Estimating the Prevalence of Prophylaxis Use among U.S. Hemophilia Patients

Q. Cathy. Zhang[1], Megan Ullman[2], Althea Grant [1]
[1] Centers for Disease Control and Prevention,
1600 Clifton Rd NE, MS E-88, Atlanta, GA 30333
[2]Gulf States Hemophilia & Thrombophilia Center
6655 Travis St., Suite 400, Houston, TX 77030

Abstract

Prevalence (PR) trend estimation is essential for public health surveillance in order to understand disease burden and uptake of interventions. This paper demonstrates three approaches for estimating PR trends: (1) LOESS smoothing (2) restricted cubic splines and (3) quartic polynomial logistic regression models.

Using a national dataset of approximately 8,000 hemophilia males aged 2-79 seen at 135 federally-funded US hemophilia treatment centers, we estimated the prevalence of prophylaxis use in order to demonstrate the unique features of each method of PR estimation. (Prophylaxis is the regularly scheduled administration of clotting factor concentrate to prevent bleeding into muscles, joints and tissue.)

We will illustrate similarities and differences of the three methods, which can be applied to estimate indicators for other public health conditions.

Key Words: LOESS smoothing, restricted cubic splines, logistic regression

1. Introduction

Prevalence rate (PR) measures are widely applied to assess disease/condition risk and occurrence in epidemiological studies [1-4]. PR is defined as the proportion of individuals in a population who have a specific characteristic in a specified time interval. The PR indicates the distribution and stability of a condition and provides information on the burden of disease.

When PR is not aligned linearly along time, a good approximation of the underlying PR function is needed. These functions can often be complex, with multiple turning points, for example. We demonstrate and evaluate here three methods of capturing, constructing and accurately visualizing the PR trend.

2. Data and Methods

2.1 Data resource

Data is from the Universal Data Collection (UDC), a national surveillance project conducted by the Centers for Disease Control and Prevention (CDC). It includes 9,725 males with moderate or severe hemophilia A or B aged 2-79 seen at 135 US hemophilia treatment centers between 1998-2011(Table 1). Prevalence of prophylaxis use (PU) was calculated. Demographic and clinical characteristics include age, ethnicity, and health insurance status and type; body mass index (BMI); geographic region; hemophilia type (A or B); disease severity and self-infusion status.

2.2 Analysis Methods

Three methods (LOESS smoothing, restricted cubic splines and quartic polynomial logistic regression models) were used to calculate the PR. Figures based on these models were developed to illustrate the mean predicted probability of using prophylaxis with advancing age among a subset of the sample population. All statistical tests were based on two-sided tests with a significance level of 0.05. SAS 9.2 was used for all calculations

LOESS smoothing (locally weighted scatterplot smoothing) is a nonparametric method for estimating local regression. It fits a regression surface to the data points within a chosen neighborhood of the point. This technique combines the simplicity of linear least squares regression with the flexibility of non-linear regression (Fig. 1).

Spline modeling creates piecewise polynomials that can be used to estimate relationships that are difficult to fit with a single function. In interpolating problems, spline interpolation is often preferred to polynomial interpolation, as it yields similar results to interpolating with higher degree polynomials while avoiding instability. In computer graphics, parametric curves whose coordinates are given by splines are popular because of their simplicity of construction, ease and accuracy of evaluation and capacity to approximate complex shapes through curve fitting and interactive design (Fig. 2).

Quartic polynomial logistic regression models fit a nonlinear relationship between the value of the independent variable x and the corresponding conditional mean of the dependent variable y. These models are used to describe nonlinear phenomena, such as the progression of disease epidemics. They can adjust for multiple covariates, such as demographic and clinical characteristics, clustered data structure and even for sampling design (Fig. 3).

3. Results:

Table 1 demonstrates that prophylaxis rates vary among all seven specified variables.

Fig. 1 Calculation of prophylaxis rates with LOESS smoothing of age demonstrates that prophylaxis rates are non-linear, monotone and decline with age.

Fig. 2 General trends in the smoothing component plots for age suggest possible quadratic dependence on this variable; the trend in age resembles a quartic polynomial. This quartic cubic polynomial can then be confirmed using the third method, quartic polynomial logistic regression.

