How to control finely for confounding using continuous variables that may have a non-linear association with the outcome

Ellen Hertzmark and Donna Spiegelman

August 23, 2012

Abstract

This is a 'cookbook' on how to determine whether a continuous confounder has a nonlinear relation with an outcome, and if it does, how to set up data to control for continuous confounders having a nonlinear relation with the outcome of interest.

1 NOTATION:

`exp` the exposure you are really interested in.
`case` the binary outcome or censoring variable
`time` the survival time, if you have a Cox model.
`conf` the continuous potential confounder of the relation between `exp` and `case` or `time`.
`confx, confmiss` the variables you will use in your multivariate analysis,
if there are any missing values for `conf`.
You want to have a new variable, so that you still have the original variable, `conf`, with its missing values.
First you need to find the mean of `conf` to use for the missing values. If `conf` is very skewed, you should use the median instead of the mean.

```
/* This is to get the mean of conf, which we will call mean_conf.
   It is a number */
proc means data=<data set> mean;  var conf;  run;
```

The coding for `confx` and `confmiss` is

```
confx=conf;  confmiss=0;
if conf eq . then do;  confmiss=1;  confx=mean_conf;  end;
/* NOTE:  you need to know mean_conf from the results of the
          previous code and use this number here. */
```

`adj` the other variables in your model.

2 Outline of procedure

Step 1. Test `conf` for a nonlinear relation with the outcome.
Use %LGTPHCURV9 if your regression is a Cox regression or a logistic regression.
Use %GLMCURV9 if you are using a log-binomial or other generalized linear model.
NOTE: Because of the difference in the links, nonlinearity of logistic models does not imply nonlinearity of log-binomial models, though if the outcome is rare these will be close.
Step 2. If the relation is not significantly nonlinear (based on the p-value in line 1 of the macro output), you are done.
If it is significantly nonlinear, rerun %LGTPHCURV9 or %GLMCURV9 to rerun the selection with rounded versions of the knots from Step 1, but use `CONFX` as your *EXPOSURE*, and include `CONFMISS` in *ADJ*. We suggest rounded values for the convenience of the user. Also, what is printed out by the macros is not to full precision, so the spline variables will be slightly different using rounded knot values, and this may affect the selection.
NOTE that if we allowed the %LGTPHCURV9 or %GLMCURV9 macro to make knots based on `CONFX` we would most likely get different knot points because of the change in the number of observations with values and the 'heaping up' of values at `mean_conf`.

3 Step 1. Check for nonlinearity

3.1 Using %LGTPHCURV9

See %LGTPHCURV9 documentation for use of this macro.

Here are the particulars you need for this specific task——for logistic or Cox models:

```
%lgtphcurv9( data= <name of dataset>,
             exposure= conf,
             select=3,
             nk=21,
             case= case,
                   name of the censoring variable
                   (coded 0=no event, 1=event),
             time= time, (if Cox model)
             model= <LOGISTIC or COX>,
             adj= adj,
                  This includes EXP, as well as all the other
                  covariates in the full model.
             refval= <reference value of CONF for the model>
   );
```

After printing out the models without `conf`, with linear `conf`, and with linear `conf` plus selected spline variables, the macro prints a summary listing the dataset, the range of the "exposure" (i.e. `conf`) in the dataset, the number of observations in the dataset, the adjusting variables, and the names of the selected spline variables, if any.
At the end of the macro output, there will be 3 test results with instructions.
The possible outcomes for the 3 lines, along with instructions, are given in the table below.

Because there is some room for discretion in these choices, we describe *p-values* as ”small” or ”large,” rather than giving a specific cutoff value.

```
Line 1     Line 2     Line 3   What to do
(non-      (overall   (linear
linear     signif.    relation-
relation-  of curve)  ship)
ship)
------     ------     ------   ------------------------------------
.          .          small    no spline variables were chosen.
                               no need to worry about controlling for
                               CONF in a nonlinear way.
.          .          large    no spline variables were chosen.
                               CONF is not significantly related to CASE.
small      small      small    Use results from "spline model with adjusters"
                               in Step 2.
small      small      large    Use results from "spline model with adjusters"
                               in Step 2.
small      large      small    impossible situation
small      large      large    If there were a relationship between
                               CONF and CASE, it would be nonlinear,
                               but there is no significant relationship.
                               If you must include CONF in your model
                               (for subject-matter reasons),
                               use results from "spline model with adjusters"
                               in Step 2.
                               Otherwise, use results from "with adjusters only
                               model in Step 2.
large      small      small    Use results from "linear model with adjusters"
                               in Step 2.
large      small      large    impossible situation
large      large      small    Use results from "linear model with adjusters"
                               in Step 2.
large      large      large    There is no significant relationship between
                               CONFX and CASE.
                               If you must include CONF in your model
                               (for subject-matter reasons),
                               use results from "linear model with adjusters"
                               in Step 2.
```

