STATISTICS IN MEDICINE, VOL. 16, 2773–2784 (1997)

THE FUTURE OF BIOSTATISTICS: EXPECTING THE UNEXPECTED

HANS C. VAN HOUWELINGEN*

Department of Medical Statistics, Leiden University, P.O. Box 9604, 2300 RC Leiden, The Netherlands

SUMMARY

The paper explores the future of (clinical) biostatistics. In the first part of the paper important trends in biostatistics are reviewed: new applications and more complex data; causal models for observational data; cross-validation-based model building; graphical chain and random effect models; faster computing and new algorithms as Markov chain Monte Carlo; generalized estimating equations and pseudo-likelihood; pooled data sets for meta-analysis and prognostic modelling. In the second part some dreams and nightmares of the author are sketched. Dreams are: implementation of prognostic and diagnostic models in the clinic; an instantaneous numeric integrator; much more and better organized follow-up data; disease mapping in space and time. 'Nightmares', that is, issues which it is hoped will go out of use in the future, are: *P*-value; rank tests; exact methods; meta-analysis; matched case-control studies. © 1997 by John Wiley & Sons, Ltd.

Statist. Med., **16**, 2773–2784 (1997)
No. of Figures: 0 No. of Tables: 1 No. of References: 43

INTRODUCTION

The paper consists of two parts. First, a number of important trends in biostatistics are reviewed. Next, a more personal view on the future is given by sketching four dreams of which it is hoped that they will be realized and become good statistical practice, and by giving a list of 'nightmares' concerning issues in current biostatistical practice, of which it is hoped that they will disappear in the future.

Of course, these are all personal ideas and everybody is most welcome to disagree. The main purpose of the paper is to trigger the discussion on the future of the discipline.

PART 1. IMPORTANT TRENDS IN BIOSTATISTICS

New applications, more complex data

(i) Patient history data in oncology and chronic diseases.
(ii) Family history data in genetic epidemiology.
(iii) Spatial and spatiotemporal data in ecological and environmental research.

The first and maybe the most important trend of every day biostatistics is that we are going to be confronted with larger data sets and more complex data. There is a growing interest in collecting

* Correspondence to: Hans C. van Houwelingen, Department of Medical Statistics, Leiden University, P.O. Box 9604, 2300 RC Leiden, The Netherlands

CCC 0277–6715/97/242773–12$17.50

and analysing follow-up data of individual patients.[1,2] Examples are markers such as CD4 counts[3] in AIDS and CA15-3[4] in cancer. These data are collected at a more or less regular time schedule and the hope is that they are predictive for the patient's future and can be used as surrogate endpoints in clinical trials. Another example is quality-of-life data for all kinds of chronic disease. The debate on what quality-of-life data should be collected is still on[5] and we definitely should not burden the patients with long and complicated questionnaires, but we can take it for granted that we are going to have more follow-up data to analyse.

Biostatisticians have a lot of experience with growth data[6] and we shall have to generalize the techniques used for growth data to more complicated data where follow-up is more irregular, dropout is a problem not to be discarded and the outcome measures are not normally distributed. Multi-level models with patient-specific regression lines could be one way of approaching these data.[7,8]

Another example of complicated data are family history and pedigree data in genetic epidemiology.[9] These data will not only be used in linkage analysis to locate the gene in simple genetic models, but they will also be used in assessing an individual's risk given his family history. We will have to come forward with something more sophisticated than simple 'positive family history', that is, one (or two) affected family members in first or second degree. We will have to take into account the whole pedigree and use all information available to give the best assessment of an individual's risk.

Finally, there is a growing interest for geographic data in epidemiology and environmental and ecological research. There is some analogy with the pedigrees in genetic epidemiology. The data consist of raw 'maps' of, for example incidence rates of a certain cancer in geographic units as the counties in Scotland (Clayton and Kaldor's famous lip cancer data[10]). We will be confronted with the task to analyse these data, to explain them by demographic and geographic covariates, to smooth the residuals and to give the best estimate of the 'map' highlighting 'hots spots' if present. We can learn a lot from the 'engineers' who developed geographical information systems and techniques such as kriging.[11] A mixture of empirical Bayes methodology and kriging can do the job here, if we can overcome the computational problems inherent in high-dimensional parameter spaces.[12]

New philosophies

(i) Causal models instead of randomization in clinical trials (with non-compliance).
(ii) Prediction ((cross-) validation) versus model fitting in prognostic models.

