ITAL J GASTROENTEROL HEPATOL 1998;30:580-3

REPORT

Chianciano Consensus Conference on Prognostic Studies in Hepatology

The Conference took place in Chianciano, Italy, from September 27 to 30, 1997. Lectures provided an overview of prognostic models (A. Morabito), validity of prognostic models (H.C. van Houwelingen), time-dependent versus time-independent models (P.K. Andersen), role of the neural networks and related statistical methods (K. Liestøl), use of quantitative tests (C Merkel), the natural history and prognosis of primary biliary cirrhosis (F.B. Bianchi), primary sclerosing cholangitis (L. Okolicsanyi), alcoholic liver disease (Dame Sheila Sherlock), ascites (M. Bernardi), chronic hepatitis (M. Rizzetto), hepatic granulomas (D.G. James), and on the prediction of death in cirrhosis (G. D'Amico), transplantation of the liver (J. Neuberger), expert systems (C.M. Leevy), conversion of empirical scores to prognostic indices (E. Christensen). Discussions followed each lecture in plenary sessions and in subgroups, aimed at consensus concerning the role of prognostic models in clinical practice and guidelines for future prognostic studies in hepatobiliary diseases. The following report integrates the issues and views raised during the consensus conference.

A primary task of the physician is to improve the patient's prognosis. Thus, prognostic assessment is most important in the evaluation of the patient. This will markedly influence the choice of therapy, and, therefore, extremely important to study prognosis and its determinants.

Definitions

Prognostic studies are intended here as investigations, in well-defined patient populations, on the relation between their characteristics (*variables*), at a given time, and the future course of disease with respect to the occurrence of a well-defined outcome (*end-point*). A prognostic study generally applies a regression analysis technique and results in a *prognostic model* which includes variables having a significant independent relation with the outcome variable. This includes variables found to be significant in the particular study, or variables known from other investigations to be of prognostic importance. The prognostic model shows how each variable in the model contributes to the prediction of the outcome. Since *a priori* knowledge and hypotheses may be incomplete, the analysis can have an exploratory or "heuristic" element, and *validation* of models using independent patient data will be necessary. Generally, a prognostic model can only provide an *approximate* prognosis in individual patients because it only explains a small part of the observed variation in outcome between patients. Biologic (genetic and environmental) variation is usually much wider than can be accounted for by any model. Therefore, prognostic models complement, but do not replace, careful clinical assessment of the individual patient.

Guidelines for future studies

Purpose of the study

The study should aim at improving our understanding of the disease by identifying determinants of its course and outcome. The prognostic information should be expressed in clinically useful terms, e.g. using prognostic indices or scores. Prognostic information is necessary for many reasons: 1) to inform the patient about the prospects for the future (this also allows for his informed consent to participate in studies), 2) to establish a more accurate diagnosis, 3) to evaluate types of technological support and to help in evaluating therapeutic intervention, 4) to adopt a rational decision-making approach concerning therapy, 5) to evaluate the contribution of a single prognostic factor amongst those which are partly controlled and partly out of controlled studies, 6) to design future randomized clinical trials, 7) to evaluate the prognosis of subgroups affected by varying severity of hepatobiliary disease. The use of prognostic indices can improve the description and allow comparison of patient groups. They are thus useful in communicating prognostic information.

The relationship between prognostic variables and the effect of therapy is particularly important. In statistical terms this is known as *interaction* between prognostic variables and therapy regarding their effect on outcome. This is best studied in randomized clinical trials. Important issues would be to predict decompensation (ascites, bleeding, encephalopathy) and survival in patients with liver disease, and to predict survival and optimal timing of liver transplantation in patients with decompensated cirrhosis.

