Published in final edited form as:

Epidemiology. 2011 November ; 22(6): 874–875. doi:10.1097/EDE.0b013e31823029dd.

Splines for trend analysis and continuous confounder control

Chanelle J. Howe^a, Stephen R. Cole^a, Daniel J. Westreich^b, Sander Greenland^c, Sonia Napravnik^{a,d}, and Joseph J. Eron Jr.^d

^aDepartment of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC

^bDepartment of Obstetrics and Gynecology and Duke Global Health Institute, Duke University, Durham, NC

^cDepartment of Epidemiology and Department of Statistics, University of California, Los Angeles, Los Angeles, CA

^dDivision of Infectious Diseases, Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC

Spline regression often represents a less biased and more efficient alternative to standard linear, curvilinear, or categorical analyses of continuous exposures and confounders. Benefits of restricted cubic and quadratic splines have been described in the epidemiologic and biomedical literature.¹⁻² Analogous to the SAS (SAS Institute, Inc., Cary, North Carolina) code provided by Harrell³ for estimating restricted cubic splines, we present straightforward SAS code for estimating restricted quadratic splines. Using data from the HIV clinical cohort at the University of North Carolina Center for AIDS Research,⁴ we illustrate use of restricted quadratic splines in regression modeling for trend analysis and control of a continuous confounder. Details regarding the functional form of the restricted quadratic splines as well as SAS code for estimating restricted quadratic spline functions are provided in the eAppendix (<http://links.lww.com>). The data and SAS code used to generate the results included in this paper are also in the eAppendix.

First, we illustrate the use of a restricted quadratic splines when estimating the association between $\log_{10}$ HIV-1 viral load centered at 2.301 $\log_{10}$ copies/ml and mortality. Figures S1 A-1C show the unadjusted association between centered $\log_{10}$ HIV-1 viral load at therapy initiation and the relative hazard of death estimated from several Cox proportional hazards models that (A) assume a log-linear relationship, (B) use indicators corresponding to quartiles of centered $\log_{10}$ HIV-1 viral load, or (C) include restricted quadratic splines with 4 equal knots based on the case distribution.

Based on the Akaike information criterion (AIC),⁵ presented in Figure 1, the restricted quadratic splines model provides the best fit to the data. The *P*-value for a joint Wald test of the three restricted quadratic splines basis functions included in the model was 0.010. The restricted quadratic splines model suggests a non-log-linear relationship between centered $\log_{10}$ HIV-1 viral load at therapy initiation and the relative hazard of death.

Second, we illustrate the use of restricted quadratic splines when controlling for centered $\log_{10}$ HIV-1 viral load as a confounder using a Cox model. The table shows the hazard ratios for the association between an indicator of CD4 cell count ≤ 350 cells/mm³ at therapy

Corresponding author: Chanelle J. Howe, Department of Epidemiology, McGavran-Greenberg Hall, Campus Box 7435, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27599-7435 (Phone: 919-966-7438, Fax: 919-966-2089, cjhowe@email.unc.edu).

initiation and hazard of death, both unadjusted and adjusted for confounding by viral load at therapy initiation. Adjusting for viral load using a log-linear term attenuated the point estimate corresponding to the CD4 cell count indicator by 26%. Adjustment using restricted quadratic splines with 4 equal knots based on the case distribution attenuated the point estimate by 30%. Attenuation upon control for viral load is expected given that higher viral load was associated with lower CD4 cell count (<http://links.lww.com>), and an elevated risk of subsequent mortality. Similar results were observed when restricted cubic splines was used instead of a restricted quadratic splines with the same degrees of freedom and comparable knot locations (<http://links.lww.com>).

For the first example, use of restricted quadratic splines rather than linear terms or indicators provided a better fit, revealing non-linear relationships that otherwise may have not been apparent. In the second example, use of a restricted quadratic spline resulted in stronger attenuation of a crude association, which likely represents better control of confounding by viral load.

The macro presented here offers users a straightforward SAS option for implementing restricted quadratic splines regression. This code is intended to aid in model selection as well as assessing robustness of inferences when comparing various modeling strategies.^{3,6-7} Furthermore, we hope the examples and code will facilitate the use of splines among researchers hesitant to employ less intuitive but largely equivalent modeling strategies,^{3,7} and in turn broaden the use of splines in applied epidemiologic research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Elizabeth Yanik for assistance with the UNC CFAR data, Petra Sander for reviewing the macro and example code, as well as participants, clinicians, investigators, and staff involved with the UNC CFAR HIV clinical cohort.

Funding Source: This research was supported by National Institutes of Health grant P30 AI50410.

References

1. Harrell FE Jr, Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst.* 1988; 80(15):1198–202. [PubMed: 3047407]
2. Greenland S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology.* 1995; 6(4):356–65. [PubMed: 7548341]
3. Harrell, FE, Jr. [April 12, 2011] DASPLINE Macro. <http://biostat.mc.vanderbilt.edu/twiki/pub/Main/SasMacros/survrisk.txt>
4. Howe CJ, Cole SR, Napravnik S, Eron JJ. Enrollment, retention, and visit attendance in the University of North Carolina Center for AIDS Research HIV clinical cohort, 2001-2007. *AIDS Research and Human Retroviruses.* 2010; 26(8):875–881. [PubMed: 20672995]
5. Akaike H. New Look at Statistical-Model Identification. *Ieee Transactions on Automatic Control.* 1974; Ac19(6):716–723.
6. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol.* 1999; 28(5):964–74. [PubMed: 10597998]
7. Desquilbet L, Mariotti F. Dose-response analyses using restricted cubic spline functions in public health research. *Statistics in Medicine.* 2010; 29(9):1037–1057. [PubMed: 20087875]

Figure 1. Unadjusted associations between centered log₁₀ HIV-1 viral load at therapy initiation and relative hazard of death among 557 male participants in the University of North Carolina Center for AIDS Research HIV clinical cohort, 1999-2010. HIV-1 viral load included as (A) log-linear, (B) indicator, and (C) restricted quadratic spline terms in Cox proportional hazards model.

Table

Hazard ratio for association between CD4 cell count less than or equal to 350 cells/mm³ versus greater than 350 cells/mm³ at therapy initiation and death among 557 male participants in the University of North Carolina Center for AIDS Research HIV clinical cohort, 1999-2010.

Model	Hazard Ratio (95% Confidence Interval)
Unadjusted	3.36 (1.22-9.28)
Adjusted for centered log ₁₀ HIV-1 viral load:	
Binary ^a	3.13 (1.13-8.68)
9 indicator terms ^b	2.56 (0.89-7.41)
Log-linear	2.48 (0.87-7.10)
Restricted quadratic spline ^c	2.34 (0.82-6.72)

^a Indicator for centered HIV-1 viral load >1 log₁₀ copies/ml.

^b Categories for 9-indicator centered log₁₀ HIV-1 viral load: >1.48 to 1.99, >1.99 to 2.27, >2.27 to 2.47, >2.47 to 2.61, >2.61 to 2.78, >2.78 to 3.04, >3.04 to 3.35, >3.35 to 3.56, and >3.56 to 3.57 log₁₀ copies/ml. 9 indicators selected to maximize fit of adjustment using indicators.

^c 4 equal knots based on the case distribution at 2.38, 2.86, 3.04, and 3.53 centered log₁₀ copies/ml.