In the future, we will also have to deal with changing philosophies. In clinical trials where randomization is hampered by non-compliance or secondary treatment depends on the patient's performance, there is a growing interest in causal models stimulated by the work of James Robins[13] and others.[14] Simple 'intention-to-treat' will not always answer the questions we have. We will have to overcome our reluctance and will have to study the literature to understand the 'counterfactual world' including the untestable assumption of 'conditional independence' or 'no remaining confounding'. There is a growing awareness that not all research problems can be answered by randomized clinical trials. How do you, for example, randomize surgical treatment if it takes a long training and continued practice to master a certain surgical procedure? Randomization within a surgical centre may render both treatments sub-optimal and endanger the patient. So we will have to rely on quasi-experiments and have to make inferences from inherently

observational data, such as treatment A in centre I and treatment B in centre II plus information on patients' characteristics even if that includes untestable assumptions.

I do believe that in the future the emphasis for our statistical analyses will shift from studying the effect of a single covariate (after correction for confounders) to building prognostic models for individuals. So we will have to assess the predictive value of models and develop instruments to compare not necessarily nested models on predictive performance. That asks for a different view on statistical model building.[15,16] Maybe Akaike's information criterion will take over completely from the *P*-values.[17] In modelling covariance structures for longitudinal data, that is already standard practice.

New models

(i) Graphical chain models.
(ii) Random effect models (fixed effects versus random effects).

More complicated data ask for new models to describe the data and to draw conclusions from the data. I see two important developments here.

First, there are the graphical chain models,[18] which give more insight in the structure of relationships. They originate from the social sciences where structural models have been popular for quite a long time. As biostatisticians we have always avoided the use of the complicated models of LISREL and the like, but the growing complexity of our data might force us to have a better look into this approach to statistical modelling.

Another very important trend in statistical modelling is the growing use of random effect models. You find them, for instance, in psychometrics in item-response-theory,[19] in educational sciences in the multi-level modelling of school tests,[20] in animal breeding to describe parents-effects for clustered data,[21] in epidemiology to describe heterogeneity,[22] in follow-up data to model dependence over time[7,8] and in genetic data to model latent traits.[9] The modelling of heterogeneous effects in meta-analysis is still another example.[23] The awareness is still growing that there is more randomness than simple measurement error or simple randomization. In The Netherlands we had a one-day workshop on this theme in 1995 under the heading 'toeval zit overal', meaning 'chance is everywhere' with David Spiegelhalter (biostatistics) and Harvey Goldstein (social sciences) as keynote speakers indicating that these so-called mixed models are important for all statisticians and that they can even bring statisticians from different fields of applications together to work on models of common interest.[24]

New computing facilities

(i) Faster computers.
(ii) New clever algorithms for integration and maximization.

Predictably, but quite unpleasantly, these new models are hard to fit to data in the sense that they need much computer time even on the modern fast computers. Computing the likelihoods of the models involves high-dimensional integration.[25] Monte Carlo methods, and, more specifically, Markov chain Monte Carlo, can be quite useful. BUGS, developed at the MRC in Cambridge by Gilks, Spiegelhalter and others,[26] is based on this technique and is quite useful to solve medium-size medium-dimensional problems. In complicated genetic data, we are still far away from practical solutions. An unexpected breakthrough in this field would be wonderful. I will come back to this issue later.

New techniques

(i) Descriptive statistics (graphical displays).
(ii) Inferential statistics:
  (a) pseudo-likelihood and generalized estimation equations,
  (b) exact methods.

The exploding capacities of our personal computers have yielded to all kind of technical possibilities. On several occasions in the summer of 1996 I have seen adaptive graphics[27] at work, where you can adapt the graphics on the screen by simply clicking with the mouse. It looks fascinating and I can imagine people sitting at the PC and adapting the regression plot or the bivariate histogram of the geographical image until they obtain the picture they like best. I think there are two big problems here. First of all, we are three-dimensional beings unable to visualize data of higher dimensions. Secondly, the result of all these 'adaptive' graphics is that they are highly, dangerously subjective. The criteria for best fit should be formalized and agreed upon and not left to the visual assessment of the analyser at the screen. Statistics should be science and not an art!

