European Journal of Epidemiology **11**: 365–371, 1995.

To use or not to use the odds ratio in epidemiologic analyses?

Markku Nurminen

Department of Epidemiology and Biostatistics, Finnish Institute of Occupational Health, and Department of Public Health, University of Helsinki, Helsinki, Finland

Accepted in revised form 3 March 1995

Abstract. This paper argues that the use of the odds ratio parameter in epidemiology needs to be considered with a view to the specific study design and the types of exposure and disease data at hand. Frequently, the odds ratio measure is being used instead of the risk ratio or the incidence-proportion ratio in cohort studies or as an estimate for the incidence-density ratio in case-referent studies. Therefore, the analyses of epidemiologic data have produced biased estimates and the presentation of results has been misleading. However, the odds ratio can be relinquished as an effect measure for these study designs; and, the application of the case-base sampling approach permits the incidence ratio and difference measures to be estimated without any untenable assumptions. For the Poisson regression, the odds ratio is not a parameter of interest; only the risk or rate ratio and difference are relevant. For the conditional logistic regression in matched case-referent studies, the odds ratio remains useful, but only when it is interpreted as an estimate of the incidence-density ratio. Thus the odds ratio should, in general, give way to the incidence ratio and difference as the measures of choice for exposure effect in epidemiology.

Key words: Biometry, Epidemiologic methods, Odds ratio, Risk difference, Risk ratio

Introduction

The odds ratio remains perhaps the most popular relative measure of the exposure-disease relation in epidemiology today. Yet, despite its widespread use the estimation and interpretation of this parameter for epidemiologic data analysis and inference raises questions [1, 2]. The main reason for its usefulness is not its intelligibility as an effect measure for epidemiologic research results, but rather its convenient mathematical properties when employed, for example, in the Cornfield chi-square statistic in unstratified analysis and in the Mantel-Haenszel odds ratio in stratified analysis [3, Section 14] and in the logistic model of multivariate analyses [4, 5]. Odds ratios also provide a connection between analyses of data from cohort and case-referent studies. However, because of recent developments in statistical methods in epidemiology [6], the need to present results referring to *average risks* for a disease outcome in populations in terms of odds ratios can be questioned if the objective is to determine the *individual responses* to exposure.

My purpose in this communication is to review briefly the justifications for using the odds ratio parameter in current epidemiologic research and to argue that many of these uses can either be bypassed or made implicit in the data analysis. Incidence difference and ratio then remain as the preferred measures of the exposure effect in most situations. Yet, thorough reflection on the statistical modelling of data should always precede the data analysis and the computation of a measure of exposure effect. In this paper, I discuss the statistical (nontechnical) issues of the modelling, estimability and presentation of results to provide an updated view of epidemiologic analyses.

Interpretation of effect measures in the modelling of epidemiologic data

The risk of disease is a probability measure pertaining to an individual. In epidemiology, average risks can be estimated in a population as an incidence-proportion. For a comparison of risk among exposed (R_1) and unexposed (R_0) populations, the simplest measures of excess risk are the risk difference, $RD = R_1 - R_0$, and the risk ratio, $RR = R_1/R_0$. The choice between the difference and ratio is made depending on the constancy of their values over potential modifiers of the effect parameter. In causal research, with high proportion-type rates in the reference population (R_0) this *principle of invariance* may theoretically call for the use of the risk-odds ratio, $OR = [R_1(1 - R_1)]/[R_0(1 - R_0)]$, and the logit difference, log(OR) (O. Miettinen, personal communication). However, empirical evidence does not

seem to support the assumption that causal effect is more constant across populations when expressed using odds ratios rather than risk ratios [7]. These more complicated measures of exposure effect enjoy some good statistical properties. The OR has generally been used as an auxiliary parameter in unstratified, stratified, and in logistic regression analyses of case-referent data, including mortality-odds ratio analyses in occupational mortality studies [3, Section 17.2, 18.4, A.2.6]. The OR may also be the most natural (i.e. time-invariant) measure in stochastic modelling of disease processes. However, it is only interpretable biologically if it estimates the incidence-proportion or the incidence-density ratio [8].

