Some recollections on the theoretical evolution of epidemiologic methods from the biostatistical perspective[*]

Markku Nurminen

Finnish Institute of Occupational Health, Helsinki, Finland

Correspondence to: Dr Markku Nurminen, Department of Epidemiology and Biostatistics,
Finnish Institute of Occupational Health, Topeliuksenkatu 41aA, FIN-00250 Helsinki,
Finland.
Email: Markku.Nurminen@ttl.fi

[*] Presented at the Sven Hernberg Symposium held in Helsinki on February 8th 2002 commemorating the 30th anniversary of the Department of Epidemiology and Biostatistics at the Finnish Institute of Occupational Health.

The year 2002 demarcates the 30th anniversary of the Department of Epidemiology and Biostatistics at the Finnish Institute of Occupational Health (FIOH). Professor Sven Hernberg, the appointed Head of Department, hired me as a young mathematician to the position of a biostatistician at the beginning of year 1972. Ever since then, I have practiced research in the field of occupational and public health, and witnessed the rapid and rich evolution of epidemiology and biostatistics. In this paper, I recollect some of the notable theoretical developments and recent advances from a subjective perspective.

Training in epidemiology and biostatistics

In 1972, I completed the first International Advanced Course on Epidemiologic Methods in Occupational Health; the certificate was signed by Professor Olli S. Miettinen, Harvard School of Public Health, Boston, and Professor Sven Hernberg, FIOH, Helsinki. This course was held under the auspices of the FIOH, and several future prominent epidemiologists participated, among whom was the dignitary, Professor Olav Axelson, from Linköping University, Sweden. This academic-level training in epidemiology has continued over a period of 30 years, and its seeds have fallen on fertile ground, as significant work has been done to promote workers' health in Finland and internationally. Much of this achievement is due to the pedagogically high-level and thought-provoking lecturing of Professor Miettinen. I will highlight below some of his contributions to the development of theoretical epidemiology in the years that followed.

In addition to the courses in epidemiology, the FIOH and the Nordic Institute for Advanced Training in Occupation Health (NIVA) organized international courses in biometry in the 1980s and early 1990s, with Professors David Oakes, Rochester University, NY, and Ben Armstrong, London School of Hygiene and Tropical Medicine, in the cathedra. They were expert teachers especially in the application of survival analysis (Oakes 2001) and measurement error modeling (Armstrong 1998), using the GLIM (Generalized Linear Interactive Modeling) program (McCullagh and Nelder 1983), to solve problems in the field of occupational health.

Some basic methods for epidemiologic research

My first professional assignment at the FIOH was a study on the mortality of workers in an anthophyllite asbestos quarry and mine in Finland (Nurminen 1972). In the data analysis, I used the standardized mortality ratio (SMR) as a summary measure of the fatalities. This statistic, as we are taught, for example, in Sven Hernberg's textbook, Introduction to Occupational Epidemiology (Hernberg 1992), summarizes the observed survival experience of a cohort relatively to that expected from the vital statistics of a "standard" population. The statistical inference using the SMR is not based solely on empirical observations. Rather, it is founded on the convolution of the data and the underlying statistical model, which, by the researcher's selection, is adopted to imitate the stochastic process that generated the data. In the case of the SMR, a viable model assumes that the observed number of deaths follows a Poisson distribution with the force of mortality (the intensity of risk of death) (Chiang 1968) as the single parameter of the model. A test of the hypothesis that there is no excess mortality can be derived as a score statistic from the likelihood function (Fisher 1973) for the intensity of the Poisson probability model. Likelihood-based intervals for the SMR are obtainable as a function of the specified intensity, given that the observed data are regarded as fixed.

Unfortunately, the simple analysis described above does not extract fully the information contained in a study of the mortality of the asbestos worker cohort. The cohort life-table technique (Chiang 1968) offers a more sophisticated approach to the description and analysis of the survival experience. This method is a stochastic representation of the process, modeled usefully as a Markov process, which produces the health realizations with death as the final ("absorbing") state. The life expectation at a given age is an easily calculated measure that has a direct probabilistic interpretation. This measure can be communicated to the decision-makers of health policies, for example, in terms of years of lives lost due to exposure to asbestos, or, alternatively, the years of life gained by the reduction of asbestos exposure following preventive measures, and through legislation. The results of the analysis can be depicted graphically by comparing the age-specific survival trend of the exposed cohort and that of the general population. One should keep in mind, however, the flattening of the true effect resulting from the so-called 'healthy worker effect, (Hernberg 1992): the people who keep on working and do not fall into the state of work disability are the surviving fitter ones. We know today that this problem can be analytically handled via the structured equations modeling approach (Robins 1986) that is rather ingenious, but difficult to grasp.

