

12-2005

A GRAPHICAL METHOD TO EVALUATE EXPOSURE-RESPONSE RELATIONSHIPS IN EPIDEMIOLOGIC STUDIES USING STANDARDIZED MORTALITY OR MORBIDITY RATIOS

Edwin van Wijngaarden

University of Rochester School of Medicine and Dentistry

Follow this and additional works at: http://scholarworks.umass.edu/dose_response

Recommended Citation

van Wijngaarden, Edwin (2005) "A GRAPHICAL METHOD TO EVALUATE EXPOSURE-RESPONSE RELATIONSHIPS IN EPIDEMIOLOGIC STUDIES USING STANDARDIZED MORTALITY OR MORBIDITY RATIOS," *Dose-Response: An International Journal*: Vol. 3 : Iss. 3 , Article 5.

Available at: http://scholarworks.umass.edu/dose_response/vol3/iss3/5

This Article is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Dose-Response: An International Journal by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Dose-Response, 3: 465–473, 2005
Formerly *Nonlinearity in Biology, Toxicology, and Medicine*
Copyright © 2006 University of Massachusetts
ISSN: 1559-3258
DOI: 10.2203/dose-response.003.04.003



A GRAPHICAL METHOD TO EVALUATE EXPOSURE-RESPONSE RELATIONSHIPS IN EPIDEMIOLOGIC STUDIES USING STANDARDIZED MORTALITY OR MORBIDITY RATIOS

Edwin van Wijngaarden, Ph.D. □ Department of Community and Preventive Medicine, University of Rochester School of Medicine and Dentistry

□ In occupational epidemiology, exposure-response analyses play an important role in the evaluation of the etiologic relevance of chemical and physical exposures. The standardized mortality or morbidity ratio (SMR) has been commonly used in occupational cohort studies. Statistical approaches to evaluate exposure-response patterns using SMRs have mostly been limited to analyses in which the exposure under investigation is categorized. Here, a graphical method for evaluating exposure-response patterns is presented based on SMR estimates across moving exposure windows. This method is demonstrated using the results of two hypothetical cohort studies. The proposed approach may be useful for graphical exploration of exposure-response trends in situations where the number of observed cases is small.

Keywords: occupational epidemiology, cohort studies, standardized mortality ratio, exposure-response analysis, moving exposure windows

INTRODUCTION

In occupational epidemiology, exposure-response analyses play a crucial role in the evaluation of the role of chemical and physical exposures in the etiology of disease. Furthermore, the results of such analyses may provide the basis for quantitative risk assessment in the process of determining regulatory exposure standards. The development of statistical methods for the evaluation of exposure-response patterns has focused on the estimation of the rate ratio, risk ratio or odds ratio based on internal comparison analyses. Examples of such methods include categorical analysis, spline regression, fractional polynomial regression, and the use of linear models [Boucher *et al.* 1998; Greenland 1995; Harrell *et al.* 1988; Witte and Greenland 1997]. The standardized mortality or morbidity ratio (SMR) based on external comparisons has been frequently used in occupational epidemiology, typically in situations where occupational exposure data are not available (i.e., a comparison of mortality in a cohort of workers to mortality in the general population). However, SMRs have also been used to explore exposure-response relationships by stratifying an occupational cohort into subgroups defined by duration of

Address correspondence to Edwin van Wijngaarden, Ph.D., Department of Community and Preventive Medicine, University of Rochester School of Medicine and Dentistry, 601 Elmwood Avenue, Box 644, Rochester, NY 14642. Email: edwin_van_wijngaarden@urmc.rochester.edu. Telephone: (585) 275-1985. Fax: (585) 461-4532.

E. van Wijngaarden

employment or exposure level. These statistical approaches to evaluate exposure-response patterns using SMRs have mostly been limited to categorical analyses. In this paper, a graphical method for evaluating exposure-response patterns is presented based on SMR calculations using moving exposure windows.