Statistical modelling depends on the characteristics of the data available. For dichotomous data with binomial distributions, the log(OR) is still the most convenient means for modelling the probability of outcome event. The RR and RD models have the unpleasant potential of producing estimated probabilities outside the zero-to-one range. Further, the log(OR) is directly related to the Bayes theorem, which acts as an explanation of the log(OR) as a natural parameter in the modelling of dichotomous data. In incidence data, Poisson-process modelling is appropriate and leads to a Poisson distribution in stable cohorts and to proportional hazards modelling in the follow-up of dynamic populations. In these models the RR is in some (intuitive) sense natural.

Greenland [8] has presented a strong theoretical argument against the utility of the OR as an effect measure in epidemiologic analyses: '. . . only incidence differences and ratios possess direct interpretations as measures of impact on average risk or hazard. Consequently, odds ratios are useful only when they serve as incidence-ratio estimates, and logistic and log-linear models are useful only insofar as they provide improved (smoothed) estimates of incidence differences or ratios.'

The choice of the comparative parameter also affects the assessment of confounding, both in stratified analyses and in logistic regression modelling. In deciding whether a covariate that is a risk factor of the disease under study is also associated with the exposure (i.e. a confounder), two criteria have been employed. First, a definition of a confounder has been based on the statistical notion of *collapsibility*. According to this definition, if the crude effect measure calculated from unstratified data equals the stratified parameter estimate (upon stratification according to a potential confounder), the parameter is collapsible and the crude measure is unconfounded. Accepting this definition, Miettinen & Cook [9] rejected the OR as a measure of intrinsic interest and instead selected the RR and RD, because, unlike the OR, both the RR and RD are collapsible whenever confounding is absent.

Second, the criterion of *comparability* between the exposed and unexposed subjects has been favored by many epidemiologists [10]. According to this criterion the control of a covariate that is only a risk factor of the disease outcome but has no association with the exposure is irrelevant for validity. Nevertheless, in his review of statistical modelling in epidemiology, Gail [6] points out: 'First, inclusion of such a "balanced" covariate can improve the precision of estimates of exposure effect. Second, except for linear and multiplicative regressions, failure to include a "balanced" covariate will lead to a biased estimate of exposure effect with bias toward the null in logistic models.'

Gails's [6] point applies in situations when the probability of the outcome event is high. (For details, see Gail [11].) This oddity of the OR has, for example, prompted prominent epidemiologists to caution others to avoid the OR as a measure of exposure effect in favor of the RD and the RR, which correspond to linear and multiplicative regressions [see, eg, 3, Section A.2.4].

Recently Savitz [12] commented that, although odds ratios are useful exclusively as a reflection of other parameters, the term odds ratio is so familiar to epidemiologists that it is the preferred term in reporting study results. However, the epidemiologist should clearly present the basis for the (statistical and biological) inference of the OR contained in the paper. Of course, this basis should not only be presented when the OR is the estimated parameter, it should always be included.

Analysis of population-based studies

Cohort population data or population cross-sectional data

Data derived from a *cohort* (closed) *population* for the assessment of excess risks are not substantively meaningful with respect to the incidence-odds ratio. Instead epidemiologic analyses of cohort data focus on the incidence-proportion (or cumulative incidence) difference and ratio. Reporting from their experience in clinical trials, Sinclair & Bracken [7] concluded: 'Because the control group's risk affects the numerical value of the odds ratio, the odds ratio cannot substitute for the risk ratio in conveying clinically important information to physicians. This is especially important when large treatment effects are shown in trials carried out in populations at high baseline risk.'

In the same vein, in occupational-epidemiology studies of the incidence of illness the concern is not with the OR in itself, even though it has been viewed by some researchers as the object of inference in prevalence data from *population cross-sections*. Other authors prefer prevalence ratios when the prevalence of the disease rate is rather high [13, 14], as in a study of a common pregnancy outcome or a

degenerative change. As commented by Axelson et al. [13]: 'However, the cross-sectional approach is usually the only possibility to study, for example, musculo-skeletal disorders in relation to various types of work-load. Rather high risk estimates are often presented from such studies in terms of odds ratios, whereas the prevalence ratios would be lower.'