"Prevalence odds = Incidence Density × Average Duration" is a basic intuitive demographic identity that is taken to hold in stationary populations (Miettinen 1976). The methodologic folklore in epidemiology has contained inaccuracies of the relation. Keiding (1991) interpreted the term incidence as intensity (hazard) and prevalence as probability, and provided a rigorous proof of the particular relation assuming that average incidence is age independent. Alho (1992) derived a more general version of the relation that permits both incidence and discounted disease duration to be age-dependent. The treatment of the epidemiologic concepts in terms of mathematical and probabilistic models strengthened the theoretical basis of the field.

Developments during the Mantel-Haenszel era in epidemiology

Cornfield (1951) was the forerunner of modern epidemiology in the application of biostatistics. His key contribution to the development of the case-control study was to point out two observations. First, the relative risk of developing the illness in exposed persons as compared to non-exposed ones can be approximated for diseases with a low incidence, by the ratio of the odds of having been exposed, contrasting the illness cases to noncases. The second observation was that the exposure-odds ratio (OR) can be estimated in a case-control study. In order for the OR and other statistics estimated from the data to be unbiased, Cornfield assumed that the case and control groups are representative samples from the case and control domains in the "general" population. His solution to the statistical problem of the interval estimation of the OR arising from "retrospective" studies was based on the likelihood function of two independent binomials for the cases and noncases (Cornfield 1956). The same likelihood-based result was derived independently by Fisher (1962). This elegant one-page paper, published 40 years after Fisher proposed the idea of likelihood (Fisher 1922), exemplifies that a publication can be concise when it is to the core of the problem.

In a landmark paper, Mantel and Haenszel (1959) clarified the relation between case-control (retrospective) and cohort (prospective) studies by observing that the only conceptual difference between these two approaches was that the former involved sampling from the cohort rather than conducting a census of its population.

For the analysis of epidemiologic data in the form of multiple fourfold contingency tables, Mantel and Haenszel (1959) derived a Chi-squared statistic with 1 degree of freedom by using an argument that involved conditioning (unecessarily) by the marginal rates of the tables. This efficient test was earlier derived by Cochran (1954) using an unconditional formulation. The test is used in a stratified analysis to control for the confounding bias by an extraneous determinant of the disease outcome. Moreover, Mantel and Haenszel (1959) gave an estimator of the summary OR parameter across the strata of the confounder. The estimator was useful for epidemiologists, and it was being used for two different types of data layouts: a small number of tables with large frequencies, and a large number of tables with small frequencies (e.g. matched series). However, it took 25 years to develop a simple and robust formula for interval estimation of the Mantel-Haenszel OR (Robins et al. 1986).

A summary risk ratio (RR) estimator for cohort studies was proposed by Miettinen in 1972 in an unpublished manuscript, which served as material for the first Epidemiologic Methods course mentioned earlier. An estimator of the common RR that is completely analogous with the Mantel-Haenszel OR estimator is almost as efficient as the corresponding iterative maximum likelihood estimator (Nurminen 1981).

The Mantel-Haenszel procedure is simple and free of assumptions, and yields a consistent estimate that converges in probability to the true OR parameter as the sample size increases. The paper (Mantel and Haenszel 1959) had a huge impact, and is still widely popular. From 1974 to 2002, it received over 5,700 citations, and it continues to be cited at the rate of about 160 per year (source: Institute for Scientific Information, Web of Science).

A shift towards modern epidemiology

In the cohort sampling scheme, according to traditional statistical theory (Cochran 1963), independent representative samples are drawn from the exposed and non-exposed populations (in statistical lingo: "infinite super-populations"). As described above, the classic case-control study inverted the cohort design by drawing independent random samples from the sub-populations of cases and noncases. In a remarkable paper, Miettinen (1976) demonstrated that the estimation of OR in case-referent studies on incidence rates can be done without any assumption about the 'rarity' of illness. For the derivation, he abandoned the classic sampling model for case-referent studies. Instead, a modern epidemiologist designs a study base by choosing from the source population the relevant experience that (s)he desires to study. The study population is either a cohort (closed) population or a dynamic one (open with population turnover), and the researcher's task is to record the cases that arise in the base population and to draw a reference sample of the study base (Miettinen 1982). The cases and the referents are then classified by the categories of the etiologic determinant. The case series provides the numerators of the compared rates, whereas the base sample provides the denominators. Since then, this (case-base) design option in epidemiologic research has become the model that underlies many modern variants of the case-referent study. Examples of designs with efficient sampling of the referents include a nested case-referent design (Langholtz and Goldstein 2001), a case-cohort design (Miettinen 1982, Nurminen 1992), two-stage sampling design (Breslow and Cain 1988, Cain and Breslow 1988), and different case-pseudocontrol sampling designs (Greenland 1999). Rothman (1986) has concluded, "The sophisticated use and understanding of case-control studies is the most outstanding development of modern epidemiology." Breslow (1996) gives a lucid review of the major contributions of statistics in epidemiology.

