Monitoring the effectiveness of treatments for alcoholic liver disease involves the use of variables that have prognostic significance and are unaffected by the treatment in unspecific ways. Here we review the value of histological and functional variables for this purpose. We conclude that histological variables, although important in defining the characteristics of the sample, have several practical problems. The functional variables are most effective when used in combinations, e.g., in global indices such as the Combined Clinical and Laboratory Index, the Child–Turcotte–Pugh Index, or the Cox model. In situations involving mortality and dropouts, functional indices cannot be used to measure changes in severity; in such cases, mortality might be the only measure for assessing the effectiveness of a treatment. In clinical trials, it is essential to determine the risk of a Type II error, to monitor compliance and drinking, and to trace appropriately all the patients who were not compliant or who dropped out of the trial.

In essence, monitoring treatment of alcoholic liver disease entails assessing whether the patients are improving or deteriorating. This is accomplished by using markers related to the severity or the prognosis of the disease. In individual patients the prognostic indicators are used not only to assess recovery or deterioration, but also to assign treatment modalities; more recently, these indicators have become important in selecting patients for whom liver transplantation offers the only chance of survival. The prognostic indicators in patients’ samples are essential for assessing treatments in clinical trials, both to compare the initial severity of the placebo and active drug treatment groups and to stratify the patients into similar disease severity groups. The indicators can also be used as a standard of comparison between patients’ samples from different centers and to determine the prognostic significance of different variables or of new liver tests.

In this review we will focus on the problem of assessing the effectiveness of treatments in clinical trials. For this purpose, the change in prognosis as a measure of the effectiveness of treatment can be assessed by using (a) histological variables; (b) functional variables, including portal hypertension; and (c) mortality.

Prognostic Value of Histological Variables

Morphologically, alcoholic liver disease has classically been divided into three strikingly different but frequently co-existing categories: (a) fatty liver, (b) alcoholic hepatitis, and (c) cirrhosis (1). Cirrhosis, defined as an alteration of liver architecture characterized by the presence of parenchymal nodules surrounded by connective tissue septa (2), has been considered the terminal, irreversible stage of liver disease (3, 4). Moreover, Popper has claimed that “the two criteria which correlate best with the functional manifestations are parenchymal nodules and septa which link portal with central canals” (2). Surprisingly, and contrary to the above concepts, the functional and prognostic consequences of these histological categories are far from obvious. Except for the clinical manifestations of portal hypertension, which are substantially more frequent in cirrhosis, the other clinical and laboratory variables have a very limited value in differentiating these three morphological categories (5).

Despite being considered the most advanced and irreversible histological stage of alcoholic liver disease, cirrhosis can be completely asymptomatic (5–8). Liver biopsies from 9% of chronic alcoholics without any clinical or laboratory evidence of liver disease demonstrate cirrhosis (5), and 50% of cirrhotics have only minimal or moderate clinical or laboratory evidence of liver disease (5).

From the above, it would appear that the presence of parenchymatous nodules surrounded by thick fibrous septa does not per se cause a serious impairment of liver function. In fact, cirrhosis not accompanied by alcoholic hepatitis has a one-year mortality of 7%, i.e., not greater than that of patients presenting only fatty liver on liver biopsy. In contrast, when cirrhosis is accompanied by alcoholic hepatitis, the mortality is 27% (9).

Epidemiological data from Pequignot and Cyrulnik (10) show that in Paris the rationing of wine during the war was followed by a 42% decrease in mortality from cirrhosis. After the war, when normal alcohol consumption was resumed, mortality from cirrhosis increased by 78% in one year. These changes in mortality rate cannot be attributed to changes in the prevalence of cirrhosis because this condition develops over many years of alcohol abuse. More probably the explanation for these abrupt changes in mortality from cirrhosis after variations in alcohol intake are related to the concomitant presence of alcoholic hepatitis, a lesion that would be expected to respond more rapidly than cirrhosis to...
changes in alcohol consumption.

The importance of alcoholic hepatitis as a determinant of prognosis was confirmed when we analyzed several of the histological variables both with regard to their relative risk for mortality and by multivariate analysis (9). Both fatty liver and inactive cirrhosis had relative risks <2, a risk that represents no significant effect on the chances for survival (Figure 1). On the other hand, necrosis (relative risk =4.4), Mallory's hyalin (relative risk =3.4), and inflammation (relative risk =3.1) were significantly related to the risk for mortality (9).