Another offspring of fast computers and clever programmers are so-called exact methods in statistics. Actually, this is one of my nightmares. I will come back to it later.

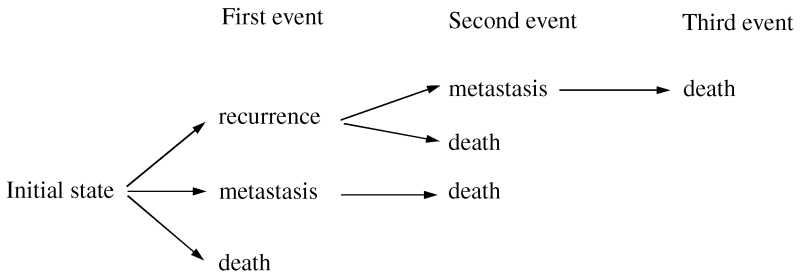
A practical solution to problems we cannot quite handle yet are generalized estimating equations,[28] GEE for short. They are quite handy in a lot of situations and their introduction started a very useful discussion on the choice between subject-specific or population-averaged models.[29] They teach us that we have to be careful in the interpretation of the parameters in non-linear mixed models. GEE can be very handy to estimate parameters. It has to compete here with pseudo- and quasi-likelihood models. However, we have to realize that it will not give us the complete high-dimensional model. In my dreams there is only little room for GEE and pseudo-likelihood.

New forms of collaboration

(i) Pooled databases for meta-analysis and prognostic modelling.
(ii) Software exchange through Internet.
(iii) Fast publication through Internet.

Finally, in this overview I have to say something about the blessings of the Internet. Undoubtedly, the Internet facilitates fast communications between researchers. It can be very helpful in setting up databases in multinational multi-centre clinical trials and for pooled-data meta-analysis. It makes central data handling possible with decentralized data entry. We can gain time here, if we are not too much distracted by all the fancy features of the Internet. We have to learn to live with the Internet. We learned to live with e-mail, that is we learned to print our e-mails and to check them before we send them out and we learned not to be distracted by incoming e-mail and wait until the end of the day before answering them. Similarly, we will adjust to Internet and we will learn again that 'more hurry' goes with 'less speed'. Of course it is nice to have users libraries of S-plus users and similar software sharing, but good software takes time to develop and, maybe, we should leave the development of software to the professionals. Publication of journals through Internet is marvellously fast and it will definitely be realized in the future, but I hope that there will be also hard copies stored on the bookshelves instead of in the back-ups of our networks, that we cannot locate anymore. Luckily,

Table I



we are not the only Internet users and we can calmly wait until the Internet system has stabilized and found its place in the scientific endeavour.

PART 2: DREAMS AND NIGHTMARES

Dream 1: On the spot prognosis/diagnosis

After this short overview of current trends in biostatistics (see *Advances in Biometry*[30] for a very broad overview), it is time to share some of my dreams with you. My first dream is to implement statistical models in such a way that they can be used at the spot in the hospital or any other relevant place. The example I will discuss in some detail is a recent experience with the modelling of oncological data. I was asked to analyse a data set of about 1500 patients with up to 10 years follow-up. It concerned breast cancer patients who underwent breast-saving surgery followed by radiotherapy. In the analysis the only patient characteristics with substantial effect on outcome turned out to be T/N-classification of the tumour, age of the patient and side of the tumour (left or right breast). The clinical researcher involved asked for traditional Cox regression analyses with local recurrence, metastasis or death as outcome measures. For the latter two outcome measures, local recurrence was used as a time-dependent covariate to check its effect on the hazard of metastasis and death. The results of this analysis will be reported in the traditional style with *P*-values and confidence intervals for relative risk.

In my view a competing-risk multi-state model[31] is much more appropriate here. The schedule for such an analysis is given in Table I.

Transitions in this model are:

intitial state	→ local recurrence, metastasis or death as competing risk for the first event;
recurrence as first event	→ metastasis or death as second event;
metastasis as first event	→ death (local recurrence after or simultaneous with metastasis is ignored)
metastasis as second event	→ death.

The rates of these transitions are easy to model. The transition rate metastasis → death does not depend on the occurrence of previous local recurrence, but does depend on the waiting time until metastasis. The effect of the covariates is different for the different transitions.

