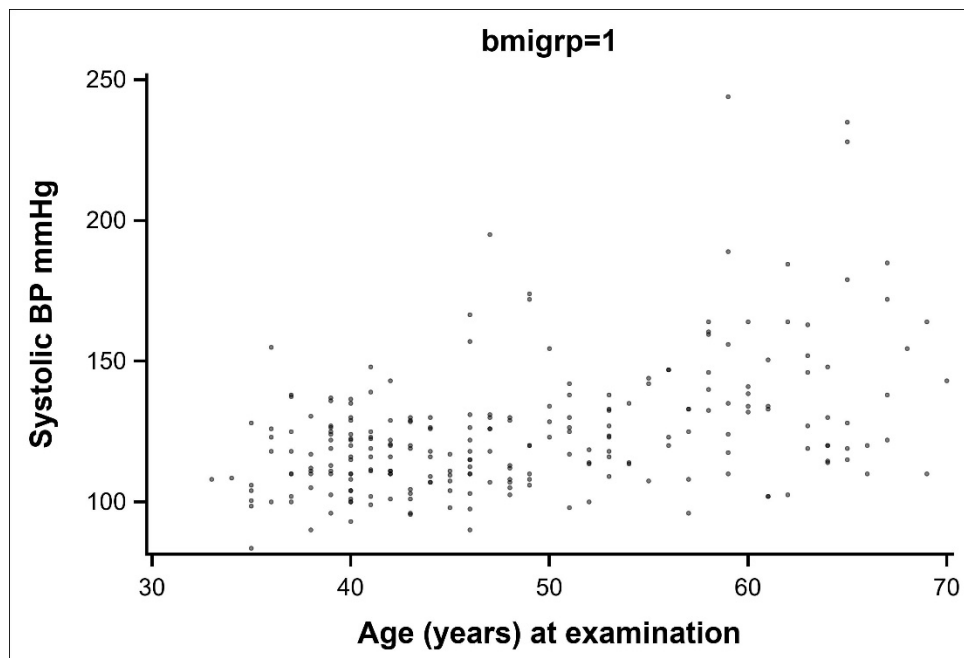
This is in contrast to the next plot, which shows the same predicted values, but where the confidence interval is based on observations. This gives much wider confidence intervals, and may be impractical with multiple groups.



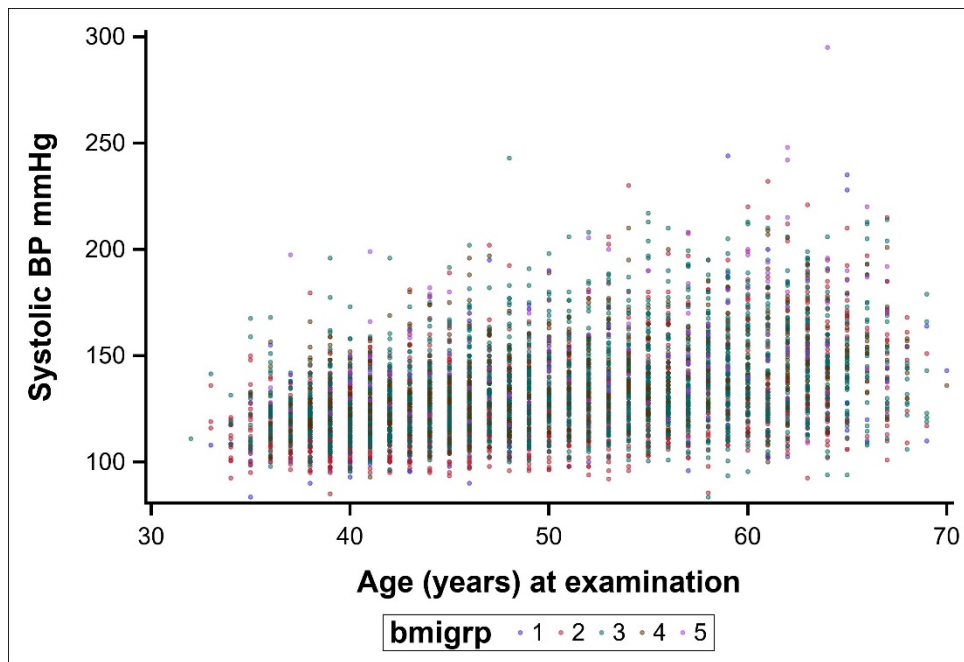
Next, we produce scatterplots showing the raw data by group, first as a panel.



Second, we produce individual scatter plots for each group.



In addition, we have a scatterplot which overlays all of the groups, with different colored circles.



In many situations, this is not very informative. Finally, we have histograms of the distribution of the outcome by group. Note that the number of bins is controlled by the BINS macro parameter.