When we assessed the independent prognostic value of the histological variables by a multivariate methodology (stepwise discriminant function analysis), necrosis was the only independent predictor of mortality. None of the other variables, including cirrhosis, contributed to the prediction independently of necrosis (unpublished data).

The predictive accuracy of a variable can be illustrated by plotting the sensitivity and the specificity of increasing grades of severity of the variable. This yields what is called a receiver-operator characteristic (ROC) curve (11). Figure 2 shows the ability of necrosis to discriminate between those who died (specificity) and those who survived (sensitivity). Adding a score for cirrhosis did not improve the prediction capacity: the maximum discrimination point [(sensitivity + specificity)/2] was 0.65 for necrosis and 0.68 for necrosis and cirrhosis.

The presence of alcoholic hepatitis is strongly associated with enlargement of hepatocytes and accumulation of collagen in the space of Disse (9, 12). These abnormalities can produce both portal hypertension and liver dysfunction (13-17). Evidence suggests that both an increase in collagen in the space of Disse and an increase in hepatocyte size can compress the sinusoids (18) and thus reduce liver perfusion and oxygen supply to hepatocytes (19). Also, portal hypertension may result from the increasing resistance to portal blood flow in the sinusoids. In fact, hepatocyte enlargement and collagen in the space of Disse have been related to portal pressure in patients with alcoholic liver disease (14, 15, 17, 20). Moreover, accumulation of collagen in the space of Disse, the microenvironment from which exchange between hepatocyte and circulation takes place, can potentially interfere seriously with this exchange process, which is vital for normal liver function (21).

Liver biopsies can provide important information in clinical trials involving alcoholic liver disease patients. This procedure allows (a) the diagnosis, or exclusion, of other non-alcohol-related liver pathologies such as hepatitis, hemochromatosis, Wilson's disease, α1-antitrypsin deficiency, viral or toxic hepatitis, etc.; (b) stratification of patients into histological categories (i.e., with and without alcoholic hepatitis); (c) comparisons between the placebo and active drug groups; and (d) comparisons between samples from patients from different centers.

On the other hand, the use of histological variables to assess the effect of a treatment on morphological abnormalities poses a series of important problems derived from the inherent limitations of the procedures currently being used to obtain liver biopsies. Although the initial biopsy is an accepted diagnostic procedure, the assessment of changes with treatment requires performing two or more successive biopsies, a fact that could present an ethical problem, because the second biopsy would not necessarily be a clinical requirement. This could be particularly problematic either in those who become completely asymptomatic or in patients who deteriorate in ways in which a liver biopsy would impose an added risk.

In many medical centers, liver biopsies are performed only by the percutaneous route, a procedure that requires the patient to be capable of acceptable coagulation (in general, a prothrombin time prolonged <4 s over that of controls). In our experience, such patients have a mortality rate of 1.5% in 45 days, whereas those with prothrombin time >3 s longer than that of the controls have, over the same period, a mortality of 23% (5). Thus, the absolute requirement of liver biopsies as a
criterion for admission to trials in these centers, even when used for diagnostic or stratification purposes, would impose a serious selection bias by excluding a group of patients who have a more serious prognosis. Although the use of transjugular liver biopsies could solve this problem (22), such biopsies are performed only in selected hospitals and require a more invasive procedure, which the patients may not consent to have repeated.

Another problem with liver biopsies, especially in the diagnosis of cirrhosis, is that they are susceptible to sampling errors. In four reported series in which two or more liver samples were obtained in one session, 30% of the samples did not show the cirrhosis that was evident in one or more of the other biopsy specimens from the same patient (23–26). Moreover, in one study in which cirrhosis was observed during laparoscopy, 52% of the liver biopsies performed by the percutaneous route failed to confirm the diagnosis (26).

Determining the Prognostic Value of Functional Variables

Prognosis in alcoholic liver disease can be assessed by using clinical and laboratory variables of known prognostic value, either individually or in combination (global indices of severity). The individual variables that are significantly related to mortality can be identified and the strength of the relationship measured by univariate analysis. The independent contribution of each abnormality in the prediction of the risk for mortality in relation to the other variables can be determined by multivariate analysis.