Example 1. Leino et al. [15] presented results showing that hairdressing is a risk factor for rhinitis symptoms. In the populations of 355 hairdressers and 583 saleswomen the frequencies of the self-reported symptoms obtained in a telephone interview were highly prevalent: 62.3 and 50.3%, respectively. In a preliminary analysis of these *cross-sectional* data, the authors computed the point estimate (and the 95% confidence interval estimate) of the prevalence-odds ratio (OR) to be 1.6 (1.2–2.1). The directly estimable prevalence rate ratio (RR) is 1.2 (1.1–1.4). Although the qualitative conclusion regarding the equality of the compared prevalences was not affected by the choice of the comparative parameter (i.e. the lower confidence limit exceeds 1 both for the OR and the RR), the quantitative estimate of the *excess risk* measured in terms of the OR-1 = 60% was three times in error compared to the more correct value, RR-1 = 20%. Moreover, the confidence interval for the OR was unduly wide (i.e. imprecise).

Example 2. Viikari-Juntura et al. [16] examined the effects of occupational factors on the risk of neck disorders in three vocational categories: machine operators (static work involving whole body vibration), carpenters (dynamic physical work), and office workers (sedentary work). By design, the study was based on the retrospective experience of a *cohort* population. The outcome considered was a possible change in the distribution of symptom statuses. The 1-year-period prevalences of recurrent severe neck trouble obtained in a questionnaire survey were as high as 77% among the 185 machine operators, 71% among the 127 carpenters, and 49% among the 51 office workers. Yet, the authors chose the prevalence-odds ratio as the effect measure and in a logistic regression analysis estimated its value (here unadjusted) for the following occupational comparisons: machine operators vs office workers, OR = 3.5 (95% confidence interval 1.8–7.1); carpenters vs office workers OR = 2.5 (1.3–4.9). However, when these cohort data with count denominators are analysed in terms of the prevalence rate ratios we obtain the corresponding estimates: RR = 1.6 (1.2–2.2) and RR = 1.4 (1.1–2.0). Thus, it is evident that the OR-based analysis estimated the exposure effects on the risk many times over because the considered outcome, as defined, was not uncommon. Besides the point estimates for the OR being location-biased (i.e. inaccurate), the associated confidence intervals were length-biased (i.e. imprecise).

In certain epidemiologic studies subjects come from an underlying *cohort* of matched pairs, but only those pairs in which an outcome event occurs are ascertained. Although the OR is used in this type of incomplete sampling design, its implications are not fully appreciated. As remarked by Greenland [17]: 'Such data can be analysed by using odds ratio modelling methods, but these methods present interpretational problems and do not use information from matched pairs in which both subjects experience the outcome.'

To remedy the situation Greenland [17] presented a method that makes use of double-outcome pairs and is a risk-ratio analogue of the odds-ratio method in conditional logistic regression analyses.

Dynamic population data

When the rate denominators represent the population time of observation for the incidence of disease events in the follow-up of a *dynamic* (open) *population*, the comparative parameters at issue are primarily the incidence-density difference and ratio and their derivatives, not including the OR [3, Section A.2]. There are two reasons for this choice. First, the notion of odds is the risk divided by its complement. Second, if the disease incidence is low and the exposure proportion remains stable, the incidence-proportion ratio (estimate of risk ratio) is approximated more accurately by the incidence-density ratio than by the incidence-odds ratio [18].

Approaches to regression analysis

Regression analyses of cohort or dynamic population-time data are typically done by means of logistic or Poisson modelling. From the estimates of regression coefficients given by logistic analysis programs [e.g., in the SPSS (Statistical Package for the Social Sciences) and the BMDP (Biomedical Computer Programs) packages] only the OR can be calculated. However, the binomial regression approach can be modeled to produce estimates of the RD and RR, and the analyses can be carried out [in the GLIM (General Linear Interactive Modelling) system] with the help of special macros [19].

