International Statistical Review (2005), **73,** 2, 259–262, Printed in Wales by Cambrian Printers

Statistics for Studies of Human Welfare

Nanny Wermuth

Department of Mathematical Statistics, Chalmers/Gothenburg University, Sweden
E-mail: wermuth@math.chalmers.se

Summary

Developments in the recent past have substantially increased our ability to measure, compute, and communicate. We take the view that a corresponding improved understanding of processes in the life sciences will come about only through more intensive studies of properties of statistical methods and algorithms and transparent, open source computing environments.

Key words: Causality; Graphical Markov models; Life sciences; Open source computing; Randomized intervention studies.

1 Introduction

Human welfare is a main topic of the life sciences. In a narrow sense this comprises health and biology, in a broader sense it includes psychological, educational, environmental and social issues. Empirical evidence is needed to understand the *status quo* and developmental processes, or to form the basis for decisions affecting individuals directly, or indirectly via changes in institutions and societies. Within the statistical sciences, methods to evaluate empirical evidence are developed and studied, methods both for the design of studies and for the analysis and interpretation of relevant data.

The innovative technical developments in the recent past have substantially increased our ability to measure, compute and communicate. We can measure many more features of individuals, institutions, regions and societies, whether it be surveillance information for a patient in an intensive care unit, the decoding of genes of individuals, performance indicators of health and educational systems or measures of stability, growth and equality in different regions. We can easily record study protocols, combine data from different sources, use complex statistical analyses, and make results available in both numerical and graphical form. Communication across different fields of specializations, across institutions, countries and continents has become feasible and can be extremely fast.

Why then is there not a corresponding increase in our understanding and in decisions that are solidly evidence-based? Some provisional answers are (1) that we need more discussion of common goals and more studies of properties of measurements, methods, and algorithms; and (2) that, for some, there may still be severe access restrictions to available knowledge and technologies due, for instance, to economic, ideological or institutional pressures.

Surely, we cannot as individuals remove the many different forms of misuse of power, but as statisticians we can contribute significantly to the transparency of empirical studies and methods of analysis as well as to the availability of open source software, tested for reliability. With the support of scientific and professional societies we can promote discussions and come to agreement on general goals and important achievements.

2 Randomized Intervention Studies

Arguably the most successful types of empirical study in the life sciences are interventions in which treatments are allocated to individuals via randomization (Mosteller & Baruch, 2002; Armitage, 2003; Zeger, Diggle & Liang, 2005). It is closest to the laboratory experiment of the natural sciences. Agreement on study protocols improves the chances for direct comparisons of results. Trust in results is increased with easily accessible documentation provided, for instance, for effects of health care by the Cochrane collaboration and for effects of educational, behavioral and social interventions by the Campbell collaboration. Both of these nonprofit organizations were started in the UK but turned rapidly into larger international collaborations.

A typical path to understanding includes individual case studies, retrospective and prospective case-control studies, randomized interventions and animal experiments. For example, this was the case with malformations in the eye of a newborn caused by a first exposition of the mother to Rubella during the initial three months of pregnancy, and for lung cancer caused by intensive cigarette smoking.

It is worth mentioning that randomization has been introduced into case studies of single individuals (Sackett *et al.*, 1996). And that in case-control studies in epidemiology, the comparison of a treatment group with a control group is preserved, which is another main principle used in intervention studies.

Even though a randomized intervention is the best strategy for studying treatment effects, results of such studies may still be misleading for a number of reasons. For instance, one main assumption is treatment-unit-additivity, that results in the lack of interactive effects of treatment and background variables. But this may not hold, especially if an illness has a genetic component or if the same symptoms can be caused by two very different types of health status.

There may also be important unobserved variables that are intermediate between treatment and the finally measured outcome variable. In this case, the overall treatment effect computed by omitting all intermediate variables may deviate strongly from the effect that describes pathways from the treatment to the response (Cox & Wermuth, 2003). Without mentioning such potential drawbacks and how they may possibly be corrected, statistics is likely to not appear as a trustworthy science to insightful collaborators from other fields.

3 Studying Causality with Potential Outcome Models

The attempt to understand possible causal effects is the motivation behind much empirical research in the life sciences. Also, there appears to be agreement that models consistent with causal interpretations are advantageous. However, there has also been an attempt to distinguish probability models from what have been termed causal models.

There is a concise description by Lindley (2002) of the effects of assuming a notional intervention in counterfactual reasoning and the potential outcome model. Let R denote a response, C a potentially causal treatment variable, and B a background variable. Then the joint density of the three variables would, in general, be different without and with intervention. More precisely, by observing or seeing, we would obtain the density for the given order of the variables as

$$f_{see} = f_{R|CB}\, f_{C|B}\, f_B,$$

whereas by intervening with C or doing, actual or notional, the potential cause becomes decoupled from the past so that the joint density is changed to

$$f_{do} = f_{R|CB}\, f_C\, f_B.$$

Both are proper probability models. In the linear case, an overall effect of C on R is $\beta_{R|C} = \beta_{R|C.B} + \beta_{R|B.C}\beta_{B|C}$ under the first, in contrast with $\beta_{R|C} = \beta_{R|C.B}$ under the second. Thus, assuming treatment-unit-additivity and a notional intervention a situation is produced in which the conditional effect of C on R given the background features B coincides with the overall effect of C on R. This is an important insight but hardly justifies speaking of a causal model, also claimed by some to be outside common probabilistic reasoning. For a detailed review of different statistical notions of causality, see Cox & Wermuth (2004).

