Quantitative Liver Function Tests As Surrogate Markers for End-Points in Controlled Clinical Trials: A Retrospective Feasibility Study

ERICH LOTTERER,[1] JOSEF HÖGEL,[2] WILHELM GAUS,[2] WOLFGANG E. FLEIG,[1] AND JOHANNES BIRCHER[3]

SEE EDITORIAL ON PAGE 1678.

Quantitative liver function tests such as the determination of galactose elimination capacity (GEC) or the aminopyrine breath test (ABT) may have the potential to serve as refined entry criteria and surrogate markers for end-points in controlled clinical trials. The magnitude of a statistically detectable difference in test results and the period of observation required to document such a difference must be known to properly design such trials. Therefore, we explored retrospectively the time course of changes in GEC and ABT and their reproducibility from a cohort of patients with alcoholic cirrhosis followed for 12 to 42 months, with a median of 34 months. In 15 patients who stopped drinking, GEC improved significantly by 0.64 mg/min/kg within 1 year (mean; 95% confidence interval [CI]: 0.42; 0.86). In contrast, it deteriorated by 0.53 mg/min/kg within 1 year (95% CI: 0.32; 0.74) in another 17 patients who continued to drink ($P < .01$). The residual standard deviation of the changes in GEC with respect to the patients' initial values was 0.43 mg/min/kg (95% CI: 0.32; 0.52). In addition, ABT improved significantly by 0.14% dose · kg/mmol CO_2 (95% CI: 0.09; 0.18) in the abstinent group, and deteriorated by 0.09% dose · kg/mmol CO_2 (95% CI: 0.06; 0.13) in the nonabstinent group ($P < .01$). The residual standard deviation in the above sense for ABT was 0.08% dose · kg/mmol CO_2 (95% CI: 0.06; 0.10). These data indicate that clinical trials with a sample size of n = 20 in each group must achieve absolute differences (ADs) in GEC of 0.6 mg/min/kg and of 0.7 mg/min/kg to reach statistical significance at the 5% and 1% level, respectively. In the present study, a period of 11 and 12 months was necessary to observe such differences. The corresponding results for the ABT are 0.11% dose · kg/mmol CO_2 (9 months of follow-up; 5% level) and 0.13% dose · kg/mmol CO_2 (11 months of observation; 1% level), respectively. Provided that patients with liver diseases treated with drugs are similar to the abstinent and nonabstinent patients with alcoholic liver disease investigated in this study, such numbers could serve for the planning of controlled clinical trials, in which the control group is likely to deteriorate and the treated group is expected to improve. Trials based on GEC or ABT would require only 37 or 30 patient years of observation compared with a median of 444 patient years (range, 50-2,100 patient years) reported for various published controlled clinical trials using survival analysis. (HEPATOLOGY 1997;26:1426-1433.)

Controlled clinical trials for patients with liver disease have been difficult to design, cumbersome to perform, and have often failed or given conflicting results.[1] This situation generally is unsatisfactory and discourages therapeutic research. Several reasons may explain these problems:

1. Disease severity at the onset of the trial generally is not adequately defined. The Child-Pugh classification reflects late disease only,[2,3] and early disease is overlooked.[4,5] The result is a heterogenous study population and consequent noise that can be overcome only by very dramatic therapeutic effects or extremely large sample sizes. With better stratification of patients included in a trial, smaller therapeutic effects might be evident.

2. Death is often the end-point in clinical trials.[6-20] Although this is an important and objective measure of outcome, it can occur for varying reasons, e.g., upper gastrointestinal hemorrhage, renal failure, portal systemic encephalopathy, septicemia, hepatocellular carcinoma, and can be unrelated to the liver. It probably is quite inefficient to lump these differing events together and relate them to a single therapeutic intervention.

3. Progression of many liver diseases is slow and variable, often requiring years or decades to evolve to a final stage.[1] Thus, it is necessary to investigate a large number of patients for a long period of time to detect treatment-related effects with death as the end-point, a procedure that is often difficult or impossible and, at the least, expensive. The time period can be reduced by studying only the late stage of a disease, although prevention of late-stage diseases is a more appropriate target of therapeutic intervention.

These issues would be addressed in principle by the use of quantitative liver function tests to define the severity of disease at the start of a trial and the progression or regression as a result of intervention. While potentially ideal surrogate markers for this purpose,[21] such tests have been employed by very few clinical investigators,[22-27] and, in some cases,

Abbreviations: GEC, galactose elimination capacity; ABT, aminopyrine breath test; SCl, sorbitol clearance; AD, absolute difference; RD, relative difference; CI, confidence interval.

From the [1]First Department of Medicine, Martin-Luther-University Halle-Wittenberg, Halle; [2]Department of Biometry and Medical Documentation, University of Ulm, Ulm; and [3]Faculty of Medicine, University Witten-Herdecke, Herdecke, Germany.

Received December 20, 1996; accepted August 7, 1997.

Address reprint requests to: Wolfgang E. Fleig, M.D., First Department of Medicine, Martin-Luther-University Halle-Wittenberg, Ernst-Grube-Str. 40, D-06097 Halle (Saale), Germany. Fax: 49-