The main rationale for the use of the logistic multiple-regression model in epidemiologic follow-up studies of disease risks is that it constrains the predicted probability of the disease occurrence to the (0, 1) range. This restriction does not hold if one were to induce a linear model for the disease risk as a function of the covariates. On the other hand, the logistic (or log-linear) model form implies a multiplicative statistical relation between the effects of the study exposures. As shown by Greenland [20], if an additive biological model holds, the logistic analysis requires three parameters to adequately summarise the joint effects of only two variables, and gives the

impression that these factors may somehow biologically interact in the base population.

The main problem in applying logistic modelling to follow-up studies is that the risk period is only implicitly taken into account in the estimated parameters. Logistic function can quantify the effect of a risk factor on the probability of a disease, but it does not naturally translate into the effect on the change with time of that risk. To cope with the problem, one may explicitly include a time variate (such as calendar time, age, or duration of exposure) in the model. The second problem with the logistic model is that it cannot deal with the effects of competing risks. This is because it represents a model relating the covariate to a probability and not to a hazard rate. In other words, it does not use person-time denominator data and thus the model cannot estimate the time to realisation of risk. Besides being unable to manage satisfactorily the problem of lost cases (they are simply excluded form the analysis), the logistic model cannot handle events that are not rare. For a technical discussion on these issues, see Manton & Stallard [21].

The Poisson regression approach can also be modeled to produce estimates of the RD and RR. Then the RR estimates the incidence-density ratio. But it is not clear how results of a Poisson model with a logit link should be interpreted. Unlike the logistic regression model considered above, the Cox regression is an appropriate method for modelling hazard (or instantaneous incidence-density) ratio in the analysis of failure (or survival) data. In tightly stratified data, the Poisson analysis using a piecewise exponential failure time model and the Cox hazard-rate regression model are expected to yield very similar results since incidence-density ratios can be interpreted as ratios of *average hazards* [8].

Another possibility in the context of multidirectional contingency tables is to apply the method proposed by Berry et al. [22] to the values of incidence-proportions, estimates of risks R_i (i = 0, 1, . . .), fitted by the logistic model, and then to present the results as standardised risks or risk ratios. This procedure can be done by a simple (e.g. GENSTAT) program with the facility to access fitted values and the table operations. Berry et al. [22] state their position germane to logistic regression analysis as follows:

'Whatever method is used, presentation of results after allowance for covariables should be in a form similar to that which would be given if no covariables were under consideration. Quoting parameter estimates in the logistic model does not achieve this and is artificial since the logit transformation would not be necessary if there had been only the one factor of interest, and no covariables.'

Thus, although results are often expressed as regression coefficients or odds ratios, these measures can appear unnatural to the investigator. On the other hand, proportions or risks are easily interpretable and have direct meaning. This *principle of intelligible reporting* of research results carries over to epidemiologic studies with case-referent sampling of the population experience.

Analysis of case-referent sampling studies

Samples from a cohort base or a cross-sectional base

The classic *raison d'être* given for the use of the OR in analyses of *case-referent series* is that, when the sample of the base population consists of the disease-free subjects (noncases) of a cohort or a population cross-section, the exposure-odds ratio approximates the incidence-proportion ratio. This condition is obtained provided that the outcome event is rare in the population-time frame – the so-called rare disease assumption (18, 23, 24) – as is often the case in cancer studies. However, if a representative sample of the entire base is drawn, the case-base data allow for the incidence-proportion ratio to be estimated whether or not the disease is rare [25, 26]. Miettinen [3, Section A.6] has given cogent reasons (conceptual clarity, informativeness, practicality) for sampling the base itself at the start of a follow-up instead of sampling the noncases at the end of a study. Various methods for the RR analyses are available for binary data [25–31]. Of course, for a rare disease, there will be very few events in the base, and thus the case-base and traditional case-referent studies will yield nearly identical results.

Example 3. Table 1 provides a numerical illustration of the analysis of unstratified case-referent data. Because the case occurrence is not uncommon, the case-control (case-noncase) odds ratio (OR = 2.25) gives a poor approximation to the incidence-proportion ratio (1.37). Note that the incidence-proportion ratio (estimate of risk ratio) is approximated more accurately by the incidence-density ratio (1.50) than by the case-control odds ratio. Remarkably, the case-base analysis also yields estimates of the risk difference. Of course, a 100% sampling of the base population will produce identical risk estimates with those obtained from the analysis of full cohort data.