Likelihood-based inference on epidemiologic parameters

Modern approaches to the analysis of epidemiologic data originate from the development of likelihood inference based on explicit probability models (Fisher 1922). Fisher (1935) introduced the likelihood inference for the OR parameter in a fourfold table for which he assumed an extended hypergeometric distribution. Likelihoods for the risk difference (RD) and RR can be modeled in a similar manner (Clayton and Hills 1993). The extended definitions of likelihood assume multiple formulations: conditional-likelihood (Bartlett 1936, Andersen 1973), partial-likelihood (Cox 1975), marginal-likelihood (Bartlett 1936, Kalbfleish and Sprott 1970), quasi-likelihood (Wedderburn 1974), and, profile-likelihood (Clayton and Hills 1993, Appendix C).

In the early 1980s, biostatistical methodology at the FIOH was carried out with Professor Olli Miettinen as the mentor. The work culminated in the publication of his textbook Theoretical Epidemiology (Miettinen 1985). Part 2 of the book considers mainly the comparative analysis of epidemiologic rates, in terms of the RD, RR, and OR parameters, both for stratified data and under a regression model. The likelihood-based inference on the comparative parameters provided a unified approach for significance testing and parameter estimation. The relative theoretical merits of the Fieller-type (Fieller 1944), likelihood score and likelihood ratio statistics were examined (Miettinen and Nurminen 1985, Nurminen 1986, Nurminen and Miettinen 1990). Simulation studies (Beal 1987, Gart and Nam 1988, Nurminen and Nurminen 1990, Farrington and Manning 1990, Chan and Zhang 1999) have evidenced that the proposed (asymmetric) interval estimation method with a constrained maximum likelihood estimate of the variance performs better than the usual asymptotic intervals in small samples in terms of actual confidence level.

Approaches to epidemiologic regression analysis

Regression analysis encompasses a vast array of techniques (Greenland 1998a, Greenland 1998b). A large variety of extensions to the linear regression model are available today for epidemiologists. In the following, only the basic methods used for modeling in epidemiology will be mentioned.

A logistic regression model can be applied to investigate the simultaneous effects of variables on disease risk. The response can be binary or ordinal-scaled. Several exposure variables, effect-modifiers and confounding factors may be accommodated. The methodology was developed in the 1960s for the needs of large cohort studies on cardiovascular disease, particularly the Framingham study in the USA.

Statistically, the methods were derived using the discriminant function (Truett at al. 1967) and maximum likelihood (Walker and Duncan 1967) approaches. The logistic method has been applied in many other fields. In the occupational health field, Alho (1990) developed a conditional logistic estimation procedure to solve a dual registration problem of the occupational disease registry at the FIOH.

The logistic model can be used to analyze case-referent data even if no external information is available to allow estimation of risks in the source population. Prentice (1976) used Cornfield's (1951) classic sampling model when he presented a binary logistic regression for case-referent data. The outcome parameter was the probability of having been exposed to a risk factor, and the illness status was entered as an explanatory variable in the regression equation. Although the causal relation was inverted in this model, it allowed the estimation of the OR as an exponential function of the model coefficients.

Epidemiologists have somewhat neglected the sensitivity of the maximum likelihood parameter estimates to model misspecification. If one posits a logistic model for the disease rates in the population that depends linearly on the determinants, but the true model form is quadratic, the regression coefficients estimated from the case-referent sample may differ markedly from the coefficients that one would estimate from a cohort study of the same population (Scott and Wild 1986). For small or imbalanced data sets, and for highly stratified data, the asymptotic maximum likelihood methods are unreliable for parameter estimation. In these situations, the software package LogXact can be used to compute exact logistic regression (Cytel Software Corporation 1999).

Cox's (1972) regression model, or the proportional hazards model, is based on the notion of partial likelihood (Cox 1975), and it is applicable to the analysis of survival data or event history data (Clayton 1988). The model is semi-parametric in that the hazard or momentary risk depends on time non-parametrically but the risk ratio is a parametric function of the covariates. Due to computational difficulties, the method was seldom used in the 1970s, but today it is applied generally. In Finland, Hakulinen (Hakulinen 1982, Hakulinen and Abeywickrama 1985) has developed analytic methods and computing procedures of survival analysis for studies on cancer epidemiology.

