MATHEMATISCHES FORSCHUNGSINSTITUT OBERWOLFACH

Tagungsbericht 10/1993

Medical Statistics: Statistical Methods in Risk Assessment

28.2. - 6.3.1993

The conference was organised by Max P. Baur (Bonn) and Robert C. Elston (New Orleans). There were 35 participants from Denmark, France, Germany, Great Britain and the USA.

The 32 presentations on the topic of statistical methodology in risk assessment dealt with the two major categories of genetical risks and environmental risks; some of the approaches presented included the treatment of gene × environment interaction in risk assessment.

From the viewpoint of statistical methodology, the specific problems of modelling genetic information and estimating the unknown parameters was stressed. The specific problems of family data structures, the corresponding likelihood formulations and the analytic methodologies of segregation analysis, linkage analysis and association analysis were presented and related to techniques of epidemiologic risk assessment by way of case control studies, cohort studies etc. The areas of application covered cancer, cardiovascular diseases and autoimmune diseases at the population level as well as for individual risk assessment in genetical counseling.

Gefördert durch
DFG Deutsche Forschungsgemeinschaft

Presentations:

Introduction to genetic epidemiology.

Max P. Baur

Genetic epidemiology centers on the clustering of diseases in families. Genetic factors transmitted from parents to children that increase the risk for the given disease have to be established. Consequently the datastructure to be analysed consists of indivduals who are related (families) with their disease status and additional information about known genetic markers. Based on the general likelihood formulation incorporating the genotype- phenotype relation, penetrance, population frequency and transition probabilities, segregation analysis tries to differentiate different modes of inheritance, linkage analysis tests for cosegregation of the disease with regard to genetic markers and maps the disease gene within the genome, and association analysis tests for independence of marker alleles among diseased individuals versus controls, thus pointing to possible linkage. Disease modelling, disease mapping and corresponding parameter estimation is the basis of individual risk assessment in prenatal diagnosis and genetical counseling.

Risk assessment in epidemiology.

Jürgen Wahrendorf

A brief overview is given about the different steps of risk assessment in epidemiology. The different types of epidemiologic studies contributing to this are reviewed and essential definitions of risk and related statistical procedures are provided. Empirical and logical justification to consider relative risks as effect measures in epidemiologic research are outlined.

Likelihood 1912-1992.

Anthony W.F. Edwards

The statistical concept of likelihood was isolated and named by R.A.Fisher in 1921, although he had in fact advocated its use as a measure 'suitable to compare point with point, but incapable of being interpreted as a probability distribution over a region, or of giving any estimate of absolute probability' in 1912 (whilst still a Cambridge undergraduate).

After a brief mention of some of the precursors of likelihood, the concept's progress from 1912 to the present is traced, with special reference to its

Gefördert durch DFG Deutsche Forschungsgemeinschaft

surprising appearance in repeated- sampling arguments in the 1930's. Can likelihood continue its separate existence, or is it destined to become merely an intermediary in the neo-Bayesian revolution? What problems does its non-Bayesian use pose?

Regressive logistic models for assessing familial aggregation of oesophageal cancer.

Maria Blettner, Heiko Becher

Oesophageal cancers have been shown to aggregate within families. To investigate the family aggregation and the genetic mechanism further, detailed data have been collected of incident cancer cases (n=250) in a high-risk area in China, including the number of family members of three successive generations, their cancer incidence, age at diagnosis and causes of deaths.

In this paper we describe the use of regressive logistic models (Bonney, 1986) for analyzing patterns of familial aggregation. By modeling the dependency of the relatives' phenotypes (oesophageal cancer) on the parents' phenotypes, their age of onset and other covariables this methods yields adjusted odds ratios that quantify familial aggregation. Estimated standard errors are obtained by taking into account possible correlations among family members following the ideas of Liang et al. (1992).

The regressive models showed that among siblings of the index case the risk was significantly elevated if the father or mother had been diagnosed with oesophageal cancer (after adjusting for age and sex). It was also found that the familial aggregation was higher if the probands (siblings) were younger than 60 years compared to those older than 60 years at diagnosis.

The usage of regressive models allows one to investigate complex patterns of family dependency and to investigate gene and environment interactions in complex pedigrees. However, several methodological problems, such as dealing with missing values and ascertainment bias, deserve further research. Some of these issues are discussed using our data and results from a small simulation study.

Bonney G.E. (1986) Regressive logistic models for familial disease and other binary traits. Biometrics, 42, 611-625.

Liang, K-Y., Zeger S.L., Qaqish B. (1992) Multivariate Regression Analysis for Categorical Data. J.R. Statist. Soc. B, 54, 3-40.

Retrospective assessment of exposure in a case-control study on lung cancer: Comparison of methods.

K.-H. Jöckel

The retrospective exposure assessment turns out to be one crucial point in the analysis of case-control studies. In the first part of the talk the exposure assessment of asbestos in a lung cancer study on occupational risk factors in Northern Germany is considered to some extent. The result of using detailed supplementary questionnaires (SQ) is compared to the application of two different job-exposure matrices (JEM) to job history. The study includes data from 391 male incident lung cancer cases and the same number of controls from the general population.

The comparison of the methods is achieved on the basis of individuals and with respect to the ability of the instruments to detect a relationship between exposure and disease (given that asbestos is carcinogenic). All three methods are able to detect such a relationship. However, on the basis of individual comparisons the relative merits are seen. So it is recommended to try to combine the methods in an efficient way. As an example, the performance of a modified dual assessment approach is investigated, suggesting an improved consistency of the exposure- disease relationship. However, additional methodological research is needed in this area.