This model can easily be used to build a dynamic prognostic model in which the distribution of the remaining lifetime or the remaining disease-free lifetime of a patient can be computed given the patient's history until that moment in time. This information can easily be visualized.

It is my dream that such a small program is incorporated in the hospital information system. The engineers in medical informatics are building quite complicated systems in which it is possible, for instance, to recall all pathological and histological information available for the patient. It would not cost much effort to implement my little program. It would enable the doctor to give a patient an update of her 'chances' at the time of every planned (or unplanned) visit to the clinic. Of course, there is an ethical question whether, for example, you should tell a patient how bad her prognosis is after diagnosing metastasis, but you can also wonder if it is not equally unethical not to inform the patient about her improved outlook if she remained free from local recurrence and metastasis during 5 years, say.

The problem, according to my experience, is not the ethical one, but the great difficulty the doctors have in understanding the appropriate dynamic multi-state model. Regrettably, there is a large gap between what we can do as experienced statistical model builders and what we can explain to the doctor. I think we should spend more energy and time on bridging this gap. One way of doing it is to publish this kind of modelling together with examples of dynamic prognosis in the medical literature. We should recognize that even *Statistics in Medicine* is very hard to read for medical people despite all efforts of the editors to keep it as applied as possible. If *Statistics in Medicine* is mainly read by statisticians, we should perhaps found a new journal, *Medicine in Statistics*, aiming at the doctors and explaining in their language the merits of new developments in statistics. I think the epidemiologist can help us to bridge the gap between statistics and medicine.

Dream 2: Instantaneous integrator

My second dream is purely theoretical. It relates to the problems we have in fitting non-linear mixed models to data. The problem I would like to solve is just the modelling of repeated dichotomous data. So the data we have per individual is a vector Y of dichotomous outcome measured at times $t_1, \ldots, t_k$. The number of observations and their timing may differ between individuals. It is my believe that these data can best be modelled by a subject-specific latent variable model:

$$\text{logit}(\Pr(Y_{it} = 1)) = X'_{it}\beta + \theta_{it} \quad i: \text{person} \quad t: \text{time}.$$

If the latent variable θ is assumed to follow a multivariate normal distribution, or, to be more precise, if it is a Wiener process, a lot of plausible models are thinkable. The simplest is that θ is independent of time. That problem is solved numerically in EGRET. A multi-level model would be $\theta_t = a + bt$ with random a and b. That is already getting harder to fit to the data. I would like to try an autoregressive process or simple Brownian motion. Anyhow, such models would describe the covariance structure of the θ_t's at the relevant moments in time with some model depending on some covariance-parameters.

First of all I would like to estimate the parameters of this model by maximum likelihood or REML (restricted maximum likelihood). Having estimated the parameters I would like to compute the predictive likelihood of future observations of an individual given his past, that is, $\Pr(Y_{\text{future}} \mid Y_{\text{past}})$. For both goals, I have to compute the distribution of the observed Y's which involves integration over high-dimensional spaces.

In early 1995 I tried to solve this problem by a combination of the EM algorithm and approximations of the integrals by multivariate normal integrals around the posterior mode. I was quite happy with the result and presented it at the Medical Statistics meeting at Oberwolfach in Spring 1995. The next speaker at that meeting was Norman Breslow who explained that these approximations can give serious bias in the estimation of the covariance parameters.[32] He had some improvements, but in the ongoing debate[33] in the literature it has become clear that all these approximations can be very bad if there is only little information per individual. So, my dream is an genuinely unexpected breakthrough in numerical computing that produces integration with speed and accuracy that does not, or hardly, depend on the dimension of the problem. In linear programming in operational research a similar miracle has happened, so why can the miracle not happen here. In the meantime we have to live with unsatisfactory approximations.

Dream 1 + 2 combined: On the spot assessment of family history information using large pedigrees

Jeanine Houwing-Duistermaat, a PhD student of mine, is working on statistical models for pedigree data.[34] Here again, the latent genetic effects are very hard to integrate out. My superdream here is that we can both solve the numerical problems and install an 'automatic genetic counsellor' that uses all information and the best model to assess genetically driven prognosis and diagnosis at the individual's level. It is still a dream, but we work hard to realize some parts of it.

