Selecting an Exposure Lag Period

Alberto Salvan, Leslie Stayner, Kyle Steenland, and Randall Smith

In epidemiology, there is an inclination to consider more credible the larger estimates of exposure effect. For example, higher relative risks or rate ratios are often emphasized as a criterion for choosing among various hypothesized exposure-lag values. An alternative criterion for this choice might be based on a goodness-of-fit measure. We present examples, based on hypothetical data, in which an exposure-lag parameter is estimated by trial and error fitting: we compare the behavior of the likelihood-ratio goodness-of-fit statistic obtained over the assigned values of the parameter with that of the relative risk. We show that there can be inconsistencies between the highest-estimate and likelihood-based goodness-of-fit criteria. Concern about the validity of the highest-estimate criterion prompts us to recommend that this criterion not be used for the estimation of exposure-weighting parameters, which should preferably be based on *a priori* biological knowledge or on goodness-of-fit criteria. (Epidemiology 1995;6:387–390)

Keywords: epidemiologic methods, data analysis, misclassification.

A common approach often employed in cancer epidemiology is to discount an exposure that is thought to be irrelevant to the outcome under study. This discounting is often achieved by simply lagging the exposure by a fixed time period of k years. In a cohort study, the effective exposure at time t will be the exposure measured at time $t - k$, whereas in a case-control study, the exposure occurring less than k years before the outcome will not be considered. There is often not enough prior information, however, for the choice of a particular value of k. In such instances, a trial-and-error approach may be used, in which estimates of exposure effect are obtained over a set of possible values of k, and the results are contrasted. An extension of this method is the "window method,"[1] in which exposures occurring outside a time window are discounted. These methods are special cases of exposure-weighting (see, for example, Thomas[2] for a general formulation); other terms used are induction-latency estimation or estimation of the empirical induction period.[1] In all cases, there is the question of how to select the best estimate of these exposure-weighting parameters. Some of the epidemiologic literature[3–8] seems to indicate that the desired objective of exposure weighting is the maximization of exposure-response relations (a maximum in the relative risk, or a steeper dose-response relation).

From the National Institute for Occupational Safety and Health, Cincinnati, OH 45226.

Address correspondence to: Alberto Salvan, Institute of Systems Science and Biomedical Engineering (LADSEB-CNR), National Research Council, Corso Stati Uniti, 4, Padova, Italy I-35127.

Submitted December 22, 1993; final version accepted December 12, 1994.

The approach based on the highest estimate was built on the argument that the elimination of causally irrelevant exposure would minimize the degree of nondifferential exposure misclassification. Quoting from Rothman: "Attenuation of the effect-estimate through a lack of appreciation for the empirical induction period is a special case of non-differential misclassification. . . . The maximum value for the incidence rate among exposed subjects obtained on repeated analyses (with changes in assumption about the timing of the etiologically relevant period) is the one that indicates the most appropriate assumption about the empirical induction period."[1] In practice, investigators often use the highest-estimate approach. For example, Aselton *et al*[4] used exposure windows and selected the window that produced the maximum point estimate of relative risk. Checkoway and colleagues[5,6] re-analyzed lung cancer in a cohort of asbestos textile plant workers[7] using several exposure-weighting schemes, including exposure lagging. Their results and conclusions are based on the assumptions producing the largest point estimates of relative risk. Caplan *et al*,[8] while illustrating the use of a log-normal distribution for induction-latency modeling, seemed to suggest looking at the point estimates (for example, for a stronger dose-response).

On the other hand, other investigators have evaluated exposure-weighting schemes in the same way as model parameters are commonly estimated; that is, parameter values are chosen that maximize the model likelihood or the likelihood-ratio statistic (or, equivalently, minimize the model deviance).[2,9–14] A brief exposition of this approach is given in the Appendix.