The second part of the talk presents a preliminary analysis of the available data on lung cancer of study subjects and their parents in a case-control setting.

An Alternative to Gail & Simon's Approach to Testing for Qualitative Interactions Between Risk Factors and Treatment Effects.

Stefan Wellek

As did Gail & Simon (1985) we assume that the effect of the therapy under investigation on the prognosis of the ith risk group is given by a real-valued parameter $\theta_i (i = 1, \ldots, k)$, say, for which there exists an asymptotically normal estimator $\hat{\theta}_i$. Furthermore, we assume that consistent estimators $\hat{\sigma}_i^2$ for the asymptotic variances σ_i^2 of the $\hat{\theta}_i$ are available.

According to Peto (1982) a qualitative interaction between the risk factor and the treatment effects is defined to be present if at least two of the θ_i are of opposite sign. If we equate the variances of the $\hat{\theta}_i$ with their consistent estimators, assessing the data with regard to a possible qualitative interaction leads to the problem of testing $H_0 : \mu_{(1)} \geq 0 \vee \mu_{(k)} \leq 0$ versus $H_1 : \mu_{(1)} < 0 < \mu_{(k)}$, where $\mu_i = E(X_i)$, $X_i \sim \mathcal{N}(\mu_i, 1)$, $\mu_{(1)} = \min_{1 \leq i \leq k} \mu_i$,

Gefördert durch
Deutsche Forschungsgemeinschaft

$\mu_{(k)} = \max_{1 \le i \le k} \mu_i$, and the X_i are mutually independent.

Gail & Simon derived a maximum likelihood ratio test for this problem which rejects H_0 for large values of the statistic $\min(Q^-(\boldsymbol{X}), Q^+(\boldsymbol{X}))$, where $Q^-(\boldsymbol{X}) \equiv \sum_{i=1}^k X_i^2 \cdot I_{(-\infty,0)}(X_i)$, $Q^+(\boldsymbol{X}) \equiv \sum_{i=1}^k X_i^2 \cdot I_{(0,\infty)}(X_i)$. The objective of the present contribution is the derivation of an alternative test for (H_0, H_1) which uses the rejection region
$\{W(\boldsymbol{X}) > w_\alpha\} \equiv \{\min(-X_{(1)}, X_{(k)}) > w_\alpha\}$.

We show that this test has, in common with Gail & Simon's test, the property of exhibiting a Schur convex power function. For the critical constant w_α making $\{W(\boldsymbol{X}) > w_\alpha\}$ an exact size α rejection region, a simple explicit expression is provided. Simulation results are presented which show that the power functions of both tests are extremely different except for very small values of k. Typical alternatives against which the new test is much more powerful than Gail & Simon's test are of the form $(\mu_1, \ldots, \mu_k) = (-\zeta, \xi, \ldots, \xi)$, $\zeta \ge 2$, $\xi \ge \zeta$. A prototype of an alternative under which the reverse holds true, is seen to be given by $(\mu_1, \ldots, \mu_k) = (\underbrace{-\xi, \ldots, -\xi}_{[k/2]}, \underbrace{\xi, \ldots, \xi}_{k-[k/2]})$.

Gail, M., Simon, R. (1985) Testing for qualitative interactions between treatment effects and patient subsets. Biometrics 41, 361-72.

Peto, R. (1982) Statistical aspects of cancer trials. In Halnan, K.E. (ed.), Treatment of Cancer. London: Chapman and Hall, 867-71.

Estimation of toxikinetic parameters.

Wolfgang Urfer

Legal regulations of dangerous chemicals are based on estimates of risks. Toxic effects of chemicals vary from one animal species to the other. Hence, it is important to know about the chemical's toxicokinetics (uptake, distribution, elimination and metabolism) in the animals and their inter-species differences. Knowledge of quantitative metabolic parameters of different species facilitates a species-specific toxicological approach to the toxic effects observed. We discuss aspects of metabolism within living organism using deterministic compartmental analysis. Based on non-invasive toxicokinetic inhalation experiments with rats, mathematical approaches and non-linear regression analysis are applied to describe the metabolic characteristic of the biological system, and to estimate toxicokinetic parameters. Pre-existing data on the hydrocarbon propylene (propene) in Sprague-Dawley rats were used to determine the rates of uptake and exhalation in this species.

M. Becka, H.M. Bolt and W. Urfer (1992): Statistical analysis of toxicokinetic data by nonlinear regression. Archives of Toxicology, 66, 450-453.

Basic genetics for non-geneticists.

Thomas F. Wienker

This presentation is intended to be an introduction to basic principles and terminology for topics covered during the meeting. In addition, recent advances are presented which call for a statistical treatment and new methodology. In order to put phenomena and terminology into place a three- dimensional approach is taken: (i) no. of genes under consideration is 1,2,3 or many, (ii) gene(s) act in an individual, a pedigree, or a population, and, (iii) view is from a genotype or phenotype level. The concepts of penetrance, genetic heterogeneity (locus h.vs. allelic h.), and polymorphic variation (neutral vs. selective mutations) are explained and illustrated on examples. Special consideration is given to highly polymorphic marker systems, which play an increasing role in epidemiologic studies on a population basis (association studies), and in genetic linkage analysis and indirect genetic diagnosis, the principles of which are briefly touched (see contribution by J. Ott).

Now a fairly dense map of the human genome is available, and the different views and approaches, i.e. physical versus genetic (i.e. based on the analysis of meiotic recombination events) are stressed. Mapping strategies, methodology, measurements, and heuristic implications are quite different, and a lot of confusion may arise, if these viewpoints are not clearly distinguished.