Dream 3: Follow-up data

My third dream is about databases, especially for follow-up data. It is my experience that proposals for clinical trials are often presented as if no, or hardly any, data are available. That concerns both phase II and phase III trials. It is my dream that we collect data, even if we run no clinical trials. Whether you call it registry or database does not matter, data are needed, not only to compare treatment but also to gain more insight in the 'natural course' of a disease, to quantify whether the situation of a patient is worsening, etc. It can be worthwhile to collect data even if the therapy is left to the discretion of physician and/or patient. The final hard proof of the superiority of one treatment over another can only be settled by a randomized trial, but if we have more data of patients not in clinical trials we may have better ways to select really promising therapies. Now, it is already happening that data sets from clinical trials are used for completely different, more epidemiological, purposes. In collecting follow-up data we should be aware of the danger of data coming from examinations prompted by the patient, potentially in response to his condition and therefore not (completely) at random. Such data are very hard to model convincingly.

So, my dream is that protocols are written about what kind of data to collect, when and for what disease. If a clinical trial is planned such protocols can be used for the measurement design. Only the choices of therapy and the sample sizes have to be filled in. So, if there are no promising new treatments at this moment for a particular disease, I would like to make a plea for a pseudo-trial of standard treatment versus standard treatment: the A-A trial. Ask the patient's consent to collect his/her data just for the case of being better documented about the disease. Try to reach consensus with colleagues about what prognostic and follow-up data should be collected. That would make databases much better comparable. Well organized databases would allow us to generalize the concept of historic controls to a broader concept of a database controlled study.

Finally, as part of this dream, the old fashioned meta-analysis in which estimated log-odds-ratios or other effect measures are compared can be abolished in favour of an integrated analysis of all available data,[35] taking into account unexplainable variation between centres.

Dream 4: Disease mapping in space and time

My last dream takes me back to statistical modelling. Analysis of spatial data has drawn considerable attention over the last years. Disease mapping of standardized epidemiological incidence rates by empirical Bayes techniques that try to smooth away the measurement error, but keeps the true variation, has become one of the standard epidemiological techniques. I have seen quite a number of these smoothed maps in *Statistics in Medicine*[36,37] and similar journals. As I said before, there is a link with geographical information systems with smoothing techniques like kriging and Bayesian image analysis[38] in medical imaging. There is still the unsolved problem of combining space with time. Disease maps can change over time and we are not so much interested in the map of last year or even this year, but more in the map of next year to see whether any consistent pattern over time can urge us, or the authorities, into action. A lot of research is done on models for space-time data[39] but the final technology has not been developed yet. One could imagine autoregressive types of models, where the prediction based on last year is a local smoothing of last year's map. It will take some time to figure it all out, but I am sure that we will be able to extrapolate maps to the future.

To show an example from which it is clear that the future is really what interests us, I will briefly describe an exercise on hospital comparison in which my department is involved. In a series of hospitals, information is collected each year on mortality of pre-term infants and the case-mix (patients characteristics) in each hospital. Based on this information, observed and expected mortality numbers in each hospital can be computed each year. The data can be analysed in each year in the spirit of a heterogeneous meta-analysis, that is by a mixed-model with a random hospital effect.[40] Empirical Bayes methods allow us to compute an estimate of the (random) hospital effect. This information is only relevant if the random fluctuation between hospitals is reasonably constant over time. That is close to my second dream where I wanted to predict future observations per individual. Here, it is wanted to predict future hospital effects. If these predictions show a certain variation between hospitals, which is only possible if the hospital effects are correlated over time, it can be considered whether measures should be taken to take away that variation between hospitals. A simple measure would be to close down a hospital with persistent high mortality.

If there is some spatial correlation between the hospitals we are close again to predicting maps. If different outcomes are correlated as well we have to consider an extra dimension which makes things even more complex.

'Nightmares' (I hope they do not come back)

(i) *P*-values.
(ii) Rank tests.
(iii) Exact methods.
(iv) Meta-analysis.
(v) Matched case-control studies.

My nightmares is a list of issues which I hope will go out of use in the future; concepts that will only haunt me in my nightmares. After such a nightmare, I hope to wake up with the happy feeling that I will not have to use them any more.

My list of nightmares may sound a little heretical, but, as I said before, this is my personal view and the reader is welcome to disagree. Moreover, the nightmares relate to the ideal future and I realize that on the road to the ideal future we have to live with the practical solutions of today.