METHODS

Data display

The approach described here can be applied to occupational cohort studies in which disease risk is evaluated in relation to quantitative measures of exposure, such as duration of employment or cumulative exposure. The first step is to create exposure categories including one death or incident case. That is, exposure categories are created based on the distribution of exposure among the observed cases N , with cut-off points at each $100/N$ percentile. For example, in a study with 10 deaths this would lead to a cut-off at each tenth percentile of the exposure distribution among the cases, resulting in 10 exposure categories under the assumption that all workers have been exposed. If a proportion of the population was not exposed, an unexposed category (with possibly multiple cases) would be added to these exposure categories. An exposure level such as the mean, median or midpoint corresponding to each exposure category is based on the exposure distribution of the person-time units in the category. Finally, for each exposure group the number of person-years and the expected numbers of cases based on external reference rates are calculated. For each individual, the amount of person-time is calculated as the time elapsed from exposure onset until the individual experiences the disease, is lost to follow-up, or reaches the end of follow-up [Checkoway *et al.* 1989b]. Subsequently, the amount person-time the individual spent in each exposure group can be derived. The total amount of person-time contributed to each exposure category by the entire study population is computed as the sum of individual person-times. Finally, the amount of person-time is multiplied by the external reference rate to yield the expected number of events.

Moving exposure windows analysis

The results for the exposure categories are combined by adding the number of observed exposed cases and expected cases across moving exposure windows. In the event that there is one unexposed case, this would be included in the moving exposure windows analysis; otherwise, the unexposed cases will be excluded. Creating moving exposure windows based on at least five cases may result in a smoother, more stable exposure-response curve. The average exposures corresponding to these SMRs are based on the exposure level (i.e., mean, median or midpoint)

Evaluating nonlinearity using standardized mortality ratios

corresponding to the collapsed exposure categories within the exposure window weighted by the number of person-years in each category.

To demonstrate the moving exposure window calculation, the results of two hypothetical cohort studies are presented in Table 1. These artificial data were chosen after an iterative search to find two distinct exposure-response patterns based on different numbers of expected events. Both studies observed ten cases of a certain disease. The calculation of the moving exposure window curve is arbitrarily based on windows with five observed cases. The first SMR_{1-5} is calculated by combining the observed and expected number of cases for the first five exposure categories, and calculating the corresponding average exposure level. The subsequent exposure window combines exposure categories 2-6, and the accompanying SMR_{2-6} and exposure level are computed as above. Hence, this SMR_{2-6} and the SMR_{1-5} have 4 deaths in common. The SMR_{3-7} can be calculated by combining exposure categories 3-7, and so on. The lower and upper two exposure categories are based on the remaining exposure windows with less than five cases. This approach results in six exposure categories with five observed cases, 2 groups with four observed cases, and 2 groups with 3 observed cases. The results from these calculations for both hypothetical scenarios are presented in Table 2.

Categorical analysis

In order to compare the results of the moving exposure window analysis to the conventional approach of evaluating exposure-response relationships, SMRs are calculated based on a categorization of exposure. Exposure groups are formed based on percentiles (e.g., tertiles or quantiles) of the exposure distribution among cases. For example, in the two

TABLE 1 Disease risk in relation to exposure in two hypothetical scenarios based on epidemiological observation

Cumulative Exposure	Mean exposure	Person-years	Scenario 1		Scenario 2	
			Observed	Expected [†]	Observed	Expected [†]
0 – 0.075	0.05	1,000	1	1.30	1	1.00
0.075 – 0.125	0.10	1,000	1	1.00	1	1.00
0.125 – 0.175	0.15	750	1	0.80	1	1.00
0.175 – 0.225	0.20	500	1	0.70	1	1.50
0.225 – 0.275	0.25	560	1	0.95	1	1.45
0.275 – 0.325	0.30	825	1	0.83	1	0.75
0.325 – 0.375	0.35	1,530	1	0.63	1	0.63
0.375 – 0.425	0.40	1,200	1	0.51	1	0.51
0.425 – 0.475	0.45	670	1	0.73	1	0.73
0.475+	0.50	710	1	0.59	1	0.59
Total	0.28	8,745	10	8.04	10	9.16