In stratified analysis of case-referent data, the Mantel-Haenszel common odds ratio estimator (OR_{MH}) is applicable in the noncase sampling of a cohort base population. The OR_{MH} does not assume homogeneity (constancy) of stratum-specific parameters. Under heterogeneity, however, the OR_{MH} has the disconcerting property that it varies according to irrelevant features of the sampling design and does not consistently estimate any meaningful population parameter [32]. But, with two additional assumptions the population OR (estimated by OR_{MH}) indirectly

Table 1. Estimates of effect measures for binary data in a case-referent study*

Data series	Exposed subcohort	Unexposed subcohort	Total
Cases	600	400	1000
Noncases	400	600	1000
Base sample	100	100	200

Size of base population = 2000
Sampling fraction of base population = 200/2000 = 0.1
Duration of follow-up time = 1 year
Total population time = 2000 years

* *Analysis*:
Case-control odds ratio =
(600/400)/(400/600) = 2.25
Case-base analysis (for methods, see reference 33):
- incidence-density ratio =
(600/100)/(400/100) = 1.50
- incidence-density difference =
(600/100 – 400/100) (200/2000 years) = 0.2/(year)
- incidence-proportion ratio =
{1 – exp [–(600/100) (200/2000 years) (1 year)]}/
{1 – exp [–(400/100) (200/2000 years) (1 year)]} =
0.45/0.33 = 1.37
- incidence-proportion difference =
{1 – exp [–(600/100) (200/2000 years) (1 year)]} –
{1 – exp [–(400/100) (200/2000 years) (1 year)]} =
0.45 – 0.33 = 0.12

approximates the population RR: (a) disease outcome is rare; (b) sampling of the study base population is representative (absence of selection bias). Instead, in the case-base sampling design the RR parameter can be estimated directly.

Samples from a dynamic population base

Characteristically, cohorts are unstable with respect to the distribution of covariables of the exposure-disease relation over time, while dynamic populations can and tend to remain stable with unchanging proportions exposed. When the study base is a stable population experience, the random base sample is drawn to obtain an estimate of the ratio of the denominator (population-times) of the compared rates and, thereby, of the incidence-density ratio [3, 26, 33] and difference [30, 33] without any assumption about the rarity of the disease. Then one can estimate the ratio but not the difference without knowledge of the base sampling fraction. If, however, the study is designed to estimate the RR and RD for incidence-proportion type rates from case-base data, and the risks are not low, the estimators for incidence-density ratio and incidence-density difference need a slight modification [30, 33]. Still the incidence-density ratio is very frequently the parameter of concern in modern epidemiology, much more so than the incidence-proportion ratio, which is generally relevant only for short latency studies, such as investigations of acute food poisoning.

However, the stability assumption may be untenable also for a dynamic population. If one stratifies tightly enough for age or some other time variable, the overall OR estimates the incidence-density ratio (assumed constant); therefore some epidemiologists tend to interpret the odds ratio for case-referent exposure as an estimate of the incidence-density ratio. A time-matched design that is used for estimating the incidence-density ratio via the OR [24] can, however, be replaced by stratification of the time period in case-base data for RR analyses [33].

Increasingly, epidemiologists are using the time-stratified sample of the study base population from which the cases were identified instead of analysing matched samples. Such a case-base design with a stratified or modelling approach to analysis offers a viable alternative that is preferable in many situations [4, p. 276]. The practical advantages of this alternative approach have been stressed by Wacholder (34, 35). In both stratified analysis or modelling, a problem persists if the matching variable has been nominally scaled with many categories (such as locality) and relatively few cases. Then it is necessary to use the methods of an individually matched OR analysis [see 36, pp. 250–251, 297–298].