Selection of patients

Well-defined *standardised, generally accepted diagnostic criteria* should be used to select the target patient group for the study. Other selection criteria should be as restrictive as possible to ensure the widest

possible applicability of the results. The time, place and routes of *enrolment* of the patients including the number of patients referred from other hospitals should be reported. *Consecutive* patients fulfilling the selection criteria should be included in the study. The *sample size* should be sufficiently large, adjusted to the number of expected events in the planned duration of the study and the number of putative variables to be investigated. Special attention should be given to the *follow-up* of the patients. This depends greatly upon the efficiency of the outpatient clinic. The follow-up should be complete, and the *dropping-out* should be reduced as much as possible. Pre-planned checks to avoid missing data and tracing of missing patients by phone should be undertaken.

Recording of variables
All variables necessary for the diagnosis and for a comprehensive description of the patients should be recorded. The highest priority should be given to easily obtainable variables with a high precision and accuracy. Variables implying a considerable observer variation, high cost, long duration (time delay) or inconvenience to the patient should only be used if deemed essential. All variables should be recorded in accordance with internationally accepted standards.

Definition of time and stage
Prognostic analyses rely on a common starting point: time zero. The definition of time zero is crucial and should, therefore, be considered very carefully. The *time of onset* may be unknown or ill-defined (e.g. chronic viral hepatitis C, alcoholic liver disease, primary biliary cirrhosis, primary sclerosing cholangitis). The patients may go to the doctor at markedly different stages of progression of the disease. Thus at the *time of diagnosis* the patients may present a very heterogeneous picture: some may be at an early stage, others at a late stage. This heterogeneity may complicate the statistical analysis and reduce the possibility of obtaining a useful prognostic model. Therefore, attempts should be made to reduce heterogeneity by synchronising courses, defining time zero as *the time of entry into a well-defined stage*. Following an insidious onset, many liver diseases (e.g. chronic hepatitis C, primary biliary cirrhosis, alcoholic liver disease) may have a gentle course with mild symptoms and laboratory abnormalities for long periods of time, with minor fluctuations but no or few signs of progression. Sooner or later this quiet phase may be followed by progression leading to decompensation and death within a short period of time. For this reason, the signs of hepatic decompensation (e.g. oedema, ascites, jaundice, encephalopathy, gastrointestinal bleeding caused by portal hypertension) are particularly important markers defining the transition to the advanced stage where more intensive therapeutic measures will be necessary and where the prognostic determinants may differ from those in the early stage.
The disease stage studied should be clearly and explicitly defined, e.g. should the cirrhotic patients included be compensated or decompensated. Other stages may be defined in special situations. When patients from more than one stage are to be analysed simultaneously, complex multistage models should be used. However, it is generally preferable and simpler to make one prognostic model for each stage of the disease.

Definition of predictor variables and end-points
Candidate variables in the prognostic model should be specified in advance by the clinician. To ensure a meaningful evolutionary scientific progress, the highest priority should be given to variables with a previously established prognostic influence, e.g. bilirubin, albumin, prothrombin time or index, ascites, encephalopathy, age, gastro-esophageal varices, variceal bleeding, serum creatinine. Important prognostic variables identified in other studies may be kept in the model even if not significant in the particular analysis. Automatic selection procedures with the computer should not be used to select the variables.
The event considered as the *end-point* should be clearly specified *a priori*. In advanced cirrhosis, important end-points would be death or, in some cases, liver transplantation. In compensated cirrhosis, the appearance of the first signs of decompensation, hepatocellular carcinoma or death would be useful end-points. So would be, for chronic hepatitis, the development of cirrhosis, revealed by signs of portal hypertension (oesophageal varices, platelets below the reference values).
For the study of the early course of chronic viral hepatitis B and C, where natural end-points are very late attempts should be made to establish whether indicators of progression of disease as well as quantitative function tests may be identified and possibly combined to a surrogate marker end-point.
The occurrence of other events (e.g. special therapeutic procedures) during the course may influence the occurrence of the end-point. For example, in primary biliary cirrhosis, the occurrence of the primary end-point of death will be influenced markedly by a liver transplant. One useful method of analysis in this complex situation would be to consider death and transplantation as *competing end-points* and to use a competing risks model for the analysis.