Finally, the population genetic dynamics of highly polymorphic DNA segments composed of reiterated short motifs with a variable repeat number are put into focus. These "new" polymorphisms call in my view for a new population genetics, i.e. an extension of classical concepts into ones "far from equilibrium" - very much analogous to thermodynamics of living systems developed some decades ago.

Introduction to segregation analysis and ascertainment-correction.

Jon Stene

Initially the one locus, two allelic situation was considered. The possible types of offspring and their probabilities for the 9 different pairs of parental matings were presented, assuming Mendelian inheritance, both for the codominant and the recessive case. For the latter case ascertainment (or identification) of parental matings able to produce recessive children could only be done through the observation of the birth of at least one recessive child. The talk discussed different ways of selecting such family data through information about recessive children and the probability models for the number of recessive children in families in the selected data, given the selection procedure. The more general case, where the probability for an abnormal offspring was unknown and non-mendelian, was then considered and methods to select the most appropriate model and estimate the parameters were discussed.

Introduction to linkage analysis.

Jurg Ott

In the course of meiosis, homologous chromosomes pair up and form points of crossing over, where each crossover randomly involves one of the two strands (chromatids) of each homolog. When one crossover has occurred, two of the four gametes resulting from each meiosis are unaffected by the crossover while each of the remaining two gametes consist of a piece of chromosome originating from one homolog and a piece of chromosome originating from the other homolog. If the crossover took place between two polymorphic loci, its effect may be recognized as a recombination in an offspring who received part of a chromosome from one grandparent and another part from another grandparent. The relative frequency of a recombination occurring between two loci is called the recombination fraction, θ. Map distance, x, is defined as the expected number of crossovers per strand between two loci. Its unit of measurement is 1 map unit or cM (centi-Morgan). The generally observable quantity, θ, is a function of map distance, $x = f(\theta)$, where f is termed a map function. It follows from the chromosomal theory of recombination that $0 \leq \theta \leq \frac{1}{2}$, and this is also true for multiple crossovers. For small distances, 1 cM $\approx$ 1% recombination fraction. Pairs of loci for which $\theta < \frac{1}{2}$ are called genetically linked. The objects of linkage analysis are 1) to estimate θ and 2) to test $H_0 : \theta = \frac{1}{2}$ versus $H_1 < \frac{1}{2}$. Various complications may make linkage analysis difficult. For example, incomplete penetrance represents of loss of linkage informativeness, which may be compensated for with an increased sample size. For example, when penetrance is only 50%, a roughly threefold larger sample size is required than with full penetrance (J. Ott, Analysis of Human Genetic Linkage, 1991). Genetic heterogeneity (admixture of "linked" and "unlinked" families) may be handled by jointly estimating the admixture proportion α and the recombination fraction θ in the "linked" families. Genetic risks are naturally calculated as

$$P(g_i \mid x_1, \ldots, x_n) = \frac{P(g_i, x_1, \ldots, x_n)}{P(x_1, \ldots, x_n)}.$$

where g_i is the genotype of the i-th individual, the denominator of the fraction is just the pedigree likelihood, and the numerator is the likelihood assuming that individual i has genotype g_i. Multiple testing between a disease and m markers is not a problem for mendelian loci because the increased type I error is more than compensated for by an increased prior probability of linkage (Ott 1985). However, for (complex) traits with an unknown mode of inheritance, when linkage analysis is used as a tool to demonstrate presence of disease loci, multiple testing must be allowed for with a more stringent test criterion, for example, with a critical lod score of $3 + log_{10}(m)$ (Kidd and Ott, 1984): Over the whole genome at most $m \approx 100$ independent comparisons may be made.

Design for the global search of the human genome by linkage analysis.

Robert C. Elston

It is now feasible to locate an autosomal disease locus on the human genome by a linkage study in which the whole autosomal genome is spanned by equally spaced polymorphic markers. Increase in either the number of sampling units or the number of markers will increase the power of the study. This paper determines the optimal spacing between consecutive markers when the sampling units are pairs of affected relatives. This spacing is asymptotically independent of the desired significance level or power, but depends critically on the "risk ratio factor" of the disease locus being sought. A two-stage procedure, in which at the first stage more widely spaced markers and a larger significance level are used, followed at the second stage by more narrowly spaced markers around those significant at the first stage, leads to much more economical designs. A method is described to find the optimal such design on the assumption that the total cost is determined by a cost per person recruited into the study and a cost per marker assay performed.

Affected sib-pair tests and their relationship to lod score analysis.

Michael Knapp

Affected sib-pair tests have been proposed to detect linkage between a marker locus and a disease with unknown mode of inheritance. In its simplest form, these tests require a sample of affected sibs where for each sib-pair the number of marker alleles ibd can be determined unequivocally.

It will be shown that the mean test is for significance level $\alpha \leq 0.5$ equivalent to the classical lod score analysis, if a recessive mode of inheritance is assumed. Additionally, for sufficient small α the application of the proportion test is equivalent to lod score analysis for a dominant mode of inheritance and a very rare disease allele.

The computational requirements for calculating probabilities on pedigrees.

Alun Thomas

The computational problem of efficient calculation of

$$\sum_{v_1}\sum_{v_2}\cdots\sum_{v_m} f(v_1v_2\ldots v_m)$$

where

$$f(v_1v_2\ldots v_m) = \prod_{i=1}^{k} f(S_i)$$

and

$$S_i \subset \{v_1v_2\ldots v_n\}$$

is discussed in the context of probability calculations on family trees. Similar calculations arise in expert systems and other complex stochastic systems.

An associated optimalisation problem of triangulating a graph so as to minimize the size of the largest clique in the resulting graph is known to be NP-complete in general. Regularity results for graphs arising from genetic applications, including 4-colourability, are derived and used for heuristic search methods.