The log-linear (or exponential) risk model (Bishop et al. 1975) is a most effective approach for the analysis of count (or aggregate) data, and especially for studying interdependencies (Nurminen et al. 1982). For example, the model was fitted to the 15-year follow-up data of a cohort of Finnish workers exposed to carbon disulfide. This intervention study was designed and conducted by Hernberg (Hernberg et al. 1970). For the analysis, the follow-up period was divided into five subperiods in which the deaths from coronary heart disease were assumed to occur according to time-homogeneous Poisson processes. A piece-wise exponential model was fitted to the data; it indicated that the declining trend in mortality reflected the reduced levels of carbon disulfide exposure (Nurminen and Hernberg 1985).

Modern ('smooth') regression methods (Venables and Ripley 1999), such as additive models and scatterplot smoothers as well as projection pursuit regression, are powerful tools, for example, to detect nonlinearities in the data. However, they are computer-intensive, and the distribution theory does not give them much support. One should be cautious in applying these new methods, because it is very easy to over-fit models and over-interpret features of the data.

Multilevel modeling and hiearchical regression in epidemiology

Many statistical problems involve multiple parameters. There is need to reflect on the complexity of observed data and different patterns of heterogeneity, dependence, mismeasurement, etc. In epidemiology, multiple parameters are involved in analyses of: 'subject effect' in growth curves; 'frailty' in correlated or familial survival data; 'center effect' in multi-center studies; relative risks for a disease outcome in different areas or time periods; and relative risks for different tumor sites in toxicological studies. In occupational and clinical epidemiology, the analysis of longitudinal data or repeated measurements (Nurminen and Hytönen 2000) involves multiparametric modeling. Environmental epidemiology uses methods such as ecologic analysis, time-series analysis, and quantitative risk assessment, for linking data on the environment and health (Nurminen 2000). The relations are often complex and fraught with uncertainties. When a model has many parameters, we may consider them as a sample from some distribution. In this way, we model the parameters with another set of ('hyper-') parameters and build a model with different levels of hierarchy.

Greenland (2000a) argues that regression models with random coefficients offer a more scientifically defensible framework for epidemiologic analysis than the fixed-effects models now prevalent in epidemiology. The data often consist of multiple levels that have effects on the results. For example, in the study of disease outcomes, there are patients involved (level 1) who are treated by physicians (level 2) who, in turn, are working in different hospitals (level 3). The characteristics of each may influence health outcomes, such as the patient's level of education, the physician's practice style, and the hospital's level of technical equipment. Sometimes characteristics from different levels interact to produce a certain outcome.

Conventional statistical methods assume that the observations are independent of each other. In a hierarchy, the observations of the same subpopulation are usually alike in some respects, that is, the data are correlated. The so-called multilevel models or hierarchical regressions offer a more realistic and flexible description of the factors that create uncertainty than do fixed-effects models. An advantage of regression with random coefficients is that it can be used to solve the often-encountered problem of under-identification of causal effects in epidemiologic data. The approach is to stochastically constrain the analysis by imposing a distribution on some parameters. The analysis of the data can be done on an individual level or on a higher aggregate level, depending on the objective of the study. Theoretically, multilevel modeling is well suited to analyzing the influence of macrolevel contexts on microlevel behavior. Statistically, hierarchical analysis solves the problems that occur when we either aggregate the data to one, higher level (loss of information) or disaggregate the data to the lower level (overestimated precision).

Multilevel modeling increases greatly the statistical precision and robustness of data analysis. A hierarchical regression is modeled in two stages. First, an ordinary (e.g. logistic) regression model is written for the effects of fixed parameters. In the second stage, a random distribution is defined for some of the parameters of the first-stage model, for example, to describe the presence of error in exposure measurement. By combining the stage 1 and stage 2 models one gets a mixed model with coefficients both for fixed-effects and for random-effects.

Multilevel models can also be estimated using a Bayesian analysis (Richardson and Gilks 1993).The Bayesian approach provides a natural framework to handle models of almost arbitrary complexity. There are many applied situations in which multilevel models and Bayesian estimation methods allow better analyses than more traditional methods. In a way, the hierarchical approach unifies the traditional and Bayesian methods (Greenland 2000b).

Statistical inference in epidemiology

Conventional ('frequentist') statisticians think of probabilities as frequencies observed in the long run of repeated experiments. Epidemiologic studies generally concentrate on nonexperimental research into causality in the health field. In these types of studies, there is little or no need for randomization and random sampling in the selection of the study base. Randomization, however, is needed for causal inferences from conventional statistics. In the context of nonrandomized studies, Greenland (1990) has questioned the interpretation of probabilistic measures such as a p-value (Nurminen 1997a) and a confidence interval as summaries of the variability of the results stemming from unidentified confounders (Nurminen 1997b). An unknown distribution of confounders cannot safely be assumed to be equivalent to what randomization would produce. According to this view, these statistics are merely rough descriptors of data variability. Causal inference should concern: (i) the search for explanations for patterns recoqnized in the data by statistical methods, and, (ii) criticism of the proposed theories about the physical mechanisms that generated the data (Greenland 1993).