The highest-estimate and the likelihood-based goodness-of-fit criterion may lead to the same choice of the exposure-lag parameter: examples of this occurrence are shown by Finkelstein[15] and by Stayner *et al*,[16] among

others. Nevertheless, the question arises: do the highest-estimate and the goodness-of-fit criterion always agree? An analysis of lung cancer risk among uranium miners[13] found that the log-likelihood statistic was maximized at a lag of 6 years, whereas the highest relative risk was obtained with unlagged data (Richard Hornung, 1993, personal communication). Occurrences of "flat" likelihoods with or without a maximum in the relative risk seem to be common.[17,18]

Methods

We generated hypothetical examples based on case-control data consisting of two-by-two tables with fixed case-control margins (50 cases and 250 controls). Each example consists of a two-by-two table in which entries are rearranged in a manner compatible with the effect of exposure lags: that is, more extreme lags imply a transfer of observations from the exposed to the unexposed category. The underlying model used is the logistic model for a single two-by-two table:

$$\log(\text{odds of disease}) = \alpha + \beta x,$$

where x is a binary exposure indicator (1 if exposed, 0 if unexposed), and β is the log(disease odds ratio). To compare models with different lags, we used a likelihood-ratio statistic which compares the fit of a model with the term of interest (lagged exposure) to that of a model without this term. The likelihood-ratio statistic provides information on the improvement in fit due to the lagged-exposure term over the range of selected exposure-lag values. We selected the examples from a computer-generated list of tables with preassigned values either of the odds ratio or of the likelihood-ratio statistic. In the examples, lags are simply identified as ordinal categories (1, 2, 3 . . .), with higher values associated with further transfer of observations from the exposed to the unexposed category. These examples can be thought of as originating either with point exposure or with sustained exposure with binary categories of duration or of cumulative exposure defined. We computed interval estimates for the odds ratio and two-sided P-values for the likelihood-ratio test assuming a fixed lag. We used exact mid-P methods[19–21] because of small or unbalanced table entries obtained with some of the more extreme lags, as is often encountered in practice.[22] We obtained the two-sided P-value for the likelihood-ratio test by doubling the one-sided P-value.

Results

In the example of Table 1, the same exposure lag maximizes both the likelihood ratio and the relative risk. In the example of Table 2, the likelihood-ratio statistic remains constant over a wide range of table rearrangements compatible with different hypothesized lags, whereas the relative risk varies from 2.0 to 5.4. In the example of Table 3, a practically constant relative risk is obtained over the exposure-lag values, whereas the likelihood-ratio statistic has a maximum. In Table 4, the lag maximizing the likelihood ratio is obtained at lag 3, with a relative risk of 6, whereas the highest estimate of the relative risk is 13.8 obtained at lag 7.

TABLE 1. Hypothetical Case-Control Data Showing Agreement between Likelihood Ratio (LR) and Highest Relative Risk (RR) (Each Row Represents a Two-by-Two Table with 50 Cases and 250 Controls)

Exposure Lag	Exposed Cases	Exposed Controls	RR	95% CI for RR (Exact)	LR (df = 1)	LR P-Value (Exact)
1	40	167	1.99	0.97–4.36	3.63	0.0633
2	35	109	3.02	1.58–5.93	11.9	0.0006
3	30	50	6.00	3.13–11.5	30.5	<0.0001
4	20	36	3.96	2.00–7.70	15.4	0.0001
5	10	19	3.04	1.27–6.96	6.13	0.0143
6	5	13	2.03	0.62–5.83	1.49	0.2214

Discussion

The examples built on hypothetical data show that the magnitude of the point estimate and goodness-of-fit criterion based on the likelihood may not agree. As mentioned in the Appendix, there is no obvious relation between a procedure that maximizes the exposure coefficient and one that provides the maximum likelihood estimates for the exposure coefficient and the lag parameter. The patterns of agreement or discrepancy described in our examples would still occur if all cell entries were large. For example, multiplying all cell entries in Table 4 by 100 would still produce the same disagreement as described above: the relative risks would stay the same,