Construction of a new genetic map of the human pseudoautosomal region.

Christine Fischer

Human sex chromosomes X and Y contain two small homologous regions on both the short and long arm adjacent to the telomeres, the so- called pseudoautosomal regions. In both regions alleles have been shown to exchange between X and Y chromosomes and are inherited as if autosomal. On average one crossover event occurs at each male meiosis in contrast to a much less recombination intensity in female meiosis. Therefore on the basis of the CEPH reference panel of families a sex specific genetic map of the pseudoautosomal region was constructed via linkage techniques using 11 physically mapped markers. Estimates for the map length are 48.8 cM in males and

Gefördert durch
DFG Deutsche Forschungsgemeinschaft

5.2 cM in females. The comparison between male and female recombination rates reveals a telomere-adjacent interval where no difference in sex specific rates can be seen.

Cross prevalences and adjustment by age: Methodological problems exemplified by family studies.

Wolfgang Maier

Epidemiological and family studies observed an excess rate of cases with cooccurrence of lifetime diagnoses of two or more disorders (comorbidity). The analysis of the strength of association between two disorders and of the cosegregation in families encounters a series of methodological problems lacking a straitforward solution. The most critical issue is age adjustment: Two disorders with excess comorbidity might reveal very different age at onset curves, however the standard measure of the extent of comorbidity - odds ratio - does not allow age adjustment. A proband or a relative classified as having only one disorder might develop a second one in the future: however, it is difficult to build differential age at onset into the available analytic techniques. These problems are discussed by reference to psychiatric data sets.

Association in SLE families: Design, methods and results.

Susanne A. Seuchter

In order to elucidate the role of the Major Histocompatibility Complex (MHC) a multicenter family study was performed to investigate the associations of Systematic Lupus Erythematosus (SLE) with MHC markers in a caucasian population.

A major problem in the analysis of association of genetic markers and diseases is the selection of an appropriate control sample. Rubinstein et al have proposed the Haplotype Relative Risk (HRR) as a possible solution to this problem. To assure that disease and control sample are obtained from the same genetic population they count the parental haplotypes which were not transmitted to the affected child as a control sample. However, this requires that all parental genotypes are well known. This assumption cannot be satisfied in many family studies. Therefore, we introduce a method where families with partially unknown genotypes and pedigrees of any size ascertained through one affected individual can also be included. The method is based on estimating the haplotype frequency difference between the

Gefördert durch DFG Deutsche Forschungsgemeinschaft

transmitted and non- transmitted group. Results obtained by applying this method to our SLE families will be presented.

(supported by BMFT project 01ZU8804/0).

Transmission tests for linkage disequilibrium.

Richard Spielman

A population association has consistently been observed between insulin-dependent diabetes mellitus (IDDM) and the "class 1" alleles of the region of tandem repeat DNA (5' flanking polymorphism or 5'FP) adjacent to the insulin gene on chromosome 11p. This finding suggests the existence of a gene or genes in that region contributing to IDDM susceptibility. However, several studies that have sought to show linkage with IDDM by testing for cosegregation in affected sib pairs have failed to find evidence for linkage. As means for identifying genes for complex diseases, both the association and the affected sib pairs approaches have limitations. It is well known that population association between a disease and a genetic marker can arise as an artifact of population structure, even in the absence of linkage. On the other hand, linkage studies with modest numbers of affected sib pairs may fail to detect linkage, especially if there is linkage heterogeneity.

We consider an alternative method to test for linkage with a genetic marker when population association has been found. Using data from families with at least one affected child, the transmission of the associated marker allele from a heterozygous parent to an affected offspring is evaluated. This approach has been used by several investigators, but the statistical properties of the method as a test for linkage have not been investigated. In the present paper we describe the statistical basis for this "transmission test for linkage disequilibrium" (transmission/disequilibrium test or TDT). We then show the relationship of this test to tests of cosegregation based on the proportion of haplotypes or genes identical by descent in affected sibs. The TDT provides strong evidence for linkage between the 5'FP and susceptibility to IDDM.

The conclusions from this analysis apply in general to the study of disease associations, where genetic markers are usually closely linked to candidate genes. When a disease is found to be associated with such a marker, the TDT may detect linkage even when haplotype sharing tests do not.

Spielman RS, MCGinnis RE, Ewens WJ (1993 Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52: in press

Ott J (1989) Statistical properties of the haplotype relative risk. Genet Epidemiol 6: 127-130

Case-control studies of association: Likelihoods and power.

David Clayton

Case-control studies of association between a candidate allele (or an allele of a marker locus in a candidate region) can be carried out by selecting population controls or more closely matched controls. The closest matching is achieved by choosing family controls and in this case there is a choice of method of analysis. The likelihood which is used to calculate relative risk estimates and hypothesis tests may be

- a the conditional (hypergeometric) likelihood of conventional matched case-control analysis, or
- b a likelihood based on the 'genetic' model.

In either case, the score test for association is the difference between the observed number of cases carrying the candidate allele (or haplotype) and the expected number under the null-hypothesis. Analyses differ according to the information available for evaluating expected frequencies and the strength of their assumptions. Correspongdingly the power of the tests varies.

Power calculations were presented for a series of case control designs and analyses. Family control designs are inefficient as compared with population based case-control studies, but less so when both parents' genotypes are known. If only a single unaffected sibling control is avaiable, a 'genetic' likelihood is more efficient than a single matched case-control analysis, but usually not dramatically so.

The relationship between the likelihood-based analysis and 'haplotype relative risk' methods was briefly indicated.