There are many alternative methods available for the description of variation in the data. In sensitivity analysis, the data can be modeled, for example, by leaving out some observations from the analysis and observing how much the results would change. In influence analysis, the model can be altered systematically to see whether the results are prone to change or whether they remain fairly similar. The uncertainty in the model specification can be reduced by the use of robust procedures. In random-effects modeling, one can enter variables into the model to stochastically limit, for example, the effects of measurement error. Semiparametric methods such as the generalized additive model (Hastie and

Tibshirani 1990) allow epidemiologists to visualize their data in novel ways, especially in the presence of nonlinear associations, leading to new insight and new hypotheses.

In Bayesian statistics, probability is used as a fundamental measure of uncertainty. Probabilities are interpreted as subjective beliefs which are modified (according to the Bayes rule) as new information accumulates. Technically, prior information is convoluted with the data at hand, and the result is presented in the form of a posterior distribution. Bayesian methods were previously seldom used by epidemiologists, who felt that the specification of the prior distribution was difficult. Even today few epidemiologists apply Bayesian methods. The finding of Bayesian solutions presents a challenge even in the case of simple problems. The exact Bayesian analysis of the comparative epidemiologic parameters RD, RR, and OR in a two-by-two table furnishes an example (Nurminen and Mutanen 1987).

The empirical Bayesian analysis is an alternative approach in which a prior distribution is most easily specified to be reciprocally related to the distribution of the data, and the parameters of the conjugate distribution are estimated from the data. The Bayesian framework offers a possibility for the hierarchical modeling of case-referent studies that can be extended to deal with any number of categorial or discretized continuous exposure variables, and to identify suitable prior distributions (Seaman and Richardson 2001). One can also perform a semi-Bayesian analysis by specifying some features of the prior distribution from existing knowledge and estimating given parameters from the data. In an epidemiologic application, for example, one can insert background information on relative risks into conjugate prior distributions (Greenland 2001).

In a Bayesian analysis, to produce exact results from the posterior distribution, it is often necessary to evaluate integrals over large-dimensional parameter spaces, and this can be computationally intractable. However, new computer programs such as AD Model Builder (Otter Research Ltd. 2000) provides feasible approximations to these integrals in the form of a profile likelihood. The profile likelihood can then be used to estimate extreme values such as the tails of Bayesian credible intervals. The program also supports the Markov Chain Monte Carlo (MCMC) simulation for an 'exact' Bayesian analysis. The development of powerful MCMC methods has meant that computational issues are no longer a major obstacle to Bayesian inference. But model convergence must nevertheless be checked carefully, for example, when using the BUGS (Bayesian Inference Using Giggs Sampling) program (fttp:/www.mrc-bsu.cam.ac.uk/bugs).

Epidemiology is approaching its limits because there is not enough variation, for example, in many life-style factors within the studied population to observe RRs of sufficient magnitude to overcome the measurement error and confounding biases (Taubes 1995). Effective solutions may be seen in randomized intervention programs, but these can be too costly and difficult to design in nonexperimental settings. In purely observational studies one can make better inferences by thinking about the causal relations among variables and by integrating causal structures into the data analysis. For example, if one wants to estimate the probability of causation for individuals in cases of liability, it is important to explicitly specify the underlying biologic model that has been assumed (Beyea and Greenland 1999). Such methods include instrumental variable analysis used in econometrics (Greenland 2000c), Rubin's (1974) causal model, Robin's (1986) G-computation algorithm for longitudinal data, and Pearl's procedures for causal reasoning based on directed, acyclic graphs (Pearl 1995, Greenland et al. 1999).

Closing remarks

Biostatisticians have contributed for a long time to the conceptualization, development, and successful usage of epidemiologic methods for the study of disease causation and prevention. The International Biometric Society was established already in 1947 (Armitage and David 1996). The Finnish Biostatistical Society was founded 40 years later in 1987, with Professors Seppo Sarna and Juha Alho, and myself as the founding members. The activity of the Biostatistical Society has fostered the application of statistical and mathematical methods in epidemiology, medicine and biology in Finland. Especially the work carried out at the Research Division of Biometry at the Rolf Nevanlinna Institute under the leadership of Professor Elja Arjas in the application of the Bayesian statistical inference and MCMC methods deserves mentioning.