Prognostic Value of Individual Variables

Univariate analysis. One method for determining the prognostic weight of individual variables is to calculate the relative risk for mortality with individual variables. This is achieved by comparing the mortality of patients that have the abnormality with the mortality of those who do not present the abnormality (27). A relative risk of 1 means that the patients with the abnormality have the same mortality as those without the abnormality. In our sample, only abnormalities with a relative risk ≥2 had prognostic significance.

In a sample of 253 patients, of whom 51 (20%) died in one year (28), the clinical abnormalities that had a prognostic weight to predict mortality were encephalopathy, ascites, collateral circulation, peripheral edema, and the presence of >10 spider nevi (Figure 3). The abnormality with the highest prognostic weight was collateral circulation (3+), with a relative risk of 6. In contrast, the prognostic weights for hepatomegaly, gynecomastia, and palmar erythema were not significant. However, in our series, splenomegaly was not accurately measured because ultrasonography was not routinely used.

Analysis of some laboratory variables that are frequently used in assessing alcoholic liver disease (Figure 4) shows that the most important prognostic abnormalities are albumin <25 g/L and bilirubin >136 µmol/L, each with a relative risk of 8. Hemoglobin <75% of normal and prothrombin time >8 s longer than the control time also had very significant prognostic weights. Aspartate aminotransferase, alanine aminotransferase, and γ-glutamyltransferase.

In this review, we have not attempted to analyze all the proposed markers of prognosis in alcoholic liver disease. We have limited our discussion to the variables identified above, the ones most commonly used in medical centers around the world. Many other endogenous markers of liver dysfunction have been proposed as good predictors of prognosis in alcoholic liver disease, e.g., creatinine (29), blood urea (29–32), bile acids (33), triiodothyronine (34), acetylcholinesterase (35), plasma tumor necrosis factor (36, 37), serum gamma globulin (29, 38), etc. Glutamate dehydrogenase was initially thought to reflect severity in alcoholic liver disease (39), but this has not been confirmed (40).

Liver function has also been measured by the use of exogenous substances cleared by the liver. Among them, aminopyrine breath test (41, 42) and indocyanine green (43–45) have been shown to be good predictors for mortality. Although galactose and antipyrine clearance tests reportedly relate well with liver function (46–48),
some investigators found that these do not relate significantly to the risk for mortality (43, 49).

Several markers of connective tissue metabolism, e.g., aminoterminal propeptide of type III collagen and type IV collagen, and markers of basement membrane formation (laminin), appear to have prognostic value. Indeed, these markers reportedly correlate significantly with the severity of liver disease (50, 51), with the histological severity of alcoholic hepatitis (50–53), and with the amount of alcohol consumption (50).

Portal hypertension is the main determinant of some of the most important causes of death in patients with alcoholic liver disease (43, 45, 49). Clearly the presence of enlarged esophageal varices with cherry-red spots has prognostic significance (54, 55). An increase in portal pressure is associated with an increase in mortality. We found that the one-year mortality was 0% in 40 patients with portal pressures <10 mmHg, 10% in 52 patients with portal pressures of 10–20 mmHg, and 30% in 56 patients with portal pressures >20 mmHg (P < 0.0001) (unpublished observation). Conversely, portal hypertension can occur with minimal abnormalities in liver function. For instance, in clinical trials of therapy directed to prevent recurrent bleeding esophageal varices, as many as 72% of the patients had only slight evidence of liver dysfunction (Pugh's severity category A) (56).

Multivariate analysis. Most individual indicators of severity are highly interrelated. Thus, although significantly related to mortality when analyzed alone, many variables have no independent relation to prognosis when analyzed in combination with other abnormalities. Multivariate analysis determines the independent contribution of a given abnormality to predict outcome. For situations in which the lengths of follow-up are similar, discriminant function analysis or logistic regression can be used. When the lengths of follow-up vary because of dropouts or late entry, the Cox proportional hazards model is particularly helpful, especially in studies with extended follow-ups of more than one year.