Approaches to regression analysis

In case-referent data (unmatched or matched), *regression analyses* can be carried out with logistic modelling or Poisson modelling. The conditional logistic model for case-noncase data yields estimates of the exposure-odds ratio [4], whereas for case-base data it provides estimates of incidence-density ratio [37]. It would be better to forget the previously advocated data-analytic practice of modelling for exposure as the outcome variate [3, Section 18.4, 38] as a conceptually inverted approach because computational reasons no longer exist for indirect risk modelling. Recently, Schouten et al. [39] have adapted the standard logistic regression on a manipulated case-cohort data set to yield pseudo-likelihood-based estimators for the RR instead of estimators of the OR, even without invoking the rare disease assumption.

Alternatively, a Poisson regression approach in case-base data allows the RR to be estimated directly if the base (cohort or dynamic) population is stable within the time strata [33]. For instantaneous time strata, essentially the same equally valid results would be expected from the time-matched logistic regression and the Poisson regression with a piecewise exponential model for the RR. Thus the modelling (likelihood formulation) and computation (likelihood estimation) can be done without the OR being resorted to as an auxiliary parameter in the estimation of the risk ratio and the incidence-density ratio.

Concluding remarks

When the disease under study is common or the prevalence of exposure in the base population is unstable, the OR can give grossly biased estimates of the underlying true RR. Even though the OR is a frequently used parameter, particularly for the conditional logistic regression for matched case-referent data, the case-base design with Poisson modelling does permit the excess risks, the RR and RD, to be estimated directly without any rare disease assumption. Therefore, the OR has become irrelevant as an estimate of the incidence-proportion ratio or risk ratio in cohort studies, and it remains useful in the case-referent sampling of a stable dynamic population only if it is interpreted as an estimate of the incidence-density ratio. Thus the odds ratio should, in general, give way to the incidence ratio and difference as the measures of choice for exposure effect in epidemiology.

This paper will have met its purpose if relevant discussion on the use of effect measures in epidemiology is stimulated and proceeds in a constructive direction.

Acknowledgements

I thank Drs T. Nurminen, J. Mandryk and O. Miettinen for their helpful comments on an earlier version of this paper.

References

1. Pearce N. What does the odds ratio estimate in a case-control study? Int J Epidemiol 1993; 22: 1189–1192.
2. Lee J. Odds ratio for cross-sectional data? Int J Epidemiol 1994; 23: 201–203.
3. Miettinen OS. Theoretical epidemiology: Principles of occurrence research in medicine. Albany, NY; Delmar Publishers, 1985.
4. Breslow NE, Day NE. Statistical methods in cancer research, Vol. 1: The design and analysis of case-control studies. Lyon: International Agency for Research on Cancer, 1980. IARC Scientific publications, No. 32.
5. Breslow NE, Day NE. Statistical methods in cancer research, Vol 2: The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, 1987. IARC Scientific Publications, No. 82.
6. Gail MH. A bibliography and comments on the use of statistical models in epidemiology in the 1980s. Stat Med 1991; 10: 1819–1885.
7. Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomised trials. J. Clin Epidemiol 1994; 47: 881–889.
8. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. Am J Epidemiol 1987; 125: 761–788.
9. Miettinen OS, Cook EF. Confounding: Essence and detection. Am J Epidemiol 1981; 114: 593–603.
10. Greenland S, Robins J. Identifiability, exchangeability and epidemiologic confounding. Int J Epidemiol 1986; 15: 413–419.
11. Gail MH, Wieand S, Piandatosi S. Biased estimates of treatment effect in randomised experiments with nonlinear regressions and omitted covariables. Biometrika 1984; 71: 431–344.
12. Savitz DA. Measurements, estimates, and inferences in reporting epidemiologic study results [editorial]. Am J Epidemiol 1992; 135: 223–224.
13. Axelson O, Fredrikson M, Ekberg K. A comment on the implications of using odds ratios or prevalence ratios in cross-sectional studies. In: Hemon D (ed), Book of abstracts of the 8th International Symposium Epidemiology in Occupational Health, 10–12 September 1991, Paris, 1991: 23.
14. Axelson O, Fredrikson M, Ekberg K. Use of the prevalence ratio *v* the prevalence odds ratio as a measure of risk in cross sectional studies. Occup and Environ Med 1994; 51: 574.
15. Leino T, Tammilehto L, Luukkonen R, Kanerva L, Nordman H. Respiratory and skin symptoms in hairdressers. In: Program résumé of the 43rd Nordic Work Environment Meeting, 28–30 August 1994, Loen, Norway, 1994: 87.
16. Viikari-Juntura E, Riihimäki H, Tola S, Videman T, Mutanen P. Neck trouble in machine operating, dynamic physical work and sedentary work: A prospective study on occupational and individual risk factors. J Clin Epidemiol 1994; 47: 1411–1422.
17. Greenland S. Modelling risk ratios from matched cohort data: An estimating equation approach. Appl Statist 1994; 43: 223–232.
18. Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. Am J Epidemiol 1982; 116: 547–553.
19. Wacholder S. Binomial regression in GLIM: Estimating risk ratios and risk differences. Am J Epidemiol 1986; 123: 174–184.
20. Greenland S. Limitations of the logistic analysis of epidemiologic data. Am J Epidemiol 1979; 110: 693–698.
21. Manton KG, Stallard E. Chronic disease modelling: Measurement and evaluation of the risks of chronic disease processes. London: Charles Griffin & Company Ltd, 1988. Mathematics in medicine, No. 2.
22. Berry G, Mandryk J, Mock P. Presentation of results of logistic regression: Analyses of cohort and cross-sectional studies. In: Francis IS, Manly BFJ, Lam FC (eds), Pacific Statistical Congress. Amsterdam: Elsevier Science Publishers (North-Holland), 1986: 79–82.
23. Cornfield J. A method of estimating comparative rates from clinical data: Applications to cancer of the lung, breast and cervix. J Natl Cancer Inst 1951; 11: 1269–1275.
24. Greenland S, Thomas DC, Morgenstern H. The rare-disease assumption revisited: A critique of 'estimators of relative risk for case-control studies'. Am J Epidemiol 1986; 124: 869–876.
25. Kupper LL, McMichael AJ, Spritas R. A hybrid