Genetic risk factors in multifactorial diseases - HLA associated diseases - Alzheimer's disease.

Francoise Clerget-Darpoux

The MASC method was proposed to demonstrate and model the role of a candidate gene in a multifactorial disease. Information on markers of the candidate gene is needed and the method uses the simultaneous information on the marker association and segregation with the disease. It allows one to test the goodness of fit of various genetic models in an easy and economical way.

The method was applied to a sample of 416 Caucasian Insulin Dependent Diabetes Mellitus (IDDM) patients and their relatives. We showed that the model which best explains all the observations assumes a cis or trans complementation of two tightly linked genes within the HLA region, an additional maternal effect, as well as other familial factors.

The results obtained by the MASC method provide rules which can be used in molecular research. In particular, we showed that the HLA molecule corresponding to the complementation of Arg52(+) and Asp57(-), recently proposed as explaining the susceptibility to IDDM, does not account for the overall observations made on the HLA marker in IDDM patients and their relatives.

The MASC method may also be applied to evaluate the risk for relatives of an affected individual depending on the available information.

Epidemiology and genetic analysis of Celiac disease.

Thomas F. Wienker

Celiac disease is a gastrointestinal disorder. A dietary gluten sensitivity is at the basis of the associated mucosal pathology, and accordingly, it is designated gluten sensitive enteropathy (GSE), too. Genetic factors do play a role, and there is a strong association with HLA-haplotypes. Prevalence among school children is about 1 in 2000. Hence, GSE is a nosologically unique multifactorial disorder. Family studies (20 pedigrees, 31 patients) are performed in order to elucidate the role of genetic components. A number of neutral polymorphisms scattered over the genome and 8 loci presumably implicated in the mucosal immune response (T-cell receptor genes) have been typed. Preliminary results of association and linkage studies are presented for discussion.

Multi-stage models of environmental determinants of lung cancer risk.

Edward Lustbader

Radon and tobacco smoke are well established risk factors for lung cancer. Although important insights into the effects of radon and tobacco have been derived from the study of uranium miners, questions about how the pattern of exposure affects risk remain. For example, there are suggestions that fractionation of a given radon dose leads to increased lung cancer risk. Such questions are difficult to assess using conventional methods of analysis. Moreover, additional difficulties arise when studying exposure patterns to more than one agent. The main purpose of this presentation is to show that biologically based models of carcinogenesis can easily incorporate age related

Gefördert durch DFG Deutsche Forschungsgemeinschaft

exposure patterns. Further, the parameters of such models are interpretable in biological terms and afford some insight into the mechanisms of action of the agents.

Quantitative risk assessment with cumulative damage models in cancer epidemiology.

Nikolaus Becker

Cumulative damage models (CD models) have been developed in reliability theory to get a macroscopic description of environment-induced wear-out processes of technical devices in order to find optimal inspection and maintenance strategies. Applied to cancer occurrence in human populations, these models reproduce the bird's eye view of epidemiology to the etiology of diseases in as far as they treat the human host as a kind of "black box" which is subject to environment-driven carcinogenic damages, whereby the precise biologic nature of these damages remains unobservable. The precise mathematical definition and some qualitative properties of these models have been introduced at a previous meeting on statistical methods in cancer research in Oberwolfach.

In the present paper, the application to quantitative risk assessment is demonstrated by two examples. The first is an occupational cancer study among stainless steel welders. ML-estimation has been carried out on the basis of individual data. Goodness-of-fit has been tested by application of a method which has been developed by econometricians but has not yet been used in biostatistics. The second example is the well-known British physicians' study. The fits of three different models (multistage, Moolgavkar-Knudsen, CD) have been compared. It turns out that all models fit the data comparably well despite the rather different model assumptions (multistage: several cellular stages, no growth; Moolgavkar-Knudsen: two cellular stages and growth; CD: wear-out, no biologic specification). This underlines the expectation that a simple wear-out mechanism might be sufficient to fit epidemiologic data in order to carry out quantitative risk assessment. The closeness of the CD model to the epidemiologic approach to disease occurrence guarantees a straightforward parameterization in terms of onset, cessation and intensity of exposure and, thus, easily allows risk extrapolations or application to, e.g., unit risk assessment.

Assessment of quantitative exposure variables.

Martin Schumacher

In epidemiology different strategies can be applied for analyzing the effect of a quantitatively measured exposure variables on the risk of developing a certain disease. The most common approach is to dichotomize exposure into present or absent or to define cutpoints reflecting different levels of exposure to analyze a dose response relationship. A less frequently applied approach consists of assuming a special functional shape of the effect of the exposure on the log-odds-ratio or log-relative risk of developing the disease, and exposure can then be modelled as a continuous covariate in a regression model. A variety of rules for classifying exposure into two or more categories are available which range from a-priori selected cutpoints to data-orientated rules. A rather common approach is to select a cutpoint for which an effect - or the most pronounced effect - of the exposure variable on the outcome is observed. This approach for selection of cutpoints can be labelled 'data and outcome orientated' and some adjustment is required to qualify the final result.

We propose a method for adjustment of results derived by variation of the cutpoint on a specified selection interval. The method should be applied to correct the P-value if the cutpoint to define different levels of exposure is selected in a way that the measure of difference between the two groups of cases and controls, such as the odds ratio or relative risk, is maximized.

The various strategies are illustrated with a case-control study on the association between exposure to magnetic fields and the risk of cancer in children which has been conducted recently in Denmark. A clinical study on the role of S-phase fraction for the prognosis of breast cancer patients serves as a second example in order to illustrate that similar problems also arise in clinical research.

Multistate survival analysis.