[†]Expected number of deaths in bold print differs between the two scenarios

E. van Wijngaarden

TABLE 2 Standard mortality ratio (SMR) in relation to exposure in two hypothetical scenarios based on rolling SMR and categorical analysis

Cumulative Exposure	Mean exposure	Person-years	Scenario 1			Scenario 2		
			Obs	Exp [†]	SMR	Obs	Exp [†]	SMR
ROLLING SMR ANALYSIS								
0 – 0.175	0.095	2,750	3	3.10	0.97	3	3.00	1.00
0 – 0.225	0.112	3,250	4	3.80	1.05	4	4.50	0.89
0 – 0.275	0.132	3,810	5	4.75	1.05	5	5.95	0.84
0.075 – 0.325	0.193	3,635	5	4.28	1.17	5	5.70	0.88
0.125 – 0.375	0.273	4,165	5	3.91	1.28	5	5.33	0.94
0.175 – 0.425	0.326	4,615	5	3.62	1.38	5	4.84	1.03
0.225 – 0.475	0.356	4,785	5	3.65	1.37	5	4.07	1.23
0.275 – 0.475+	0.389	4,935	5	3.29	1.52	5	3.21	1.56
0.325 – 0.475+	0.407	4,110	4	2.46	1.63	4	2.46	1.63
0.375 – 0.475+	0.441	2,580	3	1.83	1.64	3	1.83	1.64
Cumulative Exposure	Mean exposure	Person-years	Scenario 1			Scenario 2		
			Obs	Exp [†]	SMR	Obs	Exp [†]	SMR
CATEGORICAL ANALYSIS								
0 – 0.175	0.095	2,750	3	3.10	0.97	3	3.00	1.00
0.175 – 0.325	0.259	1,885	3	2.48	1.21	3	3.70	0.81
0.325 – 0.475+	0.407	4,110	4	2.46	1.63	4	2.46	1.63

[†]Expected number of deaths in bold print differs between the two scenarios

hypothetical cohort studies with ten exposed cases, exposure groups are formed based on 3 (SMR₁₋₃), 3 (SMR₄₋₆) and 4 (SMR₇₋₁₀) cases, and corresponding SMRs are computed (Table 2). Confidence intervals (95% CI) are calculated under the assumption of a Poisson distribution [Bailar and Ederer 1964].

Poisson regression

A linear line fitted to the categorical SMR results is visually compared with the exposure-response pattern derived from the moving exposure window SMR analysis. For the calculation of the linear slope, we assume that the number of events follows a Poisson distribution. A linear non-threshold multiplicative model is fit to the SMR results from the categorical analysis (three categories; see above) using iteratively re-weighted least-squares estimation [Hanley and Liddell 1985; Hertz-Picciotto and Smith 1993]:

$$E[obs_i] = \alpha * EXP_i * (1 + \beta * x_i)$$

Evaluating nonlinearity using standardized mortality ratios

where $E[\cdot]$ indicates the expectation of a random variable (in this case from a Poisson distribution), obs_i is the observed number of events at exposure level i , α represents any difference between the study cohort and the external referent population with respect to baseline rate, EXP_i is the expected number of events at exposure level i , β is the slope based on maximum likelihood estimation, and x is the exposure level. In addition, a goodness of fit statistic is calculated which follows a chi-square distribution with $(k - N)$ degrees of freedom, where k is the number of exposure categories and N is the number of parameters estimated.

In addition to the linear relative risk model, for each scenario we fit a fractional polynomial model [Royston *et al.* 1999] to the observed and expected events presented in Table 1 using Poisson regression:

$$\log[obs_i] = \alpha + \beta_1 * x_i^a + \beta_2 * x_i^b + 1 * \log[EXP_i]$$

where obs_i is the observed number of events at exposure level i , α represents the baseline rate, EXP_i is the expected number of events at exposure level i , β_1 and β_2 are the slope estimates based on maximum likelihood estimation, and x is the exposure level. The model which best fit the data (i.e., lowest deviance) is selected as the final model after evaluation of 36 combinations of the powers a and b , which were chosen from among -2, -1, -0.5, 0, 0.5, 1, 2, and 3 [Royston *et al.* 1999].

