Semiparametric Adjusted Exposure-Response Curves

Ashley I. Naimi,[a] *Erica E.M. Moodie,*[b] *Nathalie Auger,*[c] *and Jay S. Kaufman*[b]

Abstract: Exposure-response curves are central to understanding how continuous exposures relate to health outcomes. Common methods to plot such curves include the use of crude and adjusted regression, with the exposure specified using fractional polynomials or regression splines. These approaches are subject to important limitations. In this article, we illustrate the use of semiparametric marginally adjusted exposure-response curves obtained via inverse probability weighting. We explore the relation between interpregnancy interval and preterm birth in a cohort of over 720,000 live births in Quebec between 1989 and 2008. We include online supplementary material showing how mixed modeling routines in standard software packages can be used to implement the procedure, and how pointwise bootstrap confidence intervals can be obtained.

(*Epidemiology* 2014;XX: 00–00)

Exposure-response curves, central to understanding how continuous exposures relate to health outcomes, are often fit in 1 of 2 ways. A crude model relating the outcome to a function of the exposure may be used. Alternatively, a covariate-adjusted model may be used to obtain predicted outcome values for unique confounder values. For example, in a regression model adjusted for continuous age and an indicator of male sex, one can center the age variable at its mean and set the indicator for sex to zero to obtain an exposure-response function for average-aged females. In both approaches, the exposure is often specified using fractional polynomials[1,2] or regression splines.[3,4(p. 18–24)]

Although almost universal, these approaches are subject to important limitations. First, a poorly specified functional exposure form can lead to bias, although the severity of this problem can be lessened by judicious selection of the polynomial or spline basis functions,[1,2,5,6] or the implementation of a semi- or nonparametric approach.[5–8] Second, while not adjusting for confounders may lead to bias, researchers and policy makers may wish to interpret the exposure-response function for the entire population rather than a specific subgroup defined by unique confounder combinations.

We present a simple approach to obtain semiparametric adjusted exposure-response curves for time-fixed exposures. We show how the generalized propensity score[9] can be used to obtain exposure-response curves standardized to the total population,[10] and we use a semiparametric approach to mitigate the problem of specifying an unknown functional form for the exposure-outcome relation. We estimate the exposure-response function for the relation between interpregnancy interval and the risk of preterm birth in a cohort of over 700,000 live births in Quebec, Canada, and provide supporting details in the accompanying eAppendix (http://links.lww.com/EDE/A816).

INTERPREGNANCY INTERVAL AND THE RISK OF PRETERM BIRTH

The interpregnancy interval, or the elapsed duration between a woman's previous pregnancy and the conception date of the next pregnancy, is related to a host of adverse perinatal outcomes, including preterm birth.[11] As a potentially modifiable risk factor, the optimal timing of conception after a previous delivery is important.[11,12] In the next section, we let Y be a binary indicator of preterm birth, X refer to the number of years between the index pregnancy and the previous pregnancy, and $C = \{C_1, C_2\}$ represent 2 binary confounders for illustrative purposes.

Standard Approach: Population-specific Curves

The simplest approach to characterize exposure-response functions is to estimate:

$$\log\left\{\frac{P(Y = 1 \mid X = x)}{1 - P(Y = 1 \mid X = x)}\right\} = \alpha_0 + s(x)$$

where $s(\bullet)$ is a user-specified function. To account for confounding, one could estimate:

$$\log\left\{\frac{P(Y=1\mid X=x,\mathbf{C}=\mathbf{c})}{1-P(Y=1\mid X=x,\mathbf{C}=\mathbf{c})}\right\} = \beta_0 + s(x) + \beta_1 C_1 + \beta_2 C_2 + \beta_3 C_1 \times C_2,$$

where $\mathbf{C}$ is a vector of confounders. The parameter estimates from this model can be used to plot the risk of preterm birth against interpregnancy interval to obtain:

$$
\begin{aligned}
P(Y=1\mid X=x,\mathbf{C}=\mathbf{c}) = \text{expit}\{\beta_0 + s(x)\} & \quad \text{for } C_1 = 0, C_2 = 0, \\
\text{expit}\{\beta_0 + s(x) + \beta_1\} & \quad \text{for } C_1 = 1, C_2 = 0, \\
\text{expit}\{\beta_0 + s(x) + \beta_2\} & \quad \text{for } C_1 = 0, C_2 = 1, \\
\text{expit}\{\beta_0 + s(x) + \beta_1 + \beta_2 + \beta_3\} & \quad \text{for } C_1 = 1, C_2 = 1,
\end{aligned}
$$

where $\text{expit}\{\bullet\} = 1/\{1 + \exp(-\bullet)\}$. To plot the adjusted risk of Y as a function of X, one could choose the confounder combination that yields the lowest risk category or simply select an arbitrary subgroup (eg, with confounder values $C_1 = 0$ and $C_2 = 1$). Both cases will yield the same curve (on the log-odds scale) shifted along the y-axis.