Niels Keiding

Multistate survival analysis usually involves a series of detailed regression analyses describing transitions into various states. There is an often neglected need for the many detailed estimates resulting from such an analysis to be resynthesized into summary statements, such as prediction of various outcomes from specified patient histories, integrating the results of the specific analyses.

Arjas and Eerola recently proposed a framework for dynamic probabilistic causality which has calculation of such prediction statements as a central tool. We illustrate these procedures on data from a multicenter bone marrow transplantation study, with death while in remission and relapse as terminal

Gefördert durch
DFG Deutsche Forschungsgemeinschaft

events, and the time of recovery of the patient's platelets to a normal level and the onset of acute graft-versus-host disease as intermediate events, using Cox regression models throughout. Among the features illustrated by the resulting plots is a strong effect on death while in remission if the platelets do not recover during the first three months.

Statistical issues in risk assessment for developmental toxicity of PCBs.

Ursula Krämer

Polychlorinated Biphenyls (PCBs) are stable environmental contaminants. They are complex mixtures of congeners differing both in terms of the number and the position of chlorine atoms on the two rings. Among the broad spectrum of biological effects, developmental neurotoxicity appears to be a prominent feature of these chemical mixtures, as can be judged from experimental and epidemiological findings. Design and results of the studies in Michigan and North Carolina, from which a NOAEL (no observed adverse effect level) of 1-3 mg/kg fat is derived, are briefly summarized. The EPA (US Environmental Protection Agency) risk assessment, which leads to a reference dose for the daily intake, is critically revised. Design and goals of our planned study on PCB exposure and neurodevelopmental deficits are given. Genetic factors certainly determine neurobehavioral development and have to be considered as confounding factors.

Visual deficits as a risk for street accidents.

Jürgen Berger

To answer the question if more people with visual deficits are involved in street accidents, we are designing a case-control study. One problem is to get adequate controls. A population based sample looks adequate from the theoretical point of view but may be affected by two disadvantages. 1. Control persons may have had accidents before starting the study and do not report them. 2. For the detailed eye examination the persons have to come to the hospital and stay there for 2 to 4 hours. This could reduce the response-rate. Both facts could bias the results. Therefore we decided to formulate more specific hypothesis, e. g. accidents ignoring somebody's right of way are due to a reduced binocular visual field, and we intend to take persons involved in other road accidents as controls (however, excluding all alcohol induced accidents). Planing this study we used data of v. Hebenstreit to estimate the possible risk. The author has classified the kinds of 1920 accidents in two groups: accidents which may be related to visual deficits, such as

Gefördert durch
Deutsche Forschungsgemeinschaft

ignoring somebody's right of way, ignoring traffic signs, wrongly overtaking and those which may not be related to vision. Stratifying for age and using the Mantel-Haenszel procedure we estimated a relative risk of 6.7 (95% C.I.: 4.3 - 10.8) for persons with visual deficits to be involved in such kind of accidents.

Project: Genetic epidemiology of breast cancer.

Jenny Chang-Claude

The involvement of genetic factors in the etiology of breast cancer has long been recognized and family history has been found to be one of the most significant risk factors for breast cancer in epidemiologic studies. Results from many segregation analyses suggest that breast cancer is inherited through a rare autosomal dominant gene in some families. Close linkage of a breast cancer susceptibility gene (BRCA1) between 17q12-q21 with breast cancer, especially early onset breast cancer, has recently been reported. The Breast Cancer Linkage Consortium, a group of European and American investigators, confirmed this finding in 214 breast cancer families including 57 families with breast and ovarian cancer and provided evidence that the BRCA1 gene lies in an interval whose genetic length is estimated to be 8.3 cM in males and 18.0 cM in females.

The purpose of this project is to investigate the interaction between genetic and non-genetic factors in the etiology of breast cancer. The project includes a linkage study, first of all for the BRCA1 gene on chromosome 17q, and case control analyses, in which the long-term goal is to identify a gene or genes for breast cancer and to understand its role in the development of breast cancer in the general population.

A hereditary breast cancer family study has been initiated in collaboration with the Women's Clinic in Heidelberg to identify multiple case families throughout the country which are potentially informative for linkage analysis. Families eligible for inclusion should include at least three breast cancer cases within three generations in which at least two breast cancer cases are alive (in order to obtain the necessary biological material). In addition families with at least two ovarian cancer cases are also eligible since linkage with the BRCA1 gene was observed for both tumor types.

In the population-based genetic epidemiologic study of breast cancer newly diagnosed primary breast cancer patients (up to 50 years of age) will be compared with two control groups: a) non- diseased sister control, b) non-diseased population-based control persons from the two study regions 'Kurpfalz' and 'Freiburg' (randomly selected age-matched controls). Parents of breast cancer patients will be requested to provide a blood sample in order to be able to distinguish between a somatic mutation and a germ line mu-

Gefördert durch DFG Deutsche Forschungsgemeinschaft

tation if necessary. If a further case of breast cancer has been diagnosed in the family, all female family members will be requested to participate in the study.

Use of the regressive logistic models in linkage analysis.

Maria M. Martinez

The regressive logistic models introduced by Bonney (1984, 1986) are constructed by conditioning each individual's observations on those of his antecedents, using logistic regression for binary traits. The log of the odds of being affected is assumed to be a linear function of a major genotype, the phenotypes of antecedents, and other covariates. They allow simultaneous estimation of major-gene factors, residual covariation of unspecified origin, and environmental factors influencing the trait. Demenais (1992) has proposed an alternative formulation of the regressive logistic models. An underlying liability to the disease is assumed. The underlying liability is correlated among relatives. Then, affected persons have liabilities exceeding a threshold. The probability for a relative to be affected is expressed in terms of correlation coefficients among relative's liabilities. Under this formulation, the penetrances depend on both phenotypes and genotypes of antecedents.