epidemiologic study design useful in estimating relative risk. Am Stat Assoc 1975; 70: 524–528.
26. Miettinen OS. Estimability and estimation incase-referent studies. Am J Epidemiol 1976; 103: 226–235.
27. Miettinen O. Design options in epidemiologic research: an update. Scand J Work Environ Health 1982; 8 (suppl 1): 7–14.
28. Miettinen O, Nurminen M. Comparative analysis of two rates. Stat Med 1985; 4: 213–226.
29. Greenland S. Adjustment of risk ratios in case-base studies (hybrid epidemiologic designs). Stat Med 1986; 5: 579–584.
30. Miettinen OS, Caro JJ. Principles of nonexperimental assessment of excess risk, with special reference to adverse drug reactions. J Clin Epidemiol 1989; 42: 325–331.
31. Nurminen M. Analysis of epidemiologic case-base studies for binary data. Stat Med 1989; 8: 1241–1254.
32. Greenland S. Interpretation and estimation of summary ratios under heterogeneity. Stat Med 1982; 1: 217–227.
33. Nurminen M. Assessment of excess risk in case-base studies. J Clin Epidemiol 1992; 45: 1081–1092.
34. Wacholder S. Practical considerations in choosing between the case-cohort and nested case-control designs. Epidemiology 1991; 2: 155–158.
35. Wacholder S, Boivin J–F. External comparisons with the case-cohort design. Am J Epidemiol 1987; 126: 1208–1209.
36. Rothman KJ. Modern epidemiology. Boston, MA: Little, Brown and Company, 1986.
37. Prentice R. Use of the logistic model in retrospective studies. Biometrics 1976; 32: 599–606.
38. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 1986; 73: 1–11.
39. Schouten EG, Decker JM, Kok FJ, Le Cessie S, Van Houwelingen HC, Pool J, Wandenbroucke JP. Risk ratio and rate ratio estimation in case-cohort designs: Hypertension and cardiovascular mortality. Stat Med 1993; 12: 1733–1745.

Address for correspondence: Markku Nurminen, Topeliuksenkatu 41 a A, FIN-00250 Helsinki, Finland
Phone: 358-0-4747408; Fax: 358-0-2414634
E-mail: mnur@occuphealth.fi