IN WORDS:
If LINE 1 has a small p-value and LINE 2 has a small p-value (i.e. if the p-values for nonlinearity and for overall significance of the curve are small, such as below .05), you should include the selected spline variables to control for confounding in your main model.
If LINE 1 has a small p-value and you must include `conf` in your model even if it is not significantly related to the outcome, you should treat the situation as if LINE 2 had a small p-value.
If neither LINE 1 nor LINE 2 has a small p-value but LINE 3 has a small p-value, OR you must include `conf` in your model for subject-matter reasons, you should use `conf` in the regression model.

3.2 Using %GLMCURV9

Here are the particulars you need for this specific task for log-binomial models:

```
   %glmcurv9( data= <name of dataset>,
              select=3,
              exposure=conf,
              outcome=case,
              adj=adj,
              link=log, dist= bin,
              knot= <list of rounded knot values from %MAKESPL>
              refval= <reference value for the log-binomial regression>,
              subject= <identifying number for subject>,
              class= <list of class variables, including SUBJECT>
              reptype>= <working covariance matrix type>,
                        default is CS (compound symmetry) or EXCH (exchangeable).
                        If you only have one observation per subject, use IND.
              withinvar= <variable specifying time order of measurements,
                          if you have repeated measures and are not using the
                          default REPTYPE>,
              usegee= <T or F, depending on whether you need to use GEE>
                      If there are multiple observations for some subjects,
                      USEGEE will be automatically set to T.
                      If there is only one observaton  per subject, USEGEE=F
                      for log-binomial, but should be T if you are using
                      the Poisson approximation to the binomial.
);
```

As with %LGTPHCURV9, the macro will tell you the spline variables chosen and will print the results of 3 significance tests. If *USEGEE* = F (e.g. log-binomial model with one observation per subject), the macro gives the results of likelihood ratio tests.
If *USEGEE* = T (e.g. you have repeated measures, or you are using the Poisson distribution instead of the log-binomial model), the macro gives the results of the robust score tests.
To use models other than log-binomial, change *DIST* and *LINK* as desired.
The table in the section "Using %LGTPHCURV9" applies here too.

4 Examples

The following examples are based on a study of death after antiretroviral (ARV) initiation in the Dar es Salaam PEPFAR program. The primary analysis was done using categories/indicators for the continuous variables, but the question was raised whether closer control for potential confounding would affect the RRs of some of the truly categorical values (male sex, TB treatment, TB history, and WHO HIV stage). In this example, we will explore this question for only 2 variables, baseline CD4 count and BMI. The variables in the study are

tsurv	months of followup after ARV initiation
arvdeath	whether died after ARV initiation

msex	male sex (0=no, 1=yes)
age	age at entry to the PEPFAR program
&agecat_	indicators for age at entry to PEPFAR program
BMI	BMI at ARV initiation
&bmicat_	indicators for BMI at ARV initiation
whomaxx	WHO HIV stage (I-II, III, IV)
tbtreatbas	whether on TB treatment at ARV initiation
tbhistbas	whether reported history of TB treatment at ARV initiation
arvcat	first ARV drugs given
hgbbas	hemoglobin at ARV initiation
&hgbcat_	indicators for hemoglobin at ARV initiation
cd4bas	CD4 count at ARV initiation
&cd4bascat_	indicators for CD4 count at ARV initiation