There are, nevertheless, two current sources of concern (Armitage 2001). The first is the apparently irreversible over-mathematization of biostatistics. This trend is reflected in journals such as Biometrika and Biometrics that initially set out to be comprehensible to the less academic practitioners. Newer journals such as Statistics in Medicine and Biostatistics are more application-oriented. The second concern is that the evolution of biostatistics -- which relies increasingly on important contributions from computing -- can lead to the over-emphasis of the role of theory at the expense of practice in the teaching of epidemiologic methods for researchers. Although theory may be the best guide in practice, the stress in the application of biostatistics should be on the prefix bio.

Acknowledgments

I wish to thank Tuula Nurminen for her methodologic support in the preparation of this article and Terttu Kaustia for the English language revision.

References

Alho JM. Logistic regression in capture-recapture models. Biometrics 1990;46:623-5.

Alho JM. On prevalence, incidence, and duration in general stable populations. Biometrics 1992;48:587-92.

Andersen EB. Conditional inference and models for measuring. Copenhagen: Mental-hygienjnisk Forlag, 1973.

Armitage P, David HA, editors. Advances in Biometry. 50 years of the International Biometric Society. New York, N.Y.: Wiley, 1996.

Armitage P. Theory and practice in medical statistics. Statistics in Medicine 2001;20:2537-48.

Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occupational and Environmental Medicine 1998;55:651-6.

Bartlett MS.The information available in small samples. Proc. Camb. Phil. Soc. 1936;32;560-6.

Beal SL. Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples. Biometrics 1987;43:941-50.

Beyea J, Greenland S. The importance of specifying the underlying biologic model in estimating the probability of causation. Health Physics 1999;76:269-74.

Bishop YMM, Fienberg SE, Holland PW. Discrete multivariate analysis: Theory and practice. Cambridge, MA: The MIT Press, 1975.

Breslow NE. Statistics in epidemiology: The case-control study. Journal of the American Statistical Association 1996;91:14-28.

Breslow NE, Cain KC. Logistic regression for two-stage case-control data. Biometrika 1988;75:11-20.

Cain KC, Breslow NE. Logistic regression and efficient design for two-stage studies American Journal of Epidemiology 1988;128:1198-206.

Chan ISF, Zhang Z. Test-based exact confidence intervals for the difference of two binomial proportions. Biometrics 1999;55:1202-9.

Chiang CL. Stochastic processes in biostatistics. New York: Wiley, 1968.

Clayton D. The analysis of event history data: A review of progress and outstanding problems. Statistics in Medicine 1988;7:819-41.

Clayton D, Hills M. Statistical models in epidemiology. New York, NY: Oxford University Press, 1993.

Cochran WG. Some methods for strengthening the common X^2 tests. Biometrics 1954;10:417-51.

Cochran. WG. Sampling techniques. 2ND ED. New York, NY: Wiley, 1963.

Cornfield J. A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast, and cervix. Journal of the National Cancer Institute1951;11:1269-75.

Cornfield J. A statistical problem arising from retrospective studies. Proceedings of the Third Berkeley Symposium on Mathematical Statistics. Berkeley, CA: Berkeley University Press, 1956;4:135-48.

Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society B 1972;34:187-220.

Cox DR. Partial likelihood. Biometrika 1975;62:269-76.

Cytel Software Corporation. LogXact-4 for Windows, Version 4.1. Statistical Software for Exact Nonparametric Inference. User Manual. Cambridge, MA: Cytel Software Corporation, 1999.

Farrington CP, Manning G. Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference on non-unity relative risk. Statistics in Medicine 1990;9:1447-1454.

Fieller JL. A fundamental formula in the statistics of biological assay, and some applications. Q J Pharm 1944;17:117-23.

Fisher RA. On the mathematical foundations of theoretical statistics. Philosophical transactions of the Royal Society of London, Series A 1922;222:309-368.

Fisher RA. The logic of inductive inference. Journal of the Royal Statistical Society 1935;98:39-54.

Fisher RA. Confidence limits for a cross-product ratio. Australian Journal of Statistics1962;4:41.

Fisher RA. Statistical methods and scientific inference. 3rd ed. Edinburgh: Oliver and Boyd, 1973.

Gart JJ, Nam J. Approximate interval estimation of the ratio of binomial parameters and correction for skewness. Biometrics 1988;4:323-8.

Greenland S. Randomization, statistics, and causal inference. Epidemiology 1990;1;421-8.

Greenland S. Summarization, smoothing, and inference in epidemiologic analysis. Scand J Soc Med 1993;21;227-32.