4 Graphical Markov Models

Graphical Markov models extend the notion of a linear data generating process as studied by Sewall Wright under the name of path analysis, to more general situations using the notion of conditional independence, so permitting joint responses, sequences of treatment and intermediate outcome variables and a set of background variables. Many common statistical models are special cases. A wide range of new results have become available recently (Lauritzen & Sheehan, 2003; Wermuth & Cox, 2004; Wermuth, 2005). They provide a flexible class of models permitting extensions, for example to time series and point processes.

Stepwise data-generating processes in terms of recursive univariate regressions in a general sense can be modelled, including interactive and nonlinear effects. In particular, this makes it possible to check whether some of the assumptions of the potential outcome model can reasonably be made for a given set of data.

5 The Open Source Computing Environment *R*

Arguably the most promising development regarding transparent computing was started with the concerted work of a group of statisticians from Canada and New Zealand (Ihaka & Gentleman, 1996) and is at present the widely respected *R* project, providing open source software for many traditional statistical methods including Bayesian inference, graphical Markov models and genetic analyses. Access is via http://cran.r-project.org .

6 Other Topics

Intensive work on the decoding of genes has also led to the many studies involving statistical inference. It is particularly encouraging that statistical principles of design are being adapted to this special important area (Yang & Speed, 2002; Glonek & Solomon, 2004). Undoubtedly the next step will be a critical assessment of the properties of methods and algorithms, and discussion of permissable target populations.

Such assessments have been carried out for performance indicators (Bird *et al.*, 2005), the use of different weighting schemes in survey research (Firth & Bennett, 2002), ecological inference (Wakefield, 2004), and hierarchical modeling (Heagerty & Zeger, 2000; Firth, 2005). Many more are needed, especially with the rapid increase of methods proposed and implemented in statistical packages.

References

Armitage, P. (2003). Fisher, Bradford Hill, and randomization. *Int. J. Epidemiology*, **32**, 925–928.

Bird, S.M., Cox, D.R., Farewell, V., Goldstein, H., Holt, T. & Smith, P. (2005). Performance indicators: good, bad and ugly. *J. Roy. Statist. Soc. A*, **168**, 1–27.

Cox, D.R. & Wermuth, N. (2003). A general condition for avoiding effect reversal after marginalization. *J. Roy. Statist. Soc. B*, **56**, 934–940.

Cox, D.R. & Wermuth, N. (2004). Causality: a statistical view. *Int. Statist. Rev.*, **72**, 285–305.

Firth, D. (2005). Some topics in social statistics. In *Celebrating Statistics. Papers in Honour of Sir David Cox on his 80th Birthday*, Eds. A. Davison, Y. Dodge and N. Wermuth. Oxford University Press. To appear in October 2005.

Firth, D. & Bennett, K.E. (2002). Robust models in probability sampling. *J. Roy. Statist. Soc. B*, **60**, 3–21.

Glonek, G.F.V. & Solomon, P.J. (2004). Factorial and time course designs for cDNA microarray experiments. *Biostatistics*, **73**, 43–56.
Heagerty, P.J. & Zeger, S.L. (2000). Marginalized multilevel models and likelihood inference (with discussion). *Statistical Science*, **15**, 1–26.
Ihaka, R. & Gentleman, R. (1996). *R*: A language for data analysis and graphics. *J. Comp. and Graph. Statist.*, **5**, 299–314.
Lauritzen S.L. & Sheehan, N.A. (2003). Graphical models for genetic analyses. *Statistical Science*, **18**, 489–514.
Lindley, D.V. (2002). Seeing and doing: the concept of causation. *Int. Statist. Rev.*, **70**, 191–214.
Mosteller, F. & Baruch, R. (Eds.) (2002). *Evidence matters*. Washington, DC: Brookings Institution.
Sackett, D.L, Richardson, S., Rosenberg, W. & Haynes, R.B. (1996). *Evidence-based medicine: how to practice and teach EBM*. London: Churchill Livingstone.
Wakefield, J. (2004). Ecological inference for 2×2 tables. *J. Roy. Statist. Soc. A*, **167**, 385-445.
Wermuth, N. (2005). Graphical models: some history and recent developments. In *Encyclopedia of Behavioral Statistics*, Eds. B. Everitt and David C. Howell. New York: Wiley, to appear.
Wermuth, N. & Cox, D.R. (2004). Joint response graphs and separation induced by triangular systems. *J. Roy. Statist. Soc. B*, **66**, 687–717.
Yang, Y.H. & Speed, T. (2002). Design issues for cDNA microarray experiments. *Nature Reviews: Genetics*, **3**, 579–588.
Zeger, S.L., Diggle, P. & Liang K.-Y. (2005). Biostatistics: the near future. In *Celebrating Statistics. Papers in Honour of Sir David Cox on his 80th Birthday*, Eds. A. Davison, Y. Dodge and N. Wermuth. Oxford University Press. To appear in October 2005.

[*Received March 2005, accepted May 2005*]