To determine whether different multivariate methods select different independent variables, we analyzed by the above three methods the same sample of 253 patients with a one-year follow-up. From the 12 variables indicating significant relative risks (Figures 3 and 4), only encephalopathy, albumin, hemoglobin, and prothrombin time were selected by discriminant analysis as being independently related with the one-year mortality (28). Logistic regression, which measures on a log scale the separation of the surviving patients from those that died, selected only three of these variables: encephalopathy, albumin, and hemoglobin (28). The independent variables selected by the Cox proportional hazards model were encephalopathy, albumin, hemoglobin, and bilirubin (unpublished observation).

Methods Involving Combinations of Variables (Global Indices of Severity)

The "intact cell" theory of liver damage assumes that liver function deteriorates globally, with every function decreasing in parallel. Theoretically, it is difficult to conceive that any individual variables could really represent a global measure of the severity of liver disease (57, 58). Because chronic liver disease apparently affects differently the functions of synthesis, detoxification, and excretion as well as hepatic blood flow and liver microcirculation, it is unlikely that a single test could provide an overall estimate of total liver function (59). Moreover, serious complications, of significant prognostic value, e.g., the consequences of portal hypertension, can occur quite independently of liver function.

The global indices of severity overcome many of the possible pitfalls of individual variables. However, these indices must comply with certain requisites such as the need to include only variables of known prognostic weight that are susceptible to repeated measurements. This requirement makes impractical indices involving portal pressure measurements and histological samples, because they require invasive procedures, which are not possible for all patients.

Global methods based on clinical experience. Global methods based on clinical experience are best exemplified by the Child—Turcotte—Pugh Index (60). Initially, Child and Turcotte defined three categories (A, B, and C), using subjective variables: e.g., the classification of neurological disorders in none, minimal, or advanced coma; classifying the nutritional state as excellent, good, or poor; and grading ascites as none, easily controlled, or poorly controlled. Moreover, the cutoff points for albumin and bilirubin were set arbitrarily (61). Pugh substantially changed this index to include the variables encephalopathy, ascites, prothrombin time, albumin, and bilirubin (Table 1). Each of these variables was divided into three degrees of severity, with a range of scores for each, from 1 to 3. This scoring system has the advantage of effectively dividing severity into three different categories: A = scores 5 to 6, B = 7 to 9, and C = 10 to 15. The Pugh modification has recently been validated by both Cox regression and logistic regression as a useful method to estimate prognosis for patients with cirrhosis of various etiologies (62, 63).

Indices based on relative risk (univariate analysis). The Combined Clinical and Laboratory Index (CCLI) is

<table>
<thead>
<tr>
<th>Table 1. Child—Turcotte—Pugh Score of Severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Encephalopathy</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1–2</td>
</tr>
<tr>
<td>3–4</td>
</tr>
<tr>
<td>Ascites</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1–2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>Prothrombin, s*</td>
</tr>
<tr>
<td>&lt;4</td>
</tr>
<tr>
<td>4.0–6.0</td>
</tr>
<tr>
<td>&gt;6.0</td>
</tr>
</tbody>
</table>

*No. of seconds longer than time for control sample.
a good example of indices based on relative risk. It originally included 12 clinical and laboratory variables, each of which had previously been shown to have prognostic weight in predicting mortality in patients with alcoholic liver disease (28). The sum of the prognostic weights of each variable with a range for each of 0 to 3+ (derived from a logarithmic division of the relative risks) results in an index that relates to mortality. This index was first analyzed retrospectively and later validated prospectively (64). The original CCLI included weakness, anorexia, and the presence of >10 spider nevi, because these variables had a significant prognostic weight. Weakness and anorexia have now been deleted because their evaluation is subjective. The counting of spider nevi has also been eliminated, being impractical and subject to error. Re-evaluation of the new CCLI, without these three items, showed that this modification did not reduce its prognostic accuracy. The CCLI with nine items is shown in Table 2.