We have investigated how the linkage detection is affected by the presence of residual correlation. We have considered the simple case where the effect of the disease gene is known, and computed for different disease gene models the recurrence risks predicted in the presence of residual familial correlation. Observations are affection status (Y) and phenotypes at a marker locus (M). The generalized penetrance function is: $P[Y, M \mid g, Y_R, \theta]$ where g is the underlying genotype, Y_R represents phenotypes of preceeding relatives, and θ is the genetic distance between the disease gene and the marker locus. Our results show that the effect of ignoring residual correlations on linkage detection depends on the effect of the disease gene. When the effect of the disease gene is low (genetic variance $<$10%) ignoring residual correlation leads to both a bias on θ and to a decrease in the maximum expected lod score. We have also shown that the two formulations of the regressive logistic models do not lead to similar recurrence risks and these recurrence risks depend on the ordering of the sibs.

Complex segregation analysis models.

Robert C. Elston

Complex segregation analysis does not assume that only a single genetic mating type is possible. The evolution of models for this kind of analysis is described and the corresponding likelihoods are presented. The transmission probability (generalized major gene) model allows for non-Mendelian transmission but assumes that all familial dependencies are due to the transmission of underlying discrete types. The mixed model allows for both multifactorial and major gene transmitted dependencies, but assumes that all such transmission is governed by Mendelian laws. The unified model combines these two models into a more general model that subsumes each as a special case. This effectively decreases both Type I and Type II errors in hypothesis testing to detect major gene segregation. The regressive models allow for different forms of dependence in addition to that due to major gene transmission. In class A regressive models it is assumed that this extra dependence among the offspring is due only to the parental phenotypes. In class D regressive models it is assumed that in addition all preceding offspring phenotypes are equally predictive of an offspring's phenotype. This model subsumes the mixed model as a special case for a continuous phenotype in nuclear families. The differences for a discrete phenotype, and in larger family structures, are described.

Genotype × environment interaction in coronary heart disease.

Jean W. MacCluer and John Blangero

Among the problems that arise in identifying genes that affect risk of coronary heart disease are: (1) lack of a well-defined clinical endpoint; (2) the long time span over which disease develops; (3) clinical (and genetic) heterogeneity; and (4) inability to control (or measure) environmental risk factors. These same problems are encountered for other complex and common diseases such as cancer and autoimmune diseases. A problem of particular concern is that an individual's response to environmental factors may be a function of his/her genotype, i.e., there may be genotype by environment interaction. We are investigating the interaction of genotype and environment and the effect of such interaction on risk factors for coronary heart disease using a nonhuman primate model in which we can control, or at least measure, many of the environmental variables that are relevant to disease susceptibility.

Our analyses of lipoprotein phenotypes in pedigreed baboons have indicated many cases in which the effects of genetic variation at specific loci depend upon environmental conditions. We have found that the effects of genotype on serum concentrations of lipoproteins and apolipoproteins may be a function of diet, age, sex, or temperature. The interactions between genotype and environment involve major genes, whose identities are still unknown but which account for a large portion of the phenotypic variance in a trait; known candidate genes, such as the structural loci for apolipoproteins;

Gefördert durch
DFG Deutsche Forschungsgemeinschaft

or polygenes, which are multiple (unidentified) loci that have individually small effects. The environmental factors may be either known (and therefore measured) factors such as diet and temperature, or unknown (random, individual-specific) effects.

Examples of each of these interactions in nonhuman primates are presented, and an extension of these studies to coronary heart disease susceptibility in humans is described.

Supported by NIH grants P01 HL28972 and P01 HL45522.

Genotype environment interaction in carcinogenesis.

Catherine Bonaiti-Pellié

In the case of diseases with genotype-environment interaction, familial aggregation may be difficult to demonstrate, particularly when the environmental exposure is not very frequent in the population. In such a case, it is possible to show that taking into account the environmental exposures of both the proband and their relatives can substantially increase the power to detect familial aggregation.

Statistical methods in cancer epidemiology.

Duncan C. Thomas

Genetic epidemiology might be defined as the study of genetic and environmental determinants of disease using population-based family studies. Classical epidemiology has used studies of independent individuals (cohort and case-control studies) to examine environmental effects and, in some cases, familial aggregation through the use of family history covariates. Statistical methods based on survival analysis have been widely used in this context. Classical genetics has relied on segregation and linkage analysis methods to look at genetic determinants. We have been developing a class of survival models to look simultaneously at genetic and environmental influences in family studies. To overcome the considerable computational difficulties with standard likelihood methods, we have been exploring two approaches, one based on Markov Chain Monte Carlo methods such as the Gibbs sampler, and one based on generalized estimating equations approaches for marginal models in the means, variances, and covariances of the observed phenotypes. Both approaches are illustrated.