Weighted Approach: Population-standardized Curves

An alternative approach is to plot a marginally adjusted curve by standardizing it to a confounder distribution in the target population. If the target is the total population, one can compute a standardized curve as:

$$P(Y=1\mid X=x) = \sum_c P(Y=1\mid X=x,\mathbf{C}=\mathbf{c})P(\mathbf{C}=\mathbf{c}) \quad (1)$$

$$
\begin{aligned}
= \; & \text{expit}\{\beta_0 + s(x)\} \times P(C_1 = 0, C_2 = 0) + \\
& \text{expit}\{\beta_0 + s(x) + \beta_1\} \times P(C_1 = 1, C_2 = 0) + \\
& \text{expit}\{\beta_0 + s(x) + \beta_2\} \times P(C_1 = 0, C_2 = 1) + \\
& \text{expit}\{\beta_0 + s(x) + \beta_1 + \beta_2 + \beta_3\} \times P(C_1 = 1, C_2 = 1),
\end{aligned}
$$

which can be re-written as:

$$
\begin{aligned}
\sum_c P(Y=1\mid X=x,\mathbf{C}=\mathbf{c})P(\mathbf{C}=\mathbf{c}) & = \sum_c \frac{P(Y=1, X=x,\mathbf{C}=\mathbf{c})P(\mathbf{C}=\mathbf{c})}{f(x\mid \mathbf{C}=\mathbf{c})P(\mathbf{C}=\mathbf{c})} \\
& = \sum_c \frac{P(Y=1, X=x,\mathbf{C}=\mathbf{c})}{f(x\mid \mathbf{C}=\mathbf{c})} \quad (2)
\end{aligned}
$$

where $f(x\mid \mathbf{C}=\mathbf{c})$ is the conditional density of the exposure.

For a binary exposure, the denominator in equation (2) reduces to the probability of the observed exposure, which is defined by the propensity score for the exposed, and the complement of the propensity score for the unexposed.[13] When the exposure is continuous, the denominator is the generalized propensity score,[9,14,15] which is defined as the exposure's conditional density function evaluated at the observed exposure value. Using the generalized propensity score, equation (2) leads to an inverse probability–weighted estimator of the standardized exposure-response function defined in equation (1).[16,17]

METHODS

We extracted records on 726,294 live singleton multiparous births between 1989 and 2008 (inclusive) from Quebec birth file data to estimate exposure-response curves for the relation between interpregnancy interval (in years) and the risk of preterm birth (<37 weeks completed gestation). We used the quantile binning method to construct stabilized inverse probability weights.[18]

To obtain marginally adjusted exposure-response functions, we avoided conditioning the numerator of the weights on elements of $\mathbf{C}$.[19] Additional details are provided in eAppendix 1 (http://links.lww.com/EDE/A816).

Semiparametric Smoothing Function

We used penalized smoothing splines[20,21] to plot the exposure-response function for the relation between interpregnancy interval and preterm birth. With typical smoothing approaches (eg, fractional polynomials or restricted quadratic splines), one gains flexibility by increasing the number of polynomial terms or knot points at the expense of stability in the exposure-response function. Penalized smoothing splines use a large number of spline basis functions (yielding an unstable but highly flexible exposure-response function), with parameters estimated using penalized maximum likelihood.[22] Using penalized maximum likelihood smooths the unstable exposure-response function, and thus balances the trade-off between stability and flexibility. The approach can be implemented using standard generalized linear mixed modeling routines, as explained in eAppendix 2 (http://links.lww.com/EDE/A816).

Regression Models

We fit marginally adjusted logistic regression models to plot exposure-response functions for the overall relation between interpregnancy interval and the risk of preterm birth. We also fit marginally adjusted models for the relation between interpregnancy interval and the risk of preterm birth among mothers with 1 previous pregnancy (parity = 1) and 2 or more previous pregnancies (parity = 2+) separately. All models were marginally adjusted by weighting each individual's contribution to the penalized likelihood with the stabilized inverse probability weights defined in eAppendix 1 (http://links.lww.com/EDE/A816).

We used R version 2.15.3 for all analyses. Additional technical details, including a description of our data and models, are provided in eAppendix 1 (http://links.lww.com/EDE/A816). An illustration of the approach using SAS Software (SAS Institute, Cary, NC) is provided in eAppendix 2 (http://links.lww.com/EDE/A816).