4.1 All indicator model

Here are the results of the model with all variables as indicators.

```
----------------------------------------------------------------------------------

/udd/stleh/doctn/examples.splinecont  Program exspline   10AUG2010   15:04    stle
baseline categorical/indicator model


DATA set is analysis .    TIME is tsurv .  EVENT is arvdeath .


12830 observations  with  1682 events (13.1 %)
-2 Log Likelihood = 1605.969  with 25 degrees of freedom, p= <.0001

                           95% lower     95% upper
                           conf limit    conf limit     Wald
Variable             RR      of RR         of RR       P-value

CD4LT2001          1.54         1.30          1.82     <.0001
CD4LT200M          1.31         1.09          1.57     0.0038
MSEX               1.25         1.13          1.39     <.0001
AGEGP1             0.99         0.63          1.57     0.9770
AGEGP2             0.89         0.77          1.03     0.1222
AGEGP4             1.01         0.90          1.14     0.8178
AGEGP5             1.21         1.04          1.41     0.0161
AGEGPM             2.08         1.43          3.02     0.0001
BMICAT2            0.51         0.46          0.58     <.0001
BMICAT3            0.35         0.27          0.46     <.0001
BMICAT4            0.48         0.32          0.71     0.0003
BMICATM            2.47         2.14          2.84     <.0001
WHOMAXX2           2.32         1.85          2.90     <.0001
WHOMAXX3           4.79         3.81          6.01     <.0001
WHOMAXXM           1.95         1.51          2.51     <.0001
TBTREATBAS1        0.81         0.68          0.97     0.0254
TBTREATBASM        0.58         0.34          1.00     0.0498
TBHISTBAS1         0.78         0.69          0.88     <.0001
TBHISTBASM         1.68         0.94          2.99     0.0804
ARVCAT2            1.01         0.86          1.18     0.9217
ARVCAT3            0.89         0.68          1.17     0.4154
ARVCAT4            0.77         0.62          0.95     0.0139
ARVCATM            1.44         1.26          1.64     <.0001
HGBLT85BAS1        2.14         1.92          2.39     <.0001
HGBLT85BASM        1.27         1.09          1.48     0.0024

================================================================================
```

4.2 Step 1. Run %LGTPHCURV9 with *EXPOSURE* = `cd4bas`

The purpose of this step is to find the knot locations for `cdbas` and to see whether, using only known values of `cd4bas`, the relationship is nonlinear.
For this, we use the indicator sets for all the other continuous variables. The call to %LGTPHCURV9 is

```
title2 'using lgtphcurv9 for cd4';
%lgtphcurv9(data=analysis, exposure=cd4bas, case=arvdeath, time=tsurv,
adj=msex &agegp_ &bmicat_ &whomaxx_
&tbtreatbas_ &tbhistbas_ &arvcat_ &hgblt85bas_,
refval=200,  select=3,  klines=t,
pictname=cd4death05.ps, Hlabel=CD4 at ARV initiation,
footer=Adj for sex age BMI stage TB ARV hgb,
nk=21,
vlabelstyle=h,
vlabel=Relative Risk for Death,
graphtit=CD4 count and Mortality);
```

As usual with %LGTPHCURV9 we get information about the knots.

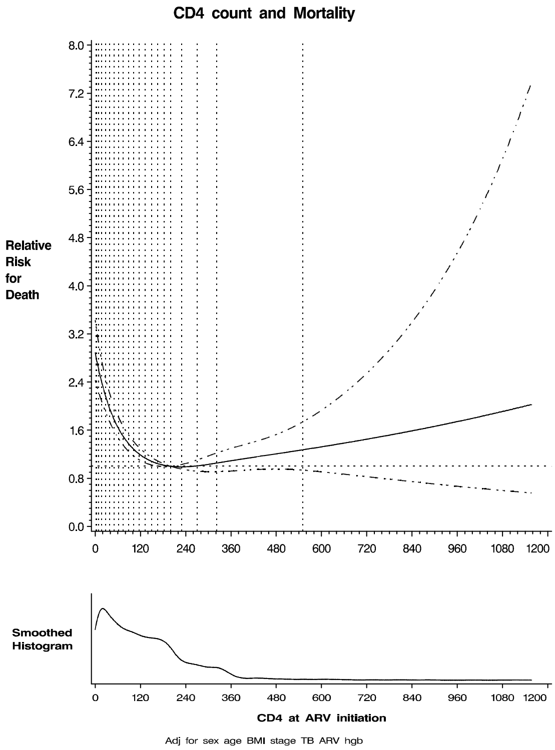
```
================================================================================

/udd/stleh/doctn/examples.splinecont  Program exspline   01SEP2010   11:34    st
using lgtphcurv9 for cd4
Percent of range of CD4BAS below the first knot is 0  .
Percent of range of CD4BAS above the last knot  is 52  .

================================================================================

/udd/stleh/doctn/examples.splinecont  Program exspline   01SEP2010   11:34    st
using lgtphcurv9 for cd4
    Knots for CD4BAS:
    2 4 9 18 27 38 50 62
    75 89 102 117 133 150 166 183
    200 230 271 323 551

================================================================================
```