Rapporteur: Rolf Fimmers

Tagungsteilnehmer

Dr. Marie-Claude Babron
INSERM U 155
Château de Longchamp
Bois du Boulogne

F-75016 Paris

Prof.Dr. Max P. Baur
Institut für Medizinische Statistik
Dokumentation und Datenverarbeitung
Universität Bonn
Sigmund-Freud-Str. 25

W-5300 Bonn 1
GERMANY

Dr. Heiko Becher
Institut für Epidemiologie und
Biometrie
Deutsches Krebsforschungszentrum
Postfach 10 19 49

W-6900 Heidelberg 1
GERMANY

Dr. Nikolaus Becker
Institut für Epidemiologie und
Biometrie
Deutsches Krebsforschungszentrum
Postfach 10 19 49

W-6900 Heidelberg 1
GERMANY

Prof.Dr. Jürgen Berger
Institut f. Mathematik und DV i. d.
Medizin, Universitätskrankenhaus
Eppendorf, Universität Hamburg
Martinistraße 52

W-2000 Hamburg 20
GERMANY

Dr. Maria Blettner
DKFFZ, Abt. Epidemiologie
Im Neuenheimer Feld 280

W-6900 Heidelberg
GERMANY

Dr. Catherine Bonaiti-Pellié
INSERM U 155
Château de Longchamp
Bois du Boulogne

F-75016 Paris

Dr. Jenny Chang-Claude
DKFZ, Abt. Epidemiologie
Im Neuenheimer Feld 280

W-6900 Heidelberg
GERMANY

Dr. David G. Clayton
Medical Research Council
Biostatistics Unit
Institute of Public Health
Robinson Way

GB- Cambridge CB2 2BW

Dr. Françoise Clerget-Darpoux
INSERM U 155
Château de Longchamp
Bois du Boulogne

F-75016 Paris

Dr. Anthony W. F. Edwards
Department of Community Medicine
Institute of Public Health
University of Cambridge
Forvie, Robinson Way

GB- Cambridge CB2 2SR

Prof.Dr. Robert C. Elston
Dept. Biometry & Genetics
Louisiana State University
Medical Centre
1901 Perdido Street

New Orleans , LA 70112-1328
USA

Dr. Rolf Fimmers
Institut für Medizinische Statistik,
Dokumentation und Datenverarbeitung
Universität Bonn
Sigmund-Freud-Str. 25

W-5300 Bonn 1
GERMANY

Christine Fischer
Institut für Humangenetik
Universität Heidelberg
Im Neuenheimer Feld 328

W-6900 Heidelberg
GERMANY

Dr. Karl-Heinz Jöckel
BIPS
Grünenstr. 120

W-2800 Bremen 1
GERMANY

Prof.Dr. Niels Keiding
Statistical Research Unit
University of Copenhagen
Blegdamsvej 3

DK-2200 Kobenhavn N

Dr. Michael Knapp
Institut für Medizinische Statistik
Dokumentation und Datenverarbeitung
Universität Bonn
Sigmund-Freud-Str. 25

W-5300 Bonn 1
GERMANY

Dr. Ursula Krämer
Medizinisches Institut für
Umwelthygiene
an der Universität
Auf'm Hennekamp 50

W-4000 Düsseldorf 1
GERMANY

Dr. Edward Lustbader
Foxchase Cancer Center
7701 Burholme Avenue

Philadelphia , PA 19111
USA

Dr. Jean W. MacCluer
Southwest Foundation for Biomedical
Research, Dept. of Genetics
P.O. Box 28147
West Loop 410 at Military Drive

San Antonio , TX 78228-0147
USA

Dr. Wolfgang Maier
Psychiatrische Klinik
Universität Mainz
Untere Zahlbacher Str. 8

W-6500 Mainz
GERMANY

Dr. Maria M. Martinez
INSERM U 155
Château de Longchamp
Bois du Boulogne

F-75016 Paris

Dr. Jurg Ott
Psychiatric Institute
Columbia University, Unit 58
722W 168TH Street

New York , NY 10032-2603
USA

Prof.Dr. Martin Schumacher
Institut für Medizinische Biometrie
und Medizinische Informatik
Klinikum der Universität
Stefan-Meier-Str. 26

W-7800 Freiburg
GERMANY

Susanne Seuchter
Institut für Medizinische Statistik,
Dokumentation und Datenverarbeitung
Universität Bonn
Sigmund-Freud-Str. 25

W-5300 Bonn 1
GERMANY

Prof. Dr. Richard Spielman
Dept. of Human Genetics
School of Medicine
University of Pennsylvania
422 Curie Boulevard

Philadelphia ,PA 19104-6145
USA

Dr. Jon Stene
Institute of Statistics
University of Copenhagen
Studiestraede 6

DK-1455 Copenhagen K

Prof. Dr. Duncan C. Thomas
Department of Prev. Medicine
University of South California
1420 San Pablo St.

Los Angeles , CA 90033-9987
USA

Dr. Alun Thomas
School of Mathematical Sciences
University of Bath
Claverton Down

GB- Bath Avon BA2 7AY

Prof.Dr. Hans-Joachim Trampisch
Institut für Medizinische
Informatik und Biomathematik
Ruhruniversität Bochum
Universitätsstr. 150

W-4630 Bochum 1
GERMANY

Prof.Dr. Wolfgang Urfer
Fachbereich Statistik
Universität Dortmund
Postfach 50 05 00

W-4600 Dortmund 50
GERMANY

Prof.Dr. Jürgen Wahrendorf
Institut für Epidemiologie und
Biometrie
Deutsches Krebsforschungszentrum
Postfach 10 19 49

W-6900 Heidelberg 1
GERMANY

Dr. Stefan Wellek
Institut für Medizinische Statistik
und Dokumentation
Universität Mainz
Langenbeckstr. 1

W-6500 Mainz 1
GERMANY

Prof.Dr. Dr. H.Erich Wichmann
GSF- Institut für Epidemiologie
Ingolstädter Landstr. 1

W-8042 Neuherberg
GERMANY

Dr. Thomas F. Wienker
Institut für Humangenetik
Universität Freiburg
Breisacher Str. 33

W-7800 Freiburg
GERMANY

Gefördert durch
DFG Deutsche Forschungsgemeinschaft