Fig. 3 The quartic polynomial logistic regression model illustrates differing mean predicted probabilities of using prophylaxis with advancing age among disease severities and insurance types after controlling for other confounders and covariates

Table 1 Association of demographic and clinical characteristics with use of prophylaxis (N=9725)

Patient Characteristic	Sample No.(%)	Prophylaxis rate %
Age		
All youth (2-19 yrs)	5573(57)	57
2-5 years	1009 (10)	54
6-11 years	1890 (19)	64
12-19 years	2674 (27)	53
All adults (20+ yrs)	4152 (43)	26
20-39 years	2958 (30)	32
40-59 years	882 (9)	11
60+ years	312 (3)	8
Ethnicity		
White	6328 (65)	41
African-american	1285 (13)	44
Hispanic	1342 (14)	52
Other	770 (8)	49
Health insurance		
Commercial	5051 (52)	45
Publicly funded	4015 (41)	47
Other/Unknown	86 (1)	16
No	573 (6)	13
BMI Status		
Normal weight or less	5133(53)	49
Overweight	2308(24)	36
Obese	2284(23)	40
Non-Exteme Obesity	1302(13)	30
Exteme Obesity	982(10)	52
Hemo Type		
Hemophilia A	7297(75)	50
Hemophilia B	2428(25)	25
Severity		
Moderate	3600(37)	16
Severe	6125(63)	60
Self-infusion		
No	6157(63)	44
Yes	3568(37)	42

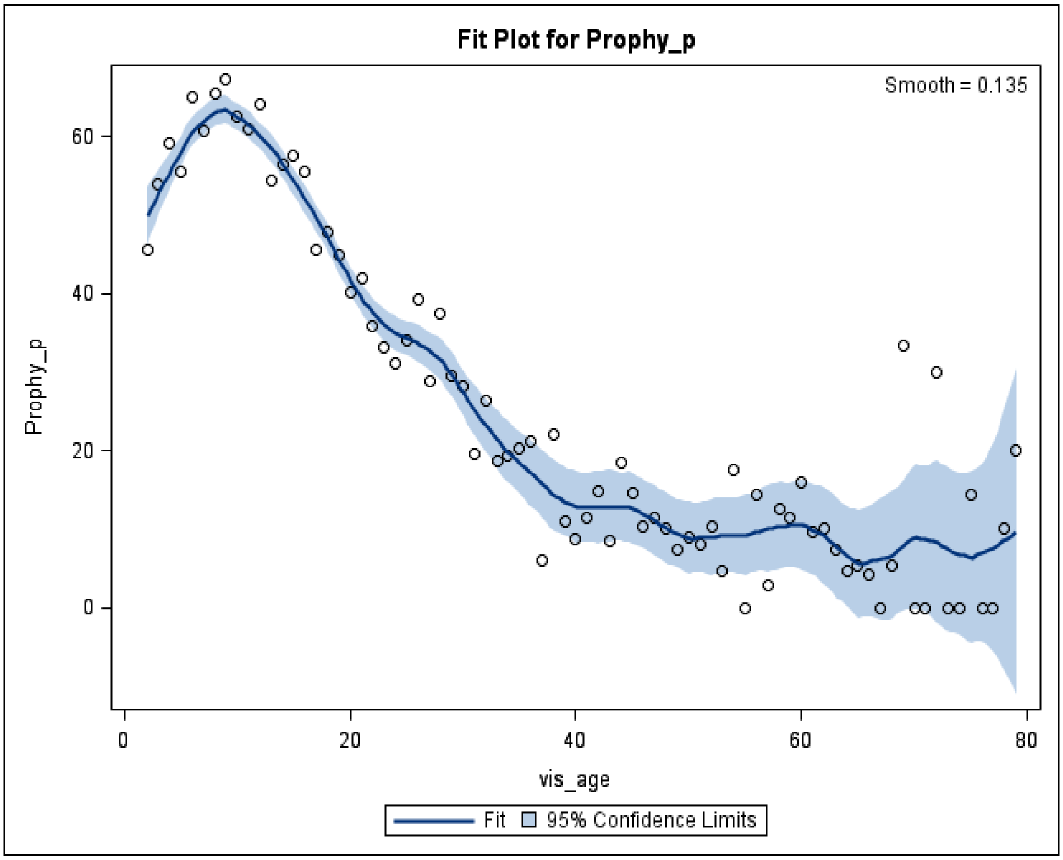

Figure 1: Actual confidence interval coverages, J = 4, ρ = .85, nominal confidence level = .75, quantile = .01