RESULTS

Table 1 presents the crude data for the two hypothetical cohort studies. For five exposure categories, a different number of expected cases were assumed whereas the data are equivalent otherwise. The total SMRs for scenario 1 and 2 are 1.24 (95% CI = 0.60-2.29) and 1.09 (95% CI = 0.52-2.01), respectively. The results of the moving exposure windows and categorical analysis are presented in Table 2, and displayed in Figures 1 and 2.

Using the three exposure categories, a linear slope estimated by iteratively re-weighted least-squares estimation provides an adequate fit to both exposure-response situations. The fit of the linear model was somewhat better for the first scenario (goodness of fit p-value = 0.91) relative to the second scenario (goodness of fit p-value = 0.47). A monotonic exposure-response pattern was seen based on visual inspection, the moving exposure window analysis, and fractional polynomial regression for scenario 1. However, the assumption of linearity appeared inappropriate for the second scenario with little indication for an exposure-response association except at relatively high levels of exposure. The moving exposure window analysis showed an abrupt increase in risk between the cumulative exposure values 0.3 and 0.4, and the fractional polynomial model showed a J-shaped exposure-response pattern. The categorical SMR was slightly elevated only in the highest exposure category.

E. van Wijngaarden

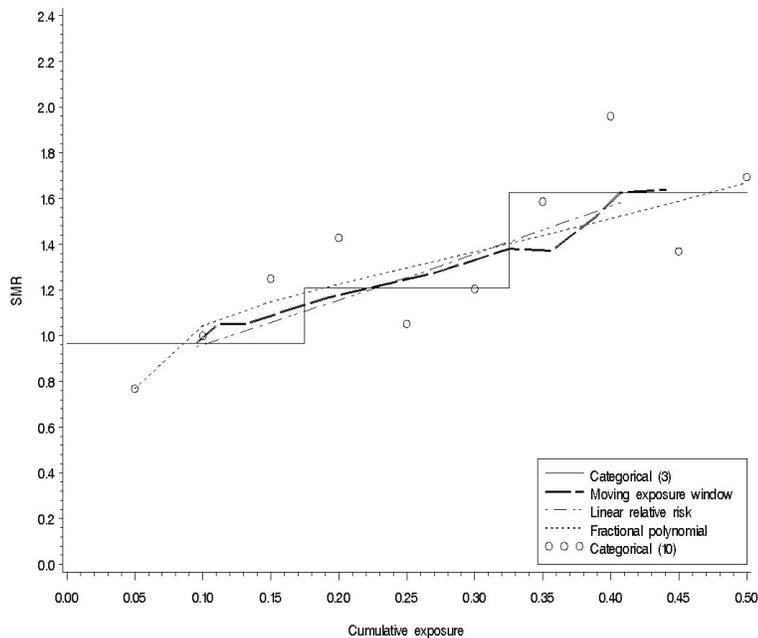


FIGURE 1 Standard mortality ratio (SMR) in relation to exposure based on moving exposure window and regression analysis: Scenario 1. Goodness of fit chi-square for linear relative risk model = 0.01 (p-value = 0.91).

DISCUSSION

It is well recognized that exposure groups selected for categorical analysis are often arbitrary and can lead to misleading results [Greenland 1995; Schulz *et al.* 2001]. Alternative methods of exposure-response analysis have been developed based on internal comparisons to improve upon these limitations [Boucher *et al.* 1998; Greenland 1995; Harrell *et al.* 1988; Witte and Greenland 1997]. Nevertheless, results of internal comparison analyses can be imprecise when the study population is small with a limited number of exposed cases [Marsh *et al.* 2001]. Exposure-response analysis incorporating an external reference population may improve the precision of the risk estimates [Rice *et al.* 2001], and may account for geographic variation in cultural or socioeconomic factors [Doll 1985].