The first one on my list is the classical *P-value*.[41] They are still abundant in medical journals. I have the honour of being the statistical editor for one such journal and I still have to fight conclusions such as 'treatments are equivalent because P is not smaller than 5 per cent'. If we are not able to explain to the medical people what P-values really mean, we should try to abandon them. I think that there is room for a concept as 'the probability that the null hypothesis is true', but that can only be formulated in a Bayesian setting. A drawback of Bayesian methods is that the prior has to be chosen by the writer of the paper. In the future of electronic publication, interaction between the reader and the 'paper' may be possible, allowing the reader to choose his personal prior and arrive at his personal probability that the null hypothesis is true.

The next item on my list is the *rank test*. My main objection is that they only give a test and no estimate of effect. I know that confidence intervals can be constructed from rank test in a semi-parametric setting, but I also know that it is never done. The software actually transforms the data to a uniform distribution and then applies a procedure fit for normal outcomes such as the t-test and the Pearson correlation coefficient. I think it is more useful to use parametric models after a Box–Cox type of transformation. In my view rank tests are only used as P-value producers and I want to get rid of both of them.

Next one is '*exact methods*'.[42] They either rely on conditioning on ancillary statistics or presume that the sampling mechanism is known. My feeling is that we hardly ever know how the data were actually collected and what stopping rules were used in the sampling process. An example I recently encountered was a data set with number of miscarriages, X, say, and number of pregnancies, n, say. Possible models are: a binomial model for X given n or a negative binomial distribution for n with stopping based on the number $(n - X)$ of successful pregnancies. The likelihoods do not differ, but the P-values do, because they are sampling-scheme dependent. So what is the use of obtaining 'exact P-values' here? I do not know. My main objection is that the term 'exact' somehow suggests some optimality of the corresponding testing procedure. Even if the sampling scheme is known, it is questionable whether tests based on exact conditional P-values are to be preferred over tests based on (approximate) unconditional P-values. 'Exact' tests might be quite conservative and, in this sense, not exact at all.

Meta-analysis is another nightmare. I wholeheartedly support the idea of combining evidence from different sources, but the popular practice of analysing summary measures from selected publications is a poor man's solution. As I said before, I hope that we will have full multi-centre multi-study databases that can be analysed by appropriate random effect models considering both random variation within and between studies and/or centres. In that ideal situation we are back to the data, there is no 'meta'-aspect on the analysis anymore and the term 'meta-analysis' can be skipped from the dictionary. In the meantime, we have to walk the poor man's path of meta-analysis and to perform the best analysis of the evaluable summary data modelling between study variation by explanatory variables at the meta-level and residual random effects.

My last nightmare is the *matched* case-control study. In a simple (non-matched) case-control study or discriminant analysis, a logistic model for the patient-specific risk can be obtained from

the logistic model fitted to the case-control data if the prevalence of the disease is known. To adjust the model we only need to add

$$\ln\left(\frac{\text{prevalence}}{1-\text{prevalence}}\right)-\ln\left(\frac{\text{number of cases}}{\text{number of controls}}\right)$$

to the constant term. In the analysis of matched case-control data by the Mantel–Haenszel method or by conditional logistic regression, only the effect of the exposure variables within the case-control strata is studied. No attempt is made to model the variation in risk between the strata. To translate the findings of a matched case-control study into a patient-specific risk model, one needs to know the stratum-specific prevalences. However, in matched case-control studies it is often unknown from how many potentially matchable controls we are sampling. Hence, the stratum specific prevalences cannot be estimated from the matched case-control data and no patient-specific risk model can be obtained, unless we have external information about the stratum-specific risk. The matched case-control study can be useful in identifying exposures that affect the disease we are interested in, but I wonder what the use is of knowing the odds-ratio for one particular exposure, if we do not know quantitatively what other factors do affect the outcome nor what is the baseline risk.

In summary, my personal dream is the biostatistician as modeller of the chances of a patient in patient-specific models that take into account all patient information and medical knowledge. To get there we need good data, realistic models with optimal discrimination between patients and access to the clinic through good co-operation with quantitatively educated clinicians. Let us all work hard to get there.

The future starts tomorrow.
Working on the future has started already!