RESULTS

The Table shows descriptive statistics for continuous variables included in our models. The numbers of live births from mothers of parity 1, 2, 3, 4, and 5+ were 473,425, 175,696, 51,356, 15,030, and 10,787, respectively. The log-mean of the stabilized inverse probability weights was close

TABLE. Descriptive Characteristics for 726,294 Live Singleton Births for Multiparous Women in the Quebec Birth File, 1989–2008

Characteristic	Median	Interquartile Range
Maternal education	13	11–16
Maternal age	30	27–33
Paternal age	32	29–36
Maternal year of birth	1967	1963–1973
Infant year of birth	1998	1993–2003

to zero across the range of exposure values, with the largest log-weight no greater than 4.
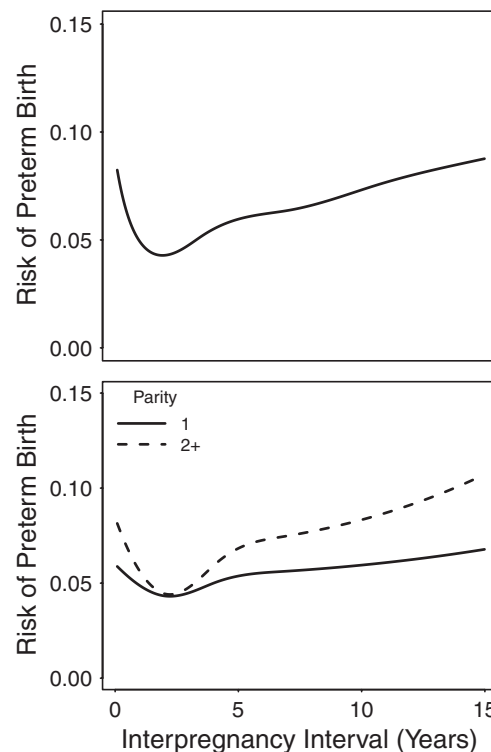
The Figure shows the marginally adjusted exposure-response functions for the total relation between interpregnancy interval and the risk of preterm birth, and the corresponding exposure-response functions among women with 1 previous pregnancy (parity = 1) and 2 or more previous pregnancies (parity = 2+). The shape of the exposure-response functions was similar across models and coincided with expectations based on previous research.[11] For all models, the lowest estimated risk of preterm birth was similar in magnitude (minimum [standard error] = 0.04 [0.005]) corresponding to an interpregnancy interval of 18.3 months.

DISCUSSION

In this article, we illustrate the use of a simple method to obtain semiparametric marginally adjusted exposure-response curves. The approach can easily be implemented using generalized linear mixed model routines (eg, proc glimmix) and yields curves that can be interpreted as standardized to the total population.

Previous reports[23–25] have outlined approaches to obtain marginally adjusted survival curves. Inverse probability weighting by the generalized propensity score has been used to plot survival curves for the relation between epoetin and mortality,[26] epoetin and cardiovascular risk,[27] and hormone replacement therapy on the risk of invasive breast cancer.[28] Although Rothman et al[29(p. 308, 444–8)] suggest standardization as a solution to the problem of confounded exposure-response curves, few applications exist. Furthermore, only one prior study[30] has combined inverse weighting with semiparametric approaches to simultaneously address issues of confounding and misspecification of the exposure function.

The propensity score for binary exposures has been generalized to the continuous exposure setting.[9,14,15] Moodie and Stephens[31] showed how exposure-response curves can be obtained by conditioning on the generalized propensity score in a study of the treatment of childhood amblyopia. Although conditioning on a scalar summary confounder score facilitates the process of plotting exposure-response curves, this approach still requires conditioning on specific generalized propensity score levels and assumes no heterogeneity across propensity score values.

FIGURE. Exposure-response relation between interpregnancy interval and the risk of preterm birth in 726,294 live singleton births in Quebec, 1989–2008. Upper panel, marginally adjusted overall risk. Lower panel, marginally adjusted risk stratified by parity.

Generalized propensity score weighting provides a simple method of obtaining standardized exposure-response curves, while use of a semiparametric approach renders curves more robust to misspecification of the exposure. Moreover, as illustrated in eAppendix 2 (http://links.lww.com/EDE/A816), exploiting the well-established relations between penalized smoothing splines and mixed effects models[32] makes fitting such models relatively straightforward. Pointwise confidence intervals for marginally adjusted curves can be obtained using the bootstrap,[33] as we illustrate in eAppendix 2 (http://links.lww.com/EDE/A816).

Weighting by the inverse of the generalized propensity score may lead to well-known issues with highly variable weights.[34] However, this problem may be improved by using normalized stabilized weights[35] or a quantile binning approach.[18] Although interpregnancy interval was highly skewed in our study, use of the quantile binning method yielded well-behaved inverse probability weights.

We combined generalized propensity score weights with semiparametric regression models to obtain exposure-response curves for the relation between interpregnancy interval and preterm birth. Our results illustrate the feasibility of obtaining marginally adjusted exposure-response curves while minimizing parametric assumptions.