Statistical analysis
The statistical analysis involved in the development of prognostic models is complex. Preferably *cooperative*

research teams including both clinicians and biostatisticians should be set up. The statisticians would benefit from a closer insight into practical clinical problems and the clinicians would benefit from the expertise of the statisticians in ensuring maximum validity of analyses and results.

Fitting a prognostic model

The *Cox regression model* for censored survival data is well established. For this reason and because it is flexible, it may be considered the model of choice, provided that the model assumptions can be met. These include proportional hazards for the entire period of follow-up and linear additive effects of the variables on the prognostic index. If the precise timing of the event is not important but only the risk in a defined short interval (e.g. 6 months) is of interest, the simpler *logistic regression model* may be used instead. If these models do not satisfactorily fit the data, the *accelerated failure time model* may be considered. A final flexible model class is provided by *neural networks*. However, since a number of parameters to fit in the internal layer of a neural network may be very large, the necessary amount of patient data sets must also be very large to avoid over-fitting. Furthermore, a neural network is essentially a complex "black box" which does not allow a clear insight into which inputs are useful and how the information is processed to predict the outcome.

Generally, models based on data at one time during the course, e.g. data at the beginning of decompensation, can only be expected to predict prognosis for a limited span of time ahead.

Short-term prediction of prognosis may be improved if follow-up information can be utilised as in the Cox regression model for *time-dependent variables*. The time-dependent model is well suited to model both deterioration and improvement. The latter is particularly important in diseases where the aetiologic agent may be eliminated as in alcoholic liver disease and chronic viral hepatitis. However, analyses involving time-dependent variables are considerably more complex. It is important to distinguish between scheduled and non-scheduled follow-up. Since the latter will tend to occur at times of deterioration, it will have a different meaning from the former. This should be taken into consideration in the analysis to allow a meaningful interpretation of the model. To obtain useful survival probabilities from prognostic indices based on standard time-dependent models, the development of the prognostic variables over time also needs to be modelled.

For all models, *over-fitting* should be avoided. Overfitting occurs when random peculiarity ("noise") in the sample is also modelled. If overfitting has taken place, prediction with this model will be poor in new patients. To minimize the risk of overfitting, the variables to be included in the model should follow the priorities specified by the clinician, and the total number of parameters to be fitted should not exceed about 10% of the number of patients reaching the end-point in the patient sample.

Validation of prognostic models

Even if relevant measures have been taken to avoid overfitting, any statistical model must be especially well adjusted to the patient sample used for its development. The observed survival in new high risk or low risk patients will tend to be closer to the average than predicted by the model. This "regression toward the mean" effect should be examined and adjusted for in the validation procedure. Furthermore, the value of a prognostic model must have been confirmed in independent patients, preferably from several other centres, before its use in clinical practice.

Clinical use of prognostic models

Since prognostic models generally only explain a small part of the variation in survival between the patients, they *cannot predict prognosis precisely in individual patients*. Thus, for an individual patient, the prognostic scores can only provide a guide to prognosis.

When applied repeatedly in a same patient, a prognostic index based on time-dependent variables may be useful in estimating the *change in short-term prognosis* during the course of disease. Such information may be valuable for the *timing of therapeutic options* such as liver transplantation in primary biliary cirrhosis or in fulminant hepatic failure. In these cases, the updated prognostic index can serve as a *guide to therapy*. A prognostic index is a summarized description of the severity of the disease expressed in one number, obtained as a combination of all the prognostic variables weighted according to their relative prognostic influence. Therefore, the prognosis for a group of patients may be very well described by the average and distribution of their prognostic indices. The use of prognostic models and indices makes it possible to perform a more qualified comparison between different groups of patients by adjusting for imbalance in prognostic variables between groups. It is then possible to determine whether any difference can be explained by a different distribution in prognostic variables between the groups or must be attributed to other variables. This can be important when groups of patients from different centres are to be compared. Since a prognostic index includes those variables which are strongly associated with the progression of the disease, it can inspire further pathogenetic studies. For medical students and doctors under training, a prognostic index may be of *educational*