**Global indices based on multivariate analysis.** Multiple regression techniques have been used to determine which prognostically significant variables are required to best predict risk of mortality in the patients studied. Because of the high interdependence between the indicators of severity, usually multivariate analysis selects a small number of tests. An index can be derived by using the regression coefficients for each of the selected variables. For example, Maddrey et al. (65) used discriminant function analysis in a clinical trial of treatment of alcoholic hepatitis with corticosteroids. Two variables were selected as having significant independent association with the short-term (five-week) mortality. The equation for this sample was prothrombin time (s) $\times 4.6 + \text{bilirubin (}\mu\text{mol/L)} \times 17$. For this specific trial, a possible problem in selecting these two variables is that, as explained below, the concentration of bilirubin can be decreased by steroids through a mechanism that appears to be independent of a change in liver function.

Others have shown that indices derived from the Cox regression model accurately predict the outcome of patients with alcoholic liver disease (32, 35, 45). Applying this method to our sample of 253 patients yielded the following expression: encephalopathy $\times 0.48 - \text{albamin (g/L)} \times 0.095 - \text{hemoglobin (}\%\text{ of normal) } \times 0.017 + \log \text{bilirubin (}\mu\text{mol/L)} \times 0.63$.

**Comparison between the different global indices of severity.** Figure 5 shows the Child–Turcotte–Pugh, the CCLI, the Cox regression, and the Maddrey indices, plotted according to sensitivity and specificity for predicting mortality (ROC curve) in the same sample of 253 patients. The indices have similar overall predictive capacities, except for Maddrey’s index, with maximum discrimination values of 0.76, 0.80, 0.78, and 0.71, respectively. The lesser accuracy of Maddrey’s index (Figure 5) highlights the difficulty of applying a specific regression equation derived from a highly selected sample of patients with increased bilirubin and (or) encephalopathy (65).

The Cox regression analysis can be used to study the relationship between the global indices of severity and the actual mortality in the patient population over time (66). This method allows for separately estimating the cumulative probability of survival for each patient and thus can produce a mean estimated survival curve for the sample (67). With this method, the indices of prognosis can be divided into several categories of severity, and the mean estimated probability of survival of these groups can be compared with the actual observed survival curves. Figure 6 shows the comparison of the actual cumulative survival with the mean predicted survival curves of three risk groups derived from the CCLI, the Child–Turcotte–Pugh, and the Cox indices. For this study, we used a sample of 445 patients with alcoholic liver disease, followed for a minimum of one year and a maximum of five years. The Cox regression index, derived retrospectively from this sample, was encephalopathy $\times 0.49 + \text{collateral circulation} \times 0.24 - \text{albumin (g/L)} \times 0.064 + \text{prothrombin time (s, patient – control)} \times 0.19$. Using the index to estimate a survival curve for the same patient population from which it was derived provides the greatest capacity for prediction of the observed mortality. However, as was done with the

---

**Table 2. Combined Clinical and Laboratory Index**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Score</th>
<th>Variable</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encephalopathy</td>
<td>2</td>
<td>Hemoglobin, % of normal</td>
<td>1</td>
</tr>
<tr>
<td>1–3</td>
<td>2</td>
<td>&lt;75</td>
<td>3</td>
</tr>
<tr>
<td>Collateral circulation</td>
<td>1</td>
<td>Albumin, g/L</td>
<td>3</td>
</tr>
<tr>
<td>1–2</td>
<td>3</td>
<td>25–29</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>&lt;25</td>
<td>3</td>
</tr>
<tr>
<td>Ascites</td>
<td>2</td>
<td>Bilirubin, µmol/L</td>
<td>2</td>
</tr>
<tr>
<td>1–3</td>
<td>2</td>
<td>&gt;136</td>
<td>3</td>
</tr>
<tr>
<td>Edema</td>
<td>1</td>
<td>Alkaline phosphatase, U/L</td>
<td>2</td>
</tr>
<tr>
<td>2–3</td>
<td>2</td>
<td>&gt;330</td>
<td>3</td>
</tr>
<tr>
<td>Prothrombin, s*</td>
<td>2</td>
<td>Score range = 0–22</td>
<td></td>
</tr>
<tr>
<td>4–5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*See footnote in Table 1.*

---

**Fig. 5. Comparison of the sensitivity and specificity of the Child–Turcotte–Pugh Index, the CCLI, Maddrey’s Index, and the one-year Cox regression index for the prediction of mortality (ROC curve) See Fig. 2 for explanation. Data from Orrego et al. (28); n = 253**

---

**CLINICAL CHEMISTRY, Vol. 37, No. 1, 1991**
CCLI and with the Child–Turcotte–Pugh Index, a prospective analysis is required to validate the accuracy of the model. Notwithstanding this difference, both the CCLI and the Child–Turcotte–Pugh Index are entirely comparable with the Cox model index. As Figure 6 shows, the three indices clearly discriminate groups of patients with significantly different prognoses and accurately predict the observed mortality in each risk group.