The summary with the significance tests is

```
================================================================================

/udd/stleh/doctn/examples.splinecont  Program exspline   31AUG2010   12:49    st
using lgtphcurv9 for cd4

    CD4 count and Mortality
    PROC PHREG
    Data set:  ANALYSIS, with 8674 observations
```

```
Time variable name:  TSURV
Censoring variable name:  ARVDEATH with 1198 events and 7476 censored
Exposure of interest: CD4 at ARV initiation
Exposure variable name: CD4BAS
Range of exposure in data used:  0  to 1157
Adjusted for:
      msex  agegp1  agegp2  agegp4  agegp5
      agegpm  bmicat2  bmicat3  bmicat4  bmicatm
      whomaxx2  whomaxx3  whomaxxm  tbtreatbas1  tbtreatbasm
      tbhistbas1  tbhistbasm  arvcat2  arvcat3  arvcat4
      arvcatm  hgblt85bas1  hgblt85basm

Reference value is  USER VALUE:  200
Number of knots: 21
You chose to select spline variables automatically, with sls=.05 and sle=.05
The following spline variables were selected:
     CD4BAS1 CD4BAS2

Name of graph file:  cd4death05.ps

Model w/o exposure of interest, -2 Log Likelihood: 20950.073117
                  Linear Model, -2 Log Likelihood: 19725.960522
                  Spline Model, -2 Log Likelihood: 19660.677138


Line Test Name     Description                        P value
-------------------------------------------------------------


1    Test for      If the P value is small, the
     curvature     relationship between the
   (i.e. non-      exposure and the outcome, if any,
    linear         is non-linear.
    relation)      SEE LINE 2.
                   If the P value is large, the
                   relationship between the
                   exposure and the outcome, if any
                   is linear
                   SEE LINE 3.
                   If the P value is missing, the
                   automatic selection procedure did
                   not select any spline variables.
                   The relationship between the expo-
                   sure and the outcome, if any, is
                   linear.  SEE LINE 3.              <.0001
-------------------------------------------------------------
2    Test for      If LINE 1 indicated a possible
     overall sig-  non-linear relation between the
     nificance     exposure and the outcome,
     of the curve  use this P value for the relation of
                   the EXPOSURE to the CASE or TIME. <.0001
    -------------------------------------------------------------
    3    Test for      If LINE 1 indicated a possible
         linear        linear relation between the
         relation      exposure and the outcome,
                       use this P value AND rerun your
                       model with the parameter
                       PWHICH=LINEAR, to get the graph
                       corresponding to the model of
                       interest (if you intend to use
                       the graph).                          <.0001

================================================================================
```

The graph is



Note that the vertical axis goes very high because of the wide confidence band on the right of the graph where there are few data points. Also note that the locations of the knots (shown by reference lines on the graph) are way over on the left side of the graph.

4.3 Step 2. Run %LGTPHCURV9 with *EXPOSURE* = cd4basx

The purpose of this step is to determine the spline variables that are chosen when cd4basx is used as the *EXPOSURE* and approximate knot locations are specified.

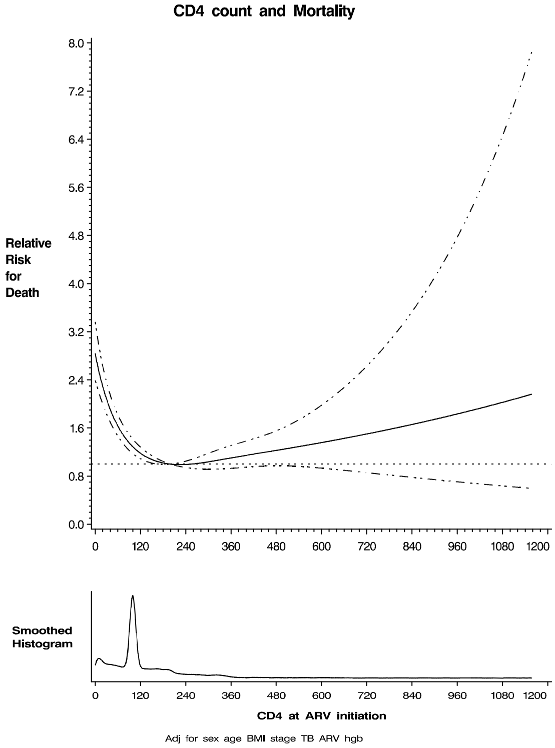
The call to %LGTPHCURV9 is

```
title2 'using lgtphcurv9 for cd4x';
%lgtphcurv9(data=analysis, exposure=cd4basx, case=arvdeath, time=tsurv,
adj=msex &agegp_ &bmicat_ &whomaxx_
&tbtreatbas_ &tbhistbas_ &arvcat_ &hgblt85bas_ cd4lt200m,
refval=200,  select=3,  klines=f,
pictname=cd4xdeath05.ps, Hlabel=CD4 at ARV initiation,
footer=Adj for sex age BMI stage TB ARV hgb,
knot=2 4 9 18 27 38 50 62 75 89 102 117 133 150 166 183 200 230  272 323 551,
testrep=short,
vlabelstyle=h,
vlabel=Relative Risk for Death,
graphtit=CD4 count and Mortality);
```