value in describing which variables are important as indicators of the course of the disease.
As survival analyses and prognostic studies become more common, they may be increasingly used by non-medical professions such as insurance companies, lawyers and administrators. The medical profession should be aware of this and should object to unqualified or unethical use. Examples of this are: 1) to consider the value of a given prognostic index as a precise estimation of the prognosis in a given patient, when, in fact, it is only an approximate indication of the course and outcome and must be considered together with other clinical information; 2) to perform crude comparisons of survival curves from different centres without adjustment for prognostic variables which may be distributed differently between the centres. Such prognostic adjustment is a necessary step in evaluating what lies behind the curves and in ensuring a fair comparison; 3) to perform crude comparisons of health expenses between centres. To ensure a fair comparison between the centres, adjustment should be made for differences in patient mix with regard to prognosis and necessary treatment.

Proposals for the future

At present, a large number of prognostic studies based on regression models have been made for many liver diseases. Nevertheless, the Child-Turcotte and Pugh scores are still the most widely used prognostic indices. Although their simplicity is attractive, they are not entirely satisfactory. They need to be succeeded by easy-to-use, up-to-date prognostic indices for the important diseases. Efforts should be made to agree on a limited number of simple prognostic indices. Commonly accepted prognostic models could be obtained by analysing combined data bases from various centres. However, models will need further adjustment in parallel with the acquisition of new knowledge in the future. One aspect could be to investigate how *quality-of-life* variables may be used and whether it is possible to predict *quality-adjusted survival*. Modern information technology allows a systematic, continuous, prospective accumulation of consecutive patient data in large data bases which can be utilised to develop better prognostic models. Prospective multicentre prognostic studies should be undertaken to establish whether heterogeneity between centres, countries and continents is real (different subtypes of disease) or can be explained by varying patient selection, resulting in a different patient mix from place to place.
The randomized clinical trial remains the best method for evaluation of new therapy. It allows analysis of the relationship between the prognostic variables and the effect of treatment, and, in particular, of whether the association of the prognostic variables with the outcome is modified by the therapy. If so, the predictive ability of the prognostic index could be changed and possibility invalidated by the therapy (example: ursodeoxycholic acid in primary biliary cirrhosis). Therefore, large multicentre randomized clinical trials should be performed not only to evaluate treatments, but also to study the interaction between treatment and prognosis (i.e., the therapy-dependent prognosis) in order to identify the characteristics of responders and non-responders. This may lead to the formulation of new hypotheses to be tested in subsequent randomized clinical trials. A pattern could thus emerge, allowing us to specify the best therapy according to the characteristics of the patient.
It is suggested that large multicentre studies should be supported, at least in principle, by International Organizations such as the World Health Organization and the International Association for the Study of the Liver. Such international studies would have limited costs if modern information technology were applied. Meetings focusing on prognosis in liver disease should develop globally accepted prognostic models and scores, reflected in guidelines and algorithms for the instruction of students and practitioners.

The Conference included the following members: P.K. Andersen, Copenhagen, Chair; M. Bernardi, Bologna; F.B. Bianchi, Bologna; J. Bircher, Herdecke; L. Bolondi, Bologna; E. Christensen, Copenhagen, Chair; G D'Amico, Palermo; H.C. van Houwelingen, Leiden; G. James, London; A.-M. Jézéquel, Ancona; K. Liestøl, Oslo; C.M. Leevy, Newark, Chair; C.B. Leevy, Newark; E. Marubini, Milano; C. Merkel, Padova; A. Morabito, Milano, Chair; J. Neuberger, Birmingham; L. Okolicsanyi, Parma; F. Orlandi, Ancona, Chair; L. Pagliaro, Palermo; M. Rizzetto, Torino; Dame Sheila Sherlock, London; J. Terblanche, Cape Town; N. Tygstrup, Copenhagen, Chair; R.K. Zetterman, Omaha. None of the participants received an honorarium. Travel and lodging expenses were covered by "Terme di Chianciano SpA", Italy.