ACKNOWLEDGEMENTS

This paper is the written version of an invited paper presented at ISCB-XVII in Budapest, 1996, in a session devoted to Philosophy, Past and Future of Clinical Biostatistics. I want to thank the Program Committee and especially its chairman Michael Schemper for inviting me. I also want to thank the colleagues of my department for fruitful discussions on this topic over lunch and during coffee-breaks. Knowing that I was to give this talk, I listened with extra attention to similar future-oriented talks at other meetings that summer. I want to mention the Presidential Address of Byron Morgan at the International Biometric Conference in Amsterdam, July 1996 and the Opening Lecture by Sir David Cox and the R.A. Fisher Lecture by Bradley Efron, both during the Joint Statistical Meetings in Chicago in August 1996. I enjoyed their broad views and my ideas are definitely influenced by theirs.

REFERENCES

1. D'Agostino, R. B., Lange, N., Ryan, L. M. and Selwyn, M. R. (eds). 'Symposium on longitudinal data analysis 19–21 June 1991', *Statistics in Medicine*, **11**, 14/15, 1997–2040 (1992).
2. Diggle, P. J., Liang, K.-F. and Zeger, S. L. *Analysis of Longitudinal Data*, Oxford University Press, 1994.
3. Lange, N., Carlin, B. P. and Gelfand, A. E. 'Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers', *Journal of the American Statistical Association*, **87**, 615–626 (1992).
4. Schlain, B. R., Lavin, P. T. and Hayden, C. L. 'Using autoregressive and random walk models to detect trends and shifts in unequally spaced tumour biomarker data', *Statistics in Medicine*, **12**, 265–279 (1993).

5. Cox, D. R., Fitzpatrick, R., Fletcher, A. E., Gore, S. M., Spiegelhalter, D. J. and Jones, D. R. 'Quality-of-life assessment: Can we keep it simple?', *Journal of the Royal Statistical Society*, *Series A*, **155**, 353–393 (1992).
6. von Rosen, D. 'The growth curve model: a review', *Communications in Statistics – Theory and Methods*, **20**, 2791–2822 (1991).
7. Vonesh, E. F. and Carter, R. L. 'Mixed-effects nonlinear regression for unbalanced repeated measures', *Biometrics*, **48**, 1–17 (1992).
8. Barry, D. 'An empirical Bayes approach to growth curve analysis', *Statistician*, **45**, 3–19 (1996).
9. Thompson, E. A. 'Genetic epidemiology: a review of the statistical basis', *Statistics in Medicine*, **8**, 291–302 (1986).
10. Clayton, D. and Kaldor, J. 'Empirical Bayes estimates of age-standardised relative risks for use in disease mapping', *Biometrics*, **43**, 671–681 (1987).
11. Matheron, G. 'Principles of geostatistics', *Economic Geology*, **58**, 1246–1266 (1963).
12. Besag, J. and Green, P. J. 'Spatial statistics and Bayesian computation', *Journal of the Royal Statistical Society*, *Series B*, **55**, 25–37 (1993).
13. Robins, J. M. and Greenland, S. 'Adjusting for differential rates of prophylaxis therapy for PCP in high-versus low-dose AZT treatment arms in an AIDS randomized trial', *Journal of the American Statistical Association*, **89**, 737–749 (1994).
14. Holland, P. 'Reader reaction: Confounding in epidemiologic studies', *Biometrics*, **45**, 1310–1316 (1989).
15. van Houwelingen, J. C. and le Cessie, S. 'Predictive value of statistical models', *Statistics in Medicine*, **8**, 1303–1325 (1990).
16. Harrel, F. E., Lee, K. L. and Mark, D. B. 'Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing error', *Statistics in Medicine*, **15**, 361–387 (1996).
17. Verweij, P. J. M. and van Houwelingen, H. C. 'Cross-validation in survival analysis', *Statistics in Medicine*, **12**, 2305–2314 (1993).
18. Cox, D. R. and Wermuth, N. *Multivariate Dependencies*: *Models*, *Analysis and Interpretation*, Chapman & Hall, London, 1996.
19. Longford, N. T. 'Multivariate variance component analysis: An application in test development', *Journal of Educational Statistics*, **15**, 91–112 (1990).
20. Aitkin, M. and Longford, N. 'Statistical modelling in school effectiveness studies (with discussion)', *Journal of the Royal Statistical Society*, *Series A*, **149**, 1–43 (1988).
21. Tempelman, R. J. and Gianola, D. 'A mixed effects model for overdispersed count data in animal breeding', *Biometrics*, **52**, 269–283 (1996).
22. Breslow, N. E. and Clayton, D. G. 'Approximate inference in generalized linear mixed models', *Journal of the American Statistical Association*, **88**, 9–25 (1993).
23. Berkey, C. S., Hoaglin, D. C., Mosteller, F. and Colditz, G. A. 'A random-effects regression model for meta-analysis', *Statistics in Medicine*, **14**, 395–411 (1995).
24. Goldstein, H. and Spiegelhalter, D. J. 'League tables and their limitations: statistical issues in comparisons of institutional performance (with discussion)', *Journal of the Royal Statistical Society*, *Series A*, **159**, 385–443 (1996).
25. Evans, M. and Swartz, T. 'Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems', *Statistical Science*, **10**, 254–272 (1995).
26. Spiegelhalter, D. J., Thomas, A., Best, N. G. and Gilks, W. R. *BUGS*: *Bayesian Inference using Gibbs Sampling*, *Version 3.0*, MRC Biostatistics Unit, Cambridge, 1994.
27. Buja, A., Cook, D. and Swayne, D. F. 'Interactive high-dimensional data visualization', *Journal of Computational and Graphical Statistics*, **5**, 78–99 (1996).
28. Zeger, S. L., Liang, K.-Y. and Albert, P. S. 'Models for longitudinal data: a generalized estimating equation approach', *Biometrics*, **44**, 1049–1060 (1988).
29. Neuhaus, J. M., Kalbfleisch, J. D. and Hauck, W. W. 'A comparison of cluster-specific and population averaged approaches for analyzing correlated binary data', *International Statistical Review*, **59**, 25–36 (1991).
30. Armitage, P. and David, H. A. (eds). *Advances in Biometry*, Wiley, New York, 1996.
31. Hansen, B. E., Thorogood, J., Hermans, J., Ploeg, R. J., van Bockel, J. H. and van Houwelingen, J. C. 'Multistate modelling of liver transplantation data', *Statistics in Medicine*, **13**, 2517–2529 (1994).