The calculation of SMRs using moving exposure windows may be useful for exploring exposure-response relationships when the number of observed cases is small, as well as for selecting appropriate cut-points for defining exposure categories for categorical analyses of exposure-response patterns. Furthermore, the technique easily accommodates the evaluation of the exposure-response relationship using excess rate models [Jarvholm 1997]. It can also be extended to internal comparison analyses by first estimating SMRs (or corresponding excess rates) using

Evaluating nonlinearity using standardized mortality ratios

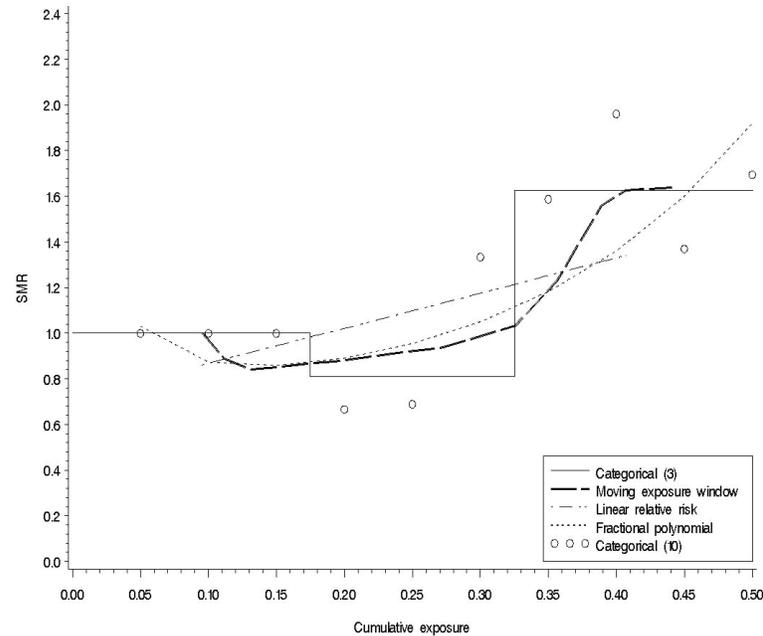


FIGURE 2 Standard mortality ratio (SMR) in relation to exposure based on moving exposure window and regression analysis: Scenario 2. Goodness of fit chi-square for linear relative risk model = 0.62 (p -value = 0.47).

the disease rate of all risk groups combined to compute the expected number of cases and subsequently computing the ratio of SMRs using a specified referent group (e.g., least exposed) [Frome and Checkoway 1985], which is appropriate when certain criteria are satisfied [Armstrong 1995].

It is recognized that the more advanced methods of exposure-response analysis developed for internal comparisons, such as cubic spline regression, could also be applied to SMR data [Rice *et al.* 2001]. However, it is unlikely that these models would substantially improve the fit to the data as compared to an intercept model (e.g., $SMR = \alpha$), or conventional linear (e.g., $SMR = \alpha * [1 + \beta * \text{exposure}]$) or log-linear (e.g., $SMR = e^{\alpha + \beta * \text{exposure}}$) models if the number of observed cases is small since the number of parameters in the regression model may approach or exceed the number of cases of disease. Nonetheless, it would be informative to compare the exposure-response curve derived using moving exposure windows with those obtained from flexible regression models when the number of observed cases is sufficiently large. In the hypothetical examples presented here, the exposure-response patterns obtained from moving exposure window analysis and fractional polynomial models were quite similar.

E. van Wijngaarden

Several limitations of the approach outlined here need to be acknowledged. First, specification of the width of the exposure window is required, and different widths may result in different shapes of the exposure-response curve. That is, the curve will become smoother with increasing width of the moving exposure windows. In addition, no parameters are estimated and since SMRs in adjacent exposure windows share observed cases, measures of variability are not easily computed. Therefore, the rolling SMR approach is most suitable for a graphical inspection of the exposure-response curve, and should be considered a preliminary step to guide more detailed analyses. Finally, it is known that comparisons of SMRs (or corresponding excess rates) between exposure categories are invalid if their confounder distributions differ [Checkoway *et al.* 1989a; Checkoway *et al.* 1989b], although the amount of bias generally tends to be small [Breslow *et al.* 1983].