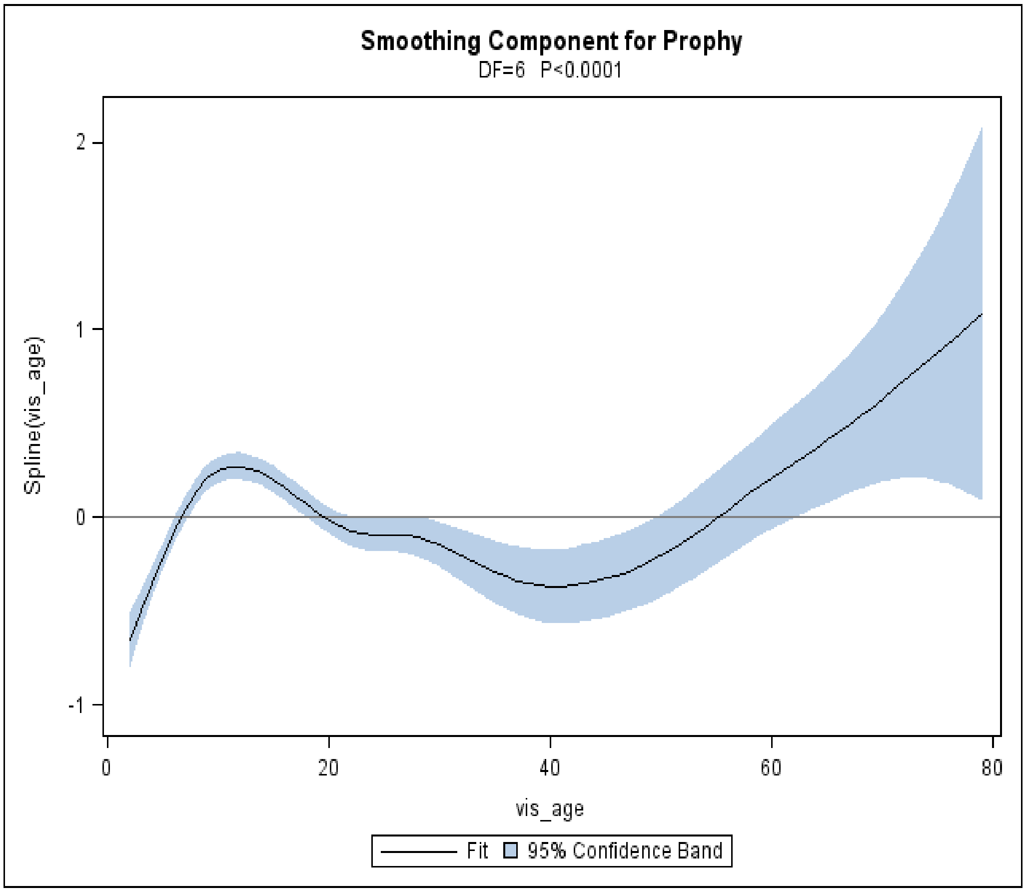

Figure 2 Cubic smoothing spline of PR

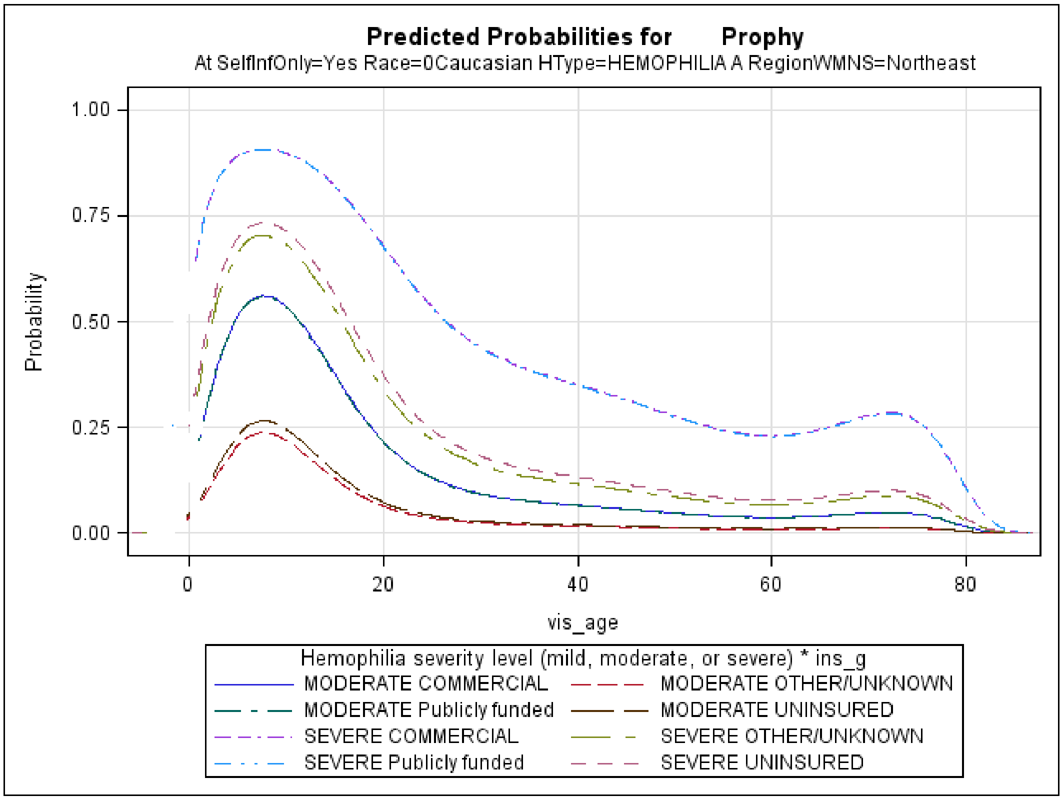

Figure 3 PR from quartic polynomial logistic regression models

4. Discussion and Conclusion

The LOESS procedure allows great flexibility because no assumptions about the parametric form of the regression surface are needed; it presents the raw relationship between two variables and is therefore often the first tool to be used. However, it cannot adjust for other covariates, and when n>5,000, completion of the analysis can be time-consuming.

Splines can adjust for multiple covariates; if assumptions are correct, this is a powerful modeling method, although analysis and interpretation are often complex. In this example, it was used as a tool to detect nonlinear forms.

Logistic regression models are useful when the functional form of the model desired is well understood and is easily interpreted, although assumptions regarding the form of non-linearity are not always correct. When data are not suited to a polynomial regression model, or when the data cannot be represented by a model with a fixed number of parameters, two other procedures can be used – LOESS and splines, e.g. logistic regression with spline.

Ideally, graphic output produced by all three methods will display a similar data pattern. The methods demonstrated here can be applied to estimate indicators for other diseases and conditions of public health importance.

Acknowledgements

The UDC project was funded by a cooperative agreement between the CDC and the Hemophilia Treatment Center Network (HTCN) comprised of 135 federally funded comprehensive care clinics located throughout the US and its territories. The authors also wish to acknowledge the staff of the HTCN for patient recruitment and data collection and the patients for their participation in the study.

References

1. Rothman, KJ, *Epidemiology*: An introduction. New York: Oxford University Press, 2002.
2. Alamanos Y, PV Voulgari, and AA Drosos, Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: a systematic review. *Seminars in arthritis and rheumatism*, 2006. 36(3): 182-8.
3. *HEDIS 2011 Measures*, 2011, National Committee for Quality Assurance.
4. P McCullagh, *Generalized Linear Models*. Second ed. London: Chapman and Hall, 1989.
5. 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]
6. Desquilbeta L, Mariottib F. Dose-response analyses using restricted cubic spline functions in public health research. *Statistics in Medicine*, Jan. 2010 (accessed on-line).