Greenland S. Introduction to regression models. Chapter 20. In: Rothman KJ, Greenland S, editors. Modern epidemiology, 2nd edition. Philadelphia, PA: Lippincott-Raven, 1998a, pp. 359-99.

Greenland S. Introduction to regression modeling. Chapter 21. In: Rothman KJ, Greenland S, editors. Modern epidemiology, 2nd edition. Philadelphia, PA: Lippincott-Raven, 1998b, pp. 401-32.

Greenland S. A unified approach to the analysis of case-distribution (case-only) studies. Statistics in Medicine 1999;18:1-15.

Greenland S. When should epidemiologic regressions use random coefficients? Biometrics 2000a;56:915-21.

Greenland S. Principles of multilevel modelling. International Journal of Epidemiology 2000b;29:158-67.

Greenland S. An introduction to instrumental variables for epidemiologists. International Journal of Epidemiology 2000c;29:722-9.

Greenland S. Putting background information about relative risks into conjugate prior distributions. Biometrics 2001;57:663-70.

Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999;10:37-48.

Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. Biometrics 1982;38:933-942.

Hakulinen T and Abeywickrama K. A computer program package for relative survival analysis. Computer. Programs. Biometrics 1985;19:197-207.

Hastie TJ, Tibshirani RJ. Generalized additive models. New York: Chapman and Hall, 1990.

Hernberg S. Introduction to occupational epidemiology. Chelsea, MI: Lewis Publishers, 1992.

Hernberg S, Partanen T, Nordman C-H, Sumari P. Coronary heart disease among workers exposed to carbon disulphide. British Journal of Industrial Medicine 1970;27:313-25.

Kalbfliesh JD, Sprott DA. Application of likelihood methods to models involving large numbers of parameters (with discussion). J Roy Satist Soc B 1970;32:175-208.

Keiding N. Age-specific incidence and prevalence: a statistical perspective. Journal of the Royal Statistical Society, Series A 1991;154:371-412.

Langholtz B, Goldstein L. Conditional logistic analysis of case-control studies with complex sampling. Biostatistics 2001;2:63-84.

Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute 1959;2:719-48.

McCullagh P, Nelder JA. Generalized Linear Models. London: Chapman and Hall, 1983.

Miettinen OS. Estimability and estimation in case-referent studies. American Journal of Epidemiology 1976;103:226-35.

Miettinen OS. Design options in epidemiologic research. Scandinavian Journal of Work Environment and Health 1982,8(suppl. 1):7-14.

Miettinen OS. Theoretical epidemiology. Principles of occurrence research in medicine. New York, NY: Wiley, 1985.

Miettinen O, Nurminen M. Comparative analysis of two rates. Statistics in Medicine 1985;4:213-26.

Nurminen M. A study of the mortality of workers in an anthophyllite asbestos factory in Finland. Work-Environment-Health 1972;9:112-8.

Nurminen M. Asymptotic efficiency of general noniterative estimators of the common relative risk. Biometrika 1981;68:525-30.

Nurminen M. Confidence intervals for the difference and ratio of two binomial proportions. Biometrics 1986;42:675-6.

Nurminen M. Assessment of excess risks in case-base studies. Journal of Clinical Epidemiology 1992;45:1081-92.

Nurminen M. Statistical significance — A misconstrued notion in medical research. Scandinavian Journal of Work, Environment and Health 1997a;23:232-5.

Nurminen M. On the epidemiologic notion of confounding and confounder identification. Scandinavian Journal of Work Environment and Health 1997b;23:64-68.

Nurminen M. Linking environment and health data: statistical and epidemiological issues. In: Corvalán C, Briggs D, Zielhuis G, editors. Decision-Making in Environmental Health. From Evidence to Action. London: E & FN Spon, 2000, 103-31.

Nurminen M, Hernberg S. Effects of intervention on cardiovascular mortality among workers exposed to carbon disulphide. A 15-year follow-up. British Journal of Industrial Medicine 1985;42:32-5.

Nurminen M, Hytönen M, Sala E. Modelling the reproducibility of acoustic rhinometry. Statistics in Medicine 2000;19:1179-89.

Nurminen M, Miettinen O. Confidence intervals for the ratio of the parameters of two independent binomials. Biometrics 1990;46:269-72.

Nurminen M, Mutanen P. Exact Bayesian analysis of two proportions. Scandinavian Journal of Statistics 1987;14:67-77.

Nurminen M, Mutanen P, Tolonen M, Hernberg S. Quantitated effects of carbon disulfide exposure, elevated blood pressure and aging on coronary mortality. American Journal of Epidemiology 1982;115:107-18.

Nurminen M, Nurminen T. Accuracy of asymptotic interval estimation methods for comparing two risks. Biometrical Journal 1990;32:195-205.