The summary with the significance tests (short version) is

```
================================================================================


using lgtphcurv9 for cd4x

    CD4 count and Mortality
    PROC PHREG
    Data set:  ANALYSIS, with 12824 observations
    Time variable name:  TSURV
    Censoring variable name:  ARVDEATH with 1681 events and 11143 censored
    Exposure of interest: CD4 at ARV initiation
    Exposure variable name: CD4BASX
    Range of exposure in data used:  0  to 1157
    Adjusted for:
          msex  agegp1  agegp2  agegp4  agegp5
          agegpm  bmicat2  bmicat3  bmicat4  bmicatm
          whomaxx2  whomaxx3  whomaxxm  tbtreatbas1  tbtreatbasm
          tbhistbas1  tbhistbasm  arvcat2  arvcat3  arvcat4
          arvcatm  hgblt85bas1  hgblt85basm  cd4lt200m

    Reference value is  USER VALUE:  200
    Number of knots: 21
    You chose to select spline variables automatically, with sls=.05 and sle=.05
    The following spline variables were selected:
         CD4BASX1 CD4BASX2

    Name of graph file:  cd4xdeath05.ps

    Model w/o exposure of interest, -2 Log Likelihood: 30635.83867
                      Linear Model, -2 Log Likelihood: 28979.715447
                      Spline Model, -2 Log Likelihood: 28914.21106


    Line Test Name                                       P value
    -------------------------------------------------------------

    1    Test for curvature (i.e. non-linear relation) <.0001
    2    Test for overall significance of curve        <.0001
    3    Test for linear relation                      <.0001

--------------------------------------------------------------------------------
```

The graph is



Note that the upper graphs are much the same for cd4bas and cd4basx, but the smoothed histogram for cd4basx shows a much more pronounced peak at 100, the level to which cd4basx was set when cd4bas was missing.

4.4 Step 3. Run %LGTPHCURV9 with *EXPOSURE* = bmibas

Now redo step 1 with *EXPOSURE*=bmibas to find the knots and see whether this variable has a nonlinear relationship with case. In this model, we need to use the spline variables for cd4basx chosen above, as well as the indicators for all the other covariates. To make the spline variables for tt cd4basx, we call %MAKESPL.

```
%makespl(data=analysis, splvbl=cd4basx, makepts=f, refval=100, outdat=analysis1,
knot1=2 4 9 18 27 38 50  62 75 89 102 117 133 150 166 183 200 230 272 323 551);

title2 'using lgtphcurv9 for bmi';
%lgtphcurv9(data=analysis1, exposure=bmibas, case=arvdeath, time=tsurv,
adj=msex &agegp_ &whomaxx_
&tbtreatbas_ &tbhistbas_ &arvcat_ cd4basx cd4basx1 cd4basx2 cd4lt200m &hgblt85bas_
refval=18.5,  select=3,  klines=t,
pictname=bmideath05.ps, Hlabel=BMI at ARV initiation,
footer=Adj for sex age stage TB ARV CD4 hgb,
nk=21,
testrep=short,
vlabelstyle=h,
vlabel=Relative Risk for Death,
graphtit=BMI and Mortality);
```