The value of liver tests based on hepatic metabolism or excretion of a single exogenous substance such as aminopyrine, indocyanine green, or antipyrine as independent predictors of mortality has been analyzed by using the Cox regression model. According to this method, these quantitative liver-function tests have less predictive value than do the Child–Turcotte–Pugh Index or the CCLI and offer no additional independent prognostic information (42, 43, 49). This further demonstrates that single tests of liver function cannot possibly replace global indices of severity. These indices combine several variables that relate to liver function or complications such as portal hypertension. Of course, the individual variables have important prognostic weight, but do not necessarily represent a global and parallel decrease in liver function.

Use of Functional Variables in Clinical Trials

Global indices based on functional variables are essential for assessing initial severity in clinical trials, which allows the stratification of patients into groups of severity and provides comparisons between the placebo and the active-drug groups. Also, the comparability of the disease severity among the groups of patients in clinical trials from different centers can be assessed. This information would facilitate studies involving meta-analysis.

For trials in which no patients are lost through dropout or mortality, the functional indicators of prognosis are the best criteria for assessing the effectiveness of treatment.

In addition, a good indicator of severity should not be influenced by treatments via effects that are not dependent on changes in liver function. An example of this problem is bilirubin. Bilirubin, through a mechanism that has not been clearly defined, can be decreased by corticosteroids, even in patients with complete biliary obstruction from pancreatic carcinoma (68). These hormones can also produce an unspecific increase in the concentration of serum albumin (69). Also, prognostic indicators such as triiodothyronine could show an apparent change in patients treated with propylthiouracil.

Global indices of severity have the advantage of providing in a single number a score that gives an estimate of the risk for mortality. These indices are also less susceptible to changes attributable to mechanisms unrelated to liver function.

The variables selected by multivariate models can be quite different, depending on the selection criteria of the patient population. Ideally, in clinical trials, one should calculate the relationship of the variables for each sample rather than use previously published equations, especially for patients with heterogeneous etiologies of liver disease.

For clinical trials in which patients are lost to mortality or incomplete follow-up, the use of a global index or individual functional variables to monitor recovery entails a serious methodological problem. Very commonly, the results are presented in graphs in which the mean scores are plotted over time and the means of the placebo and active-drug groups are compared to define the efficacy of the treatment. This procedure does not take into consideration the fact that both mortality and dropouts affect the mean values unpredictably. In essence, analysis of the functional variables as mean values for the group requires the same number of patients initially and finally, and also the same time of observation for every patient. This requirement is especially relevant in long-term trials.

Use of Mortality to Assess Clinical Trials

For clinical trials in which patients are lost to follow-up through mortality or dropout, the only way to properly analyze the effect of a treatment is by determining its impact on mortality. The use of mortality as the criterion for effectiveness in clinical trials requires avoiding the possibility of a Type II error, i.e., obtaining nonsignificant effects because of the inadequate size of the sample. For example, in a disease such as alcoholic liver disease, with a one-year mortality of ~20%, if one is to avoid a Type II error with an 80% power, a study involving a treatment that reduces mortality by 50% would require 438 patients (70).

In smaller samples, to avoid a Type II error, researchers have been forced to use groups with very high mortality rates. The flaw with this procedure is that such types of samples will necessarily include patients with many serious complications, e.g., spontaneous bacterial peritonitis, hepatorenal syndrome, bleeding esophageal varices, and encephalopathy—a situation that could certainly obscure the effects of therapies because a treatment effective in the basal alcoholic liver disease might not affect patients with complications such as the hepatorenal syndrome. On the other hand,
Follow-up Requirements for a Long-Term Clinical Trial