In conclusion, a straightforward approach to graphically explore non-linearity in sparse epidemiological data is proposed. Further evaluation of this method is needed using empirical data. Meanwhile, it is recommended that researchers follow the structure of Table 1 to display SMR estimates based on a small number of cases, which would allow the reader to explore the exposure-response curve under various specifications of the width of moving exposure windows. Furthermore, it may be helpful to graphically display the original data (from Table 1) and relationships fitted with categorical, linear, fractional polynomial or moving exposure window models (as was done in Figures 1 and 2) to evaluate the fit of such models to the data.

ACKNOWLEDGMENTS

I am grateful to Drs. David Richardson and Dana Loomis for their thoughtful review of an earlier draft of this manuscript.

REFERENCES

- Armstrong BG. 1995. Comparing standardized mortality ratios. *Ann Epidemiol* 5(1):60-4.
- Bailar JC, 3rd, Ederer F. 1964. Significance factors for the ratio of a Poisson variable to its expectation. *Biometrics* 20:639-643.
- Boucher KM, Slattery ML, Berry TD, Quesenberry C, Anderson K. 1998. Statistical methods in epidemiology: a comparison of statistical methods to analyze dose-response and trend analysis in epidemiologic studies. *J Clin Epidemiol* 51(12):1223-33.
- Breslow NE, Lubin JH, Marek P, Langholz B. 1983. Multiplicative models and cohort analysis. *J Am Stat Assoc* 78(381):1-12.
- Checkoway H, Pearce N, Dement JM. 1989a. Design and conduct of occupational epidemiology studies: I. Design aspects of cohort studies. *Am J Ind Med* 15(4):363-73.
- Checkoway H, Pearce N, Dement JM. 1989b. Design and conduct of occupational epidemiology studies: II. Analysis of cohort data. *Am J Ind Med* 15(4):375-94.
- Doll R. 1985. Occupational cancer: a hazard for epidemiologists. *Int J Epidemiol* 14(1):22-31.
- Frome EL, Checkoway H. 1985. Epidemiologic programs for computers and calculators. Use of Poisson regression models in estimating incidence rates and ratios. *Am J Epidemiol* 121(2):309-23.

Evaluating nonlinearity using standardized mortality ratios

- Greenland S. 1995. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology* 6(4):356-65.
- Hanley J, Liddell D. 1985. Fitting relationships between exposure and standardized mortality ratios. *J Occup Med* 27(8):555-60.
- Harrell FE, Jr., Lee KL, Pollock BG. 1988. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst* 80(15):1198-202.
- Hertz-Picciotto I, Smith AH. 1993. Observations on the dose-response curve for arsenic exposure and lung cancer. *Scand J Work Environ Health* 19(4):217-26.
- Jarvholm B. 1997. Relative risk and excess rate models in exposure-response analysis. *Am J Ind Med* 31(4):399-402.
- Marsh GM, Youk AO, Collins JJ. 2001. Reevaluation of lung cancer risk in the acrylonitrile cohort study of the National Cancer Institute and the National Institute for Occupational Safety and Health. *Scand J Work Environ Health* 27(1):5-13.
- Rice FL, Park R, Stayner L, Smith R, Gilbert S, Checkoway H. 2001. Crystalline silica exposure and lung cancer mortality in diatomaceous earth industry workers: a quantitative risk assessment. *Occup Environ Med* 58(1):38-45.
- Royston P, Ambler G, Sauerbrei W. 1999. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 28(5):964-74.
- Schulz MR, Hertz-Picciotto I, van Wijngaarden E, Hernandez JC, Ball LM. 2001. Dose-response relation between acrylamide and pancreatic cancer. *Occup Environ Med* 58(9):609.
- Witte JS, Greenland S. 1997. A nested approach to evaluating dose-response and trend. *Ann Epidemiol* 7(3):188-93.