The summary with the significance tests is

```
================================================================================


using lgtphcurv9 for bmi

    BMI and Mortality
    PROC PHREG
    Data set:  ANALYSIS1, with 12116 observations
    Time variable name:  TSURV
    Censoring variable name:  ARVDEATH with 1422 events and 10694 censored
    Exposure of interest: BMI at ARV initiation
    Exposure variable name: BMIBAS
    Range of exposure in data used:  6.8571395874  to 47.562408447
    Adjusted for:
          msex  agegp1  agegp2  agegp4  agegp5
          agegpm  whomaxx2  whomaxx3  whomaxxm  tbtreatbas1
          tbtreatbasm  tbhistbas1  tbhistbasm  arvcat2  arvcat3
          arvcat4  arvcatm  cd4basx  cd4basx1  cd4basx2
          cd4lt200m  hgblt85bas1  hgblt85basm

    Reference value is  USER VALUE:  18.5
    Number of knots: 21
    You chose to select spline variables automatically, with sls=.05 and sle=.05
    The following spline variable was selected:
         BMIBAS2

    Name of graph file:  bmideath05.ps

    Model w/o exposure of interest, -2 Log Likelihood: 25726.710976
                      Linear Model, -2 Log Likelihood: 24545.750517
                      Spline Model, -2 Log Likelihood: 24499.353215


    Line Test Name                                       P value
    -------------------------------------------------------------

    1    Test for curvature (i.e. non-linear relation) <.0001
    2    Test for overall significance of curve        <.0001
    3    Test for linear relation                      <.0001

================================================================================
```

4.5 Step 4. Run %LGTPHCURV9 with *EXPOSURE* = bmibasx

The purpose of this step is to determine the spline variables that are chosen when bmibasx is used as the *EXPOSURE* and approximate knot locations are specified. Before we call %LGTPHCURV9 with exposure bmibasx we have to make the spline variables for CD4. In the call to

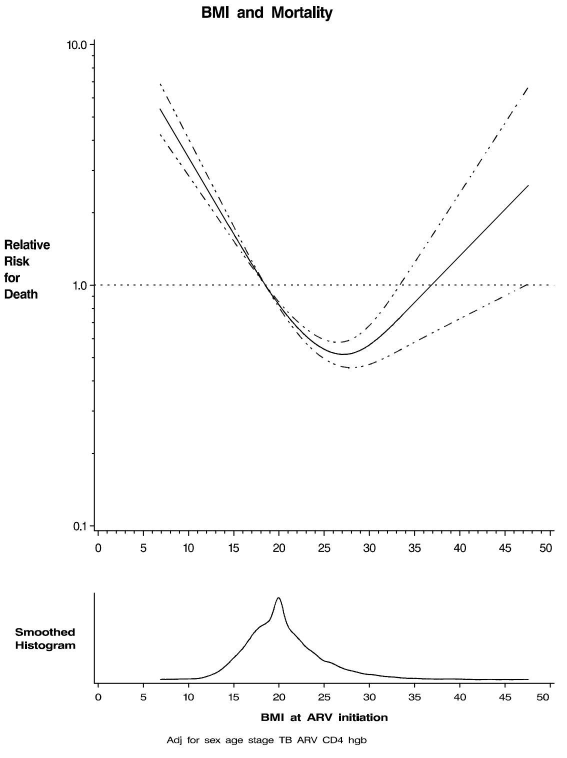
```
%makespl(data=analysis, splvbl=cd4basx, makepts=f, refval=100, outdat=analysis1,
knot1=2 4 9 18 27 38 50  62 75 89 102 117 133 150 166 183 200 230 272 323 551);

title2 'using lgtphcurv9 for bmi';
%lgtphcurv9(data=analysis1, exposure=bmibasx, case=arvdeath, time=tsurv,
adj=msex &agegp_ &whomaxx_
&tbtreatbas_ &tbhistbas_ &arvcat_ cd4basx cd4basx1 cd4basx2 cd4lt200m &hgblt85bas_
refval=18.5,  select=3,  klines=f,
pictname=bmixdeath05.ps, Hlabel=BMI at ARV initiation,
footer=Adj for sex age stage TB ARV CD4 hgb,
knot=12.6 14.3 15.6 16.4 17.1 17.6 18.1 18.6 19 19.5 19.9 20.3 20.9
21.4 21.9 22.6 23.4 24.4 25.8 27.9 33.8,
testrep=short,
vlabelstyle=h,
axordvlog10=t,
vlabel=Relative Risk for Death,
graphtit=BMI and Mortality);
```