To avoid possibly biasing results by sporadic loss or withdrawal of patients, one must attempt to minimize this effect. Patients should be entered into the randomization process only after diagnosis is confirmed and the trial admission criteria are satisfied (71). Alcoholic patients being treated for a long time tend to have problems maintaining compliance. It is essential to determine whether the number of noncompliant or dropout patients are similar in the active-drug and placebo groups, and whether their reasons for dropping out, their severity of disease, their mortality rate, and their rate of dropping out throughout the study are similar. This determination entails a rigorous tracing of patients after they have become noncompliant or have dropped out. Failure to do so can introduce serious biases into the statistical analysis of the results of the treatment. For example, in a clinical trial on the effectiveness of treatment with propylthiouracil on alcoholic liver disease, the investigators monitored compliance by adding riboflavin to the capsules and measuring fluorescence in daily mailed urines. Also in this trial, patients who did not comply by taking the capsules and those who dropped out of the study were traced for at least two years after they abandoned the study (72).

Assessment of alcohol intake is extremely important because of its possible effects on the treatment outcome. In the alcoholic population with liver disease, complete abstinence continues to be a desirable but unrealistic goal. In the propylthiouracil study, 95% of the patients had at least occasional alcohol-positive urine samples (72). Self-reporting of alcohol intake is very unreliable for ascertaining abstinence. In a study of patients who provided daily urine samples, 50% of the patients whose urines tested positive for alcohol convincingly denied their alcohol consumption (73).

Although not perfect from the scientific point of view, determination of the concentration of alcohol in the urine provided each morning is still the best available method for assessing alcohol intake. Significant differences between patients drinking small amounts and patients drinking more heavily have been demonstrated with respect to improvement in clinical severity (CCLI), markers of connective tissue metabolism, and mortality in the patients being treated with propylthiouracil (50, 72).

Conclusions

The effectiveness of a treatment for alcoholic liver disease can be assessed by analyzing histological or functional variables and (or) the rate of mortality. Histologically, hepatocellular necrosis appears to be the best independent predictor of the risk for mortality. Because repeated liver biopsies are required for the use of histological variables, histology is an impractical tool for assessing the effects of treatments.

Several methodological procedures are available for determining the prognostic weight of various functional (clinical and laboratory) abnormalities commonly present in patients with alcoholic liver disease. These procedures allow the development of global indices of severity, e.g., the Child–Turcotte–Pugh Index, the CCLI, and the Cox regression, which provide a more effective overall measurement of the degree of liver dysfunction than do single-factor indices. These three severity indices have approximately the same value for predicting mortality. Maddrey's index, which depends on only two variables and is more susceptible to nonspecific effects of treatments, is less reliable; it was derived from a select sample of patients who might not be representative of the full spectrum of severity of the disease.

Functional variables have clear advantages over histological abnormalities for assessing the effectiveness of treatment in clinical trials. However, these methods for monitoring treatment are not perfect. They are extremely good for predicting an all or none (dichotomous) phenomenon, e.g., the risk for mortality; however, because the indices are not linearly related to the severity of the disease, assessing improvement in patients who survive continues to be a problem. Changes in the upper range of severity, as assessed by an index, probably have a different significance than changes in the lower range; e.g., the effects of a change from 20 to 15 in the CCLI scale may differ from those of a change from 5 to 0. This problem could introduce a serious artifact in trials that assess the effectiveness of a treatment by comparing the changes from the initial to the final scores of the global severity index in the drug and placebo patients. Future studies should be directed toward finding an index of liver dysfunction that more closely reflects a linear relationship with the severity of the disease. Such an index would incorporate variables that reflect changes occurring early in the disease (high sensitivity), as well as those associated with a high risk for mortality (high specificity).

The use of the functional indices of severity should be restricted to trials in which no subjects die or drop out; i.e., the number of patients should not change during the course of the study. Also, the time of observation should be similar for all patients. When these criteria cannot be met, mortality becomes the only accurate method for determining the effectiveness of a treatment. An accurate analysis of mortality rates should include determining the statistical power to avoid a Type II error, knowledge of the fate of dropouts, an accurate method to assess alcohol intake and compliance with treatment, and the ability to control for confounding factors such as differences in initial severity, age, sex, etc. Powerful statistical methods now available are
designed to minimize the effects of different times of follow-up and other factors that interfere with the evaluation of the effectiveness of a treatment to reduce the rate of mortality.

References