TABLE 2. Hypothetical Case-Control Data Showing a Constant Likelihood Ratio (LR) Whereas the Relative Risk (RR) Has a Maximum (Each Row Represents a Two-by-Two Table with 50 Cases and 250 Controls)

Exposure Lag	Exposed Cases	Exposed Controls	RR	95% CI for RR (Exact)	LR (df = 1)	LR P-Value (Exact)
1	41	167	2.26	1.07–5.14	4.91	0.0306
2	30	107	2.01	1.08–3.76	4.96	0.0278
3	15	40	2.25	1.10–4.47	4.93	0.0274
4	8	15	2.98	1.13–7.42	4.88	0.0287
5	6	9	3.65	1.15–10.8	4.91	0.0294
6	4	4	5.35	1.16–24.3	4.88	0.0327

TABLE 3. Hypothetical Case-Control Data Showing a Constant Relative Risk (RR), Whereas the Likelihood Ratio (LR) Has a Maximum (Each Row Represents a Two-by-Two Table with 50 Cases and 250 Controls)

Exposure Lag	Exposed Cases	Exposed Controls	RR	95% CI for RR (Exact)	LR (df = 1)	LR P-Value (Exact)
1	40	143	2.99	1.46–6.53	9.84	0.0021
2	35	109	3.02	1.58–5.93	11.9	0.0006
3	30	83	3.02	1.61–5.69	12.4	0.0005
4	25	62	3.03	1.61–5.68	11.9	0.0007
5	20	45	3.04	1.56–5.81	10.6	0.0013
6	15	31	3.03	1.45–6.14	8.57	0.0038
7	10	19	3.04	1.27–6.96	6.13	0.0143
8	5	9	2.98	0.863–9.27	3.13	0.0808
9	4	7	3.02	0.744–10.8	2.58	0.1139

TABLE 4. Hypothetical Case-Control Data Showing Disagreement between the Highest-Estimate Criterion and the Likelihood-Ratio (LR) Statistic (Each Row Represents a Two-by-Two Table with 50 Cases and 250 Controls)

Exposure Lag	Exposed Cases	Exposed Controls	RR	95% CI for RR (Exact)	LR (df = 1)	LR *P*-Value (Exact)
1	40	160	2.25	1.09–4.92	5.16	0.0264
2	35	110	2.97	1.55–5.83	11.5	0.0008
3	30	50	6.00	3.13–11.5	30.4	<0.0001
4	10	8	7.56	2.74–20.9	15.3	0.0001
5	7	4	10.0	2.75–40.0	12.8	0.0006
6	5	3	9.15	2.02–47.4	8.77	0.0043
7	5	2	13.8	2.60–104	10.6	0.0018

whereas the likelihood ratios would be multiplied by 100.

The argument for choosing the highest relative risk based on nondifferential misclassification used as a basis for the validity of the window method[1] has been weakened by recent work showing that, under nondifferential misclassification, it is possible to create a bias away from the null.[23,24] Also, Flegal *et al*[25] have shown that a dichotomous exposure variable formed from an underlying continuous variable subject to nondifferential measurement error may be affected by differential misclassification. Although simulations have shown that pure random error in the original variable seems to result in bias toward the null, a bias in either direction can be obtained if a systematic component of error is also present.[26] It seems prudent to conclude that if the exposure is not inherently categorical, the highest estimate may be affected by bias of unpredictable direction, and therefore there would be no basis for the selection of induction time based on this criterion. In practical situations, the presence of misclassified covariates would add sources of bias of unpredictable direction.[27] In addition, the highest estimate can be biased away from the null. Such bias could arise if (1) the lag is a determinant of the outcome, in the sense that more events are reassigned to the unexposed category for increasing values of the lag parameter, and (2) the determination of the relevant exposure depends on the lag parameter itself.