the summary with the significance tests is

```
================================================================================


using lgtphcurv9 for bmi

    BMI and Mortality
    PROC PHREG
    Data set:  ANALYSIS1, with 12817 observations
    Time variable name:  TSURV
    Censoring variable name:  ARVDEATH with 1681 events and 11136 censored
    Exposure of interest: BMI at ARV initiation
    Exposure variable name: BMIBASX
    Range of exposure in data used:  6.8571395874  to 47.562408447
    Adjusted for:
          msex  agegp1  agegp2  agegp4  agegp5
          agegpm  whomaxx2  whomaxx3  whomaxxm  tbtreatbas1
          tbtreatbasm  tbhistbas1  tbhistbasm  arvcat2  arvcat3
          arvcat4  arvcatm  cd4basx  cd4basx1  cd4basx2
          cd4lt200m  hgblt85bas1  hgblt85basm  bmicatm

    Reference value is  USER VALUE:  18.5
    Number of knots: 21
    You chose to select spline variables automatically, with sls=.05 and sle=.05
```

```
The following spline variable was selected:
     BMIBASX2

Name of graph file:  bmixdeath05.ps

Model w/o exposure of interest, -2 Log Likelihood: 30633.563795
                  Linear Model, -2 Log Likelihood: 28909.737481
                  Spline Model, -2 Log Likelihood: 28866.051854


Line Test Name                                       P value
-------------------------------------------------------------

1    Test for curvature (i.e. non-linear relation) <.0001
2    Test for overall significance of curve        <.0001
3    Test for linear relation                      <.0001
```

The graph is



4.6 Summary of results

In the table below, we summarize the RRs and their 95% confidence intervals for the truly categorical variables for 3 models (all categorical, CD4 splines and BMI categorical, CD4 and BMI splines).

predictor	all categorical	CD4 splines BMI categorical	both splines

```
                 RR (CI)            RR (CI)              RR (CI)

Male sex      1.25 (1.13-1.39)   1.22 (1.10-1.35)      1.25 (1.13-1.39)
TB treatment  0.81 (0.68-0.97)   0.81 (0.68-0.98)      0.80 (0.34-1.01)
TB history    0.78 (0.69-0.88)   0.80 (0.71-0.91)      0.80 (0.71-0.90)
Stage III     2.32 (1.85-2.90)   2.19 (1.75-2.74)      2.15 (1.72-2.69)
Stage IV      4.79 (3.81-6.01)   4.26 (3.39-5.35)      3.99 (3.17-5.02)
```

Finer control using spline variables made very little difference in the RRs for male sex, TB treatment, and TB history, but it decreased the RRs for HIV stages 3 and 4 by 7% and 17%, respectively. This suggests that the purely categorical analysis contained residual confounding by CD4 and BMI in estimating the effects of WHO stages III and IV.

5 Frequently Asked Questions

5.1 Q: %LGTPHCURV9 chose spline variables for the model with the original continuous variable, but it did not choose spline variables for the model with the missing indicator.

A: This could happen if you have a lot of missing values, because of the "bunching up" at tt mean_conf. Use the model with the spline variables chosen in the original model, because that is a better reflection of the relationship between tt conf and tt case.

5.2 Q: %LGTPHCURV9 chose 2 spline variables with *SLS*=.05=*SLE*, but the test for nonlinearity (LINE 1) was not significant at the .05 level

A: This is possible, but the p-value should still be relatively small. Use the chosen spline variables.

5.3 Q: The p-values in LINES 1 and 2 are small, but none of the spline variables is significant

A: With highly collinear variables like the spline variables, this can happen. Use the chosen spline variables.

6 Credits

Written by Ellen Hertzmark and Donna Spiegelman for the Channing Laboratory. The spline-making procedure is based on a macro written by Frank Harrell. Questions can be directed to Ellen Hertzmark, `stleh@channing.harvard.edu`, (617) 432-4597.

7 References

Harrell, Frank E, Jr., Lee, Kerry L., Pollock, Barbara G.: Regression models in clinical studies: determining relationships between predictors and response. JNCI 80: 1198-1202, 1988.

Govindarajulu, U.S., Malloy, E.J., Ganguli, B., Spiegelman, D., Eisen, E.A.: The comparison of alternative smoothing methods for fitting non-linear exposure-response relationships with Cox models in a simulation study. The International Journal of Biostatistics 5(1): article 2, 2009.

8 See Also

Other relevant Channing macros (available at `/usr/local/channing/sasautos` and with documentation available on the Channing intranet website) are %MAKESPL, %INT2WAY, %LGTPHCURV9.