Oakes D. Biometrika Centenary: Survival Analysis. Biometrika 2001;88:99-142.

Otter Research Ltd. P.O. Box 2040 Sidney, B.C., V8L 3S3 Canada. AD Model Builder. Sidney, B.C.: Otter Research Ltd, 2000.

Pearl J. Causal diagrams for empirical research. Biometrika 1995;82:669-709.

Prentice R. The use of logistic model in retrospective studies. Biometrics 1976;32:599-605.

Richardson S, Gilks WR. A Bayesian approach to measurement error problems in epidemiology using conditional independence models. American Journal of Epidemiology 1993;138:430-42.

Robins JM. A new approach to causal inference in mortality studies with sustained exposure period — application to control of the healthy worker survivor effect. Math Model 1986;7:1393-512.

Robins J, Breslow N, Greenland S. Estimators of the Mantel-Haenszel variance consistent in both sparse-data and large-strata limiting models. Biometrics 1986;42:311-23.

Rothman KJ. Modern epidemiology. Boston: Little, Brown, 1986.

Rubin DR. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 1974;66:688-701.

Scott AJ, Wild CJ. Fitting logistic models under case-control or choice based sampling. Journal of the Royal Statistical Society, Ser B 1986;48:170-82.

Seaman SR, Richardson S. Bayesian analysis of case-control studies with categorical covariates. Biometrika 2001;88:1073-88.

Truett J, Cornfield J, Kennell W. A multivariate analysis of the risk of coronary heart disease in Framingham. Journal of Chronic Diseases 1967;20:511-24.

Taubes G. Epidemiology faces its limits. Science 1995;269:164-9.

Venables WN, Ripley BD. Modern applied statistics with S-PLUS, 3rd edition. Chapter 9. Smooth regression. New York: Springer, 1999, pp. 281-302.

Walker SH, Duncan DB. Estimation of the probability of an event as a function of several independent variables. Biometrika 1967;54:167-79.

Wedderburn RWM. Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. Biometrika 1974;61:439-47.

Testimonial epidemiology â€“ giving depositions and testifying in court or in legislative hearings on the state of epidemiologic evidence on a matter of dispute. Social epidemiology â€“ interpersonal and community-level factors influencing health at the population level. Global epidemiology â€“ assessing the effects of human activity on the ecosystem that supports life on Earth. @article{Zhang2004EvolutionOE, title={Evolution of epidemiologic methods and concepts in selected textbooks of the 20th century}, author={Fang F. Zhang and Desireï¿½ C. Michaels and B. Mathema and S. Kauchali and A. Chatterjee and D. Ferris and T. James and J. Knight and M. Dounel and Hebatullah O. Tawfik and Janet A. Frohlich and Li Kuang and.Â By reviewing eight epidemiology textbooks published over the course of a century, we have attempted to trace the evolution of five epidemiologic concepts and methods: study design (cohort studies and case-control studies), confounding, bias, interaction and causal inference. Overall, these eight textbooks can be grouped into three generations. Section 2: Historical Evolution of Epidemiology. Although epidemiology as a discipline has blossomed since World War II, epidemiologic thinking has been traced from Hippocrates through John Graunt, William Farr, John Snow, and others. The contributions of some of these early and more recent thinkers are described below.(5). Circa 400 B.C. Epidemiologyâ€™s roots are nearly 2,500 years old. Hippocrates attempted to explain disease occurrence from a rational rather than a supernatural viewpoint. In his essay entitled â€œOn Airs, Waters, and Places,â€ Hippocrates suggested that environmental and host f Theoretical Perspectives. Learning Objectives. By the end of this section, you will be able to: Explain what sociological theories are and how they are used. Understand the similarities and differences between structural functionalism, conflict theory, and symbolic interactionism. Sociologists develop theories to explain social occurrences such as protest rallies.Â Just as structural functionalism was criticized for focusing too much on the stability of societies, conflict theory has been criticized because it tends to focus on conflict to the exclusion of recognizing stability. Many social structures are extremely stable or have gradually progressed over time rather than changing abruptly as conflict theory would suggest. Symbolic Interactionist Theory. epidemiological transition. stems from epidemiology: branch of medical science concerned with incidence/distribution/control of disease. stage 1. Pestilence and Famine: infectious and parasitic diseases are principal causes of human deaths as well as accidents/attacks by animals and other humans. stage 2.Â Evolution, Poverty, Increased Connections: infectious diseases evolve and return or emerge, spread easier in poor areas with unsanitary conditions, improved transportation allows for spread of disease. This set is often saved in the same folder as AP Human Geography-Types of Agricultureâ€¦ 13 terms.