REFERENCES

1. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol*. 1999;28:964–974.
2. Greenland S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology*. 1995;6:356–365.
3. Howe CJ, Cole SR, Westreich DJ, Greenland S, Napravnik S, Eron JJ Jr. Splines for trend analysis and continuous confounder control. *Epidemiology*. 2011;22:874–875.
4. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York, NY: Springer; 2001.
5. Steenland K, Deddens JA. A practical guide to dose-response analyses and risk assessment in occupational epidemiology. *Epidemiology*. 2004;15:63–70.
6. Govindarajulu US, Spiegelman D, Thurston SW, Ganguli B, Eisen EA. Comparing smoothing techniques in Cox models for exposure-response relationships. *Stat Med*. 2007;26:3735–3752.
7. Hastie T, Tibshirani R. Generalized Additive Models. London; New York: Chapman & Hall; 1990.
8. Hardle WK, Muller M, Sperlich S, Werwatz A. Nonparametric and Semiparametric Models. Berlin, Germany: Springer-Verlag; 2004.
9. Imbens G. The role of the propensity score in estimating dose-response functions. Biometrika. 2000;87:706–710.
10. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology*. 2003;14:680–686.
11. Conde-Agudelo A, Rosas-Bermúdez A, Kafury-Goeta AC. Birth spacing and risk of adverse perinatal outcomes: a meta-analysis. *JAMA*. 2006;295:1809–1823.
12. Zhu B. The effect of inter-pregnancy interval on birth outcomes: findings from three recent US studies. *Int J Gynaecol Obstet*. 2005;89:S25–S33.
13. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55.
14. Imai K, van Dyk DA. Causal inference with general treatment regimes: generalizing the propensity score. J Am Stat Assoc. 2004;99:854–866.
15. Hirano K, Imbens G. The propensity score with continuous treatments. In: Gelman A, Meng XL, eds. *Applied Bayesian Modeling and Causal Inference from Incomplete-data Perspectives*. Oxford: Wiley; 2004:73–84.
16. Robins JM. Marginal structural models. In: Proceedings of the Section on Bayesian Inference. Alexandria, VA: American Statistical Association;1998:1–10.
17. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60:578–586.
18. Naimi AI, Moodie EE, Auger N, Kaufman JS. Constructing inverse probability weights for continuous exposures: a comparison of methods. *Epidemiology*. 2014;25:292–299.
19. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168:656–664.
20. Green P, Silverman B. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. London: Chapman & Hall; 1994.
21. Wood SN. Generalized Additive Models: An Introduction with R. Boca Raton, FL: Chapman & Hall/CRC; 2006.
22. Cole SR, Chu H, Greenland S. Maximum likelihood, profile likelihood, and penalized likelihood: a primer. *Am J Epidemiol*. 2014;179:252–260.
23. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed*. 2004;75:45–49.
24. Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med*. 2005;24:3089–3110.
25. Nieto FJ, Coresh J. Adjusting survival curves for confounders: a review and a new method. *Am J Epidemiol*. 1996;143:1059–1068.
26. Zhang Y, Thamer M, Cotter D, Kaufman J, Hernán MA. Estimated effect of epoetin dosage on survival among elderly hemodialysis patients in the United States. *Clin J Am Soc Nephrol*. 2009;4:638–644.
27. Zhang Y, Thamer M, Kaufman JS, Cotter DJ, Hernán MA. High doses of epoetin do not lower mortality and cardiovascular risk among elderly hemodialysis patients with diabetes. *Kidney Int*. 2011;80:663–669.
28. Toh S, Hernández-Díaz S, Logan R, Robins JM, Hernán MA. Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomization. *Epidemiology*. 2010;21:528–539.
29. Rothman KJ, Greenland S, Lash T. Modern Epidemiology. 3rd ed. Philadelphia, PA: Wolters Kluwer; 2008.
30. Xiao Y, Abrahamowicz M, Moodie EEM, Weber R, Young J. Flexible marginal structural models for estimating the cumulative effect of a time-dependent treatment on the hazard: reassessing the cardiovascular risks of didanosine treatment in the Swiss HIV cohort study. *J Am Stat Assoc*. 2014;109:455–464.
31. Moodie EE, Stephens DA. Estimation of dose-response functions for longitudinal data using the generalised propensity score. *Stat Methods Med Res*. 2012;21:149–166.
32. Ruppert D, Wand M, Carroll R. Semiparametric Regression. New York, NY: Cambridge University Press; 2003.
33. Kauermann G, Claeskens G, Opsomer JD. Bootstrapping for penalized spline regression. J Comput Graph Stat. 2009;18:126–146.
34. Robins J, Sued M, Lei-Gomez Q, Rotnitzky A. Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable. Stat Sci. 2007;22:544–559.
35. Xiao Y, Abrahamowicz M, Moodie EE. Accuracy of conventional and marginal structural Cox model estimators: a simulation study. *Int J Biostat*. 2010;6:Article 13.