Epidemiologic data are often not extensive enough to validate a particular choice of an exposure-weighting scheme. This limitation often results in indistinguishable values of the exposure-weighting parameters. Some authors, however, have based this conclusion on unchanging point estimates of the exposure effect (for example, Checkoway *et al*[28]), whereas other authors have reached the same conclusion based on the impossibility of discriminating between estimates of exposure effect in a statistical sense, that is, using a goodness-of-fit criterion.[17,18] The example of Table 3 shows that the point estimate for relative risk does not vary over the lag values, whereas there is a concomitant, nontrivial variation in goodness-of-fit. One value for the lag may still be preferable over other values, based on precision of the relative risk estimate. On the other hand, in the "flat" likelihood example of Table 2, in which the exposure effect does, however, show a maximum, a reasonable conclusion is to accept that it is not possible to select an exposure lag, given the data at hand.

One argument against the use of goodness-of-fit criteria for exposure-lag estimation is that these criteria are model dependent: if the model is misspecified, the goodness-of-fit criterion may not be valid. These considerations, however, apply equally to the relative risk, and therefore to the highest estimate of it, which may not be valid when estimated based on a misspecified model. In addition, a fixed lag itself may be a misspecified parameterization, thus rendering invalid any criterion of selection. For example, in a situation with sustained exposure, it can be argued that the effect estimates obtained under different values of the lag are noncomparable, since they can be thought of as the effect of exposures defined over different time intervals or windows. Other approaches have been suggested.[29] In addition, other more complex weighting schemes may be more appropriate than a simple lag.[2,9,10]

The findings of this paper, compounded by the existence of scenarios in which the highest estimate may not have validity, suggest that the estimation of exposure-weighting parameters should be based on *a priori* biological knowledge or on likelihood-based goodness-of-fit statistics and not on the highest-estimate criterion. Goodness-of-fit information should be reported when estimating exposure-weighting parameters.

Acknowledgments

We thank Harvey Checkoway, Richard Hornung, David Kriebel, and Dana Loomis for their comments on earlier drafts of this manuscript. Duncan Thomas and Jamie Robins provided comments during presentations of this work. The comments of an anonymous reviewer led to improvements in the paper.

References

1. Rothman KJ. Induction and latent periods. Am J Epidemiol 1981;114:253–259.
2. Thomas DC. Statistical methods for analyzing effects of temporal patterns of exposure on cancer risks. Scand J Work Environ Health 1983;9:353–366.
3. Rothman KJ. Modern Epidemiology. Boston: Little, Brown, 1986;58.
4. Aselton P, Jick H, Chentow SJ, Perera DR, Hunter JR, Rothman KJ. Bendectin during early gestation and pyloric stenosis. Am J Epidemiol 1984;120:251–256.
5. Checkoway H, Pearce N, Crawford-Brown DJ. Research Methods in Occupational Epidemiology. New York: Oxford University Press, 1989.
6. Checkoway H, Pearce N, Hickey JLS, Dement JM. Latency analysis in occupational epidemiology. Arch Environ Health 1990;45:95–100.
7. Dement JM, Harris RL, Symons MJ, Shy CM. Exposure and mortality among chrysotile asbestos workers. Part II. Mortality. Am J Ind Med 1983;4:421–433.
8. Caplan RJ, Marsh GM, Enterline PE. A generalized effective exposure modeling program for assessing dose-response in epidemiologic studies. Comput Biomed Res 1983;16:587–596.
9. Breslow NE, Lubin JH, Marek P, Langholz B. Multiplicative models and cohort analysis. J Am Stat Assoc 1983;78:1–12.
10. Breslow NE, Day NE. Statistical Methods in Cancer Research. vol. 2. The Design and Analysis of Cohort Studies. IARC Scientific Pub. No. 82. Lyon: International Agency for Research on Cancer, 1987.
11. Lundin FE Jr, Archer VE, Wagoner JK. An exposure-time-response model for lung cancer mortality in uranium miners: effects of radiation exposure, age, and cigarette smoking. In: Breslow NE, Whittemore AS, eds. Energy and Health. Philadelphia: Society for Industrial and Applied Mathematics, 1979;243–264.