32. Lin, X. and Breslow, N. E. 'Bias correction in generalized linear models with multiple components of dispersion', *Journal of the American Statistical Association*, **91**, 1007–1016 (1996).
33. Lee, Y. and Nelder, J. A. 'Hierarchical generalized linear models (with discussion)', *Journal of the Royal Statistical Society*, *Series B*, **58**, 619–678 (1996).
34. Houwing-Duistermaat, J. J., Derkx, B. H. F., Rosendaal, F. R. and van Houwelingen, H. C. 'Testing familial aggregation', *Biometrics*, **51**, 1292–1301 (1995).
35. Stewart, L. A. and Clarke, M. J. 'Practical methodology of meta-analysis (overviews) using individual patient data', *Statistics in Medicine*, **19**, 2057–2079 (1995).
36. Jacquez, G. M. (ed.) 'Workshop on Statistics and Computing in Disease Clustering, Port Jefferson, NY, USA, 23–24 July 1992', *Statistics in Medicine*, **12**, 19/20, 1751–1968 (1993).
37. Lawson, A., Waller, L. and Biggert, A. (eds). 'Spatial disease patterns', *Statistics in Medicine*, **14**, 21/22, 2289–2508 (1995).
38. Besag, J., Green, P. J., Higdon, D. and Mengersen, K. 'Bayesian computing and stochastic systems', *Statistical Science*, **10**, 3–41 (1995).
39. Bernardinelli, L., Clayton, D., Pascutto, C., Montomelli, C., Ghislandi, M. and Songini, M. 'Bayesian analysis of space-time variation in disease risk', *Statistics in Medicine*, **14**, 2433–2443 (1995).
40. Thomas, N., Longford, N. T. and Rolph, J. E. 'Empirical Bayes methods for estimating hospital-specific mortality rates', *Statistics in Medicine*, **9**, 889–903 (1994).
41. Freeman, P. R. 'The role of *P*-values in analysing trial results', *Statistics in Medicine*, **12**, 1443–1452 (1993).
42. Mehta, C. R. and Patel, N. R. 'Exact logistic regression: theory and applications', *Statistics in Medicine*, **14**, 2143–2160 (1995).
43. Breslow, N. E. 'Statistics in epidemiology: the case-control study', *Journal of the American Statistical Association*, **91**, 14–28 (1996).