12. Crump KS, Allen BC. Methods for quantitative risk assessment using occupational studies. Am Statistician 1985;39:442–450.
13. Hornung RW, Meinhardt TJ. Quantitative risk assessment of lung cancer in U.S. uranium miners. Health Phys 1987;52:417–430.
14. Hodgson JT, Jones RD. Mortality of a cohort of tin miners 1941–86. Br J Ind Med 1990;47:665–676.
15. Finkelstein MM. Use of time windows to investigate lung cancer latency intervals at an Ontario steel plant. Am J Ind Med 1991;19:229–235.
16. Stayner L, Smith R, Thun M, Schnorr T, Lemen R. A dose-response analysis and quantitative assessment of lung cancer risk and occupational cadmium exposure. Ann Epidemiol 1992;2:177–194.
17. Gilbert ES, Marks S. An analysis of the mortality of workers in a nuclear facility. Radiat Res 1979;79:122–148.
18. Berry G, Gilson JC, Holmes S, Lewinsohn HC, Roach SA. Asbestosis: a study of dose-response relationships in an asbestos textile factory. Br J Ind Med 1979;36:98–112.
19. Stat-Xact, version 2.04. Boston: Cytel Software, 1992.
20. Lancaster HO. Significance tests in discrete distributions. J Am Stat Assoc 1961;56:223–234.
21. Miettinen OS. Comment. J Am Stat Assoc 1974;69:380–382.
22. Checkoway H, Pearce N, Crawford-Brown DJ, Cragle DL. Radiation doses and cause-specific mortality among workers at a nuclear material fabrication plant. Am J Epidemiol 1988;127:255–266.
23. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? Am J Epidemiol 1990;132:746–748.
24. Wacholder S, Dosemeci M, Lubin JH. Blind assignment of exposure does not always prevent differential misclassification. Am J Epidemiol 1991;134:433–437.
25. Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. Am J Epidemiol 1991;134:1233–1244.
26. Brenner H, Loomis D. Varied forms of bias due to nondifferential error in measuring exposure. Epidemiology 1994;5:510–517.
27. Greenland S. The effect of misclassification in the presence of covariates. Am J Epidemiol 1980;112:564–569.
28. Checkoway H, Mathew RM, Shy CM, Watson JE Jr, Tankersley WG, Wolf SH, Smith JC, Fry SA. Radiation, work experience, and cause specific mortality among workers at an energy research laboratory. Br J Ind Med 1985;42:525–533.
29. Moulton HL, Lê MG. Latency and time-dependent exposure in a case-control study. J Clin Epidemiol 1991;44:915–923.

Appendix

A formulation of the lag interval as a parameter to be estimated is given below. Following Thomas,[2] a weighted cumulative exposure index can be written in a general form as follows:

$$D(T) = \int_{t_0}^{T} u(t) f(T - t) dt,$$

where

t = time (eg, age)
t_0 = time of hire
T = time of risk
$D(T)$ = cumulative exposure index at time T
$u(t)$ = exposure level at time t
$f(T - t)$ = weight function.

The fixed lag can be expressed by choosing $f(T - t)$ as:

$$f(T - t) = 0 \quad \text{if} \quad T - t \leq k$$
$$= 1 \quad \text{otherwise.}$$

$D(T)$ can then be written as:

$$D(T) = \int_{t_0}^{T-k} u(t) dt.$$

$D(T)$ can be incorporated as such or in a categorized form into a log-linear relative-risk (RR) model:

$$RR = e^{\beta D(T)}.$$

The likelihood is then a function of both β and k. Usually, for each fixed value of k, the likelihood (or its log) is maximized. By repeating this over several values of k, the maximum likelihood estimates for β and k are located. Note that a procedure that only maximizes β bears no obvious relation to the maximum likelihood estimates for β and k. In this context, it is easy to see that when the likelihood changes little as k varies, this will imply that the likelihood surface is flat. Although the lag parameter creates discontinuities[18] that prevent the computation of a confidence interval for this parameter based on asymptotic normal theory, it is clear that the data at hand do not contain enough information to make it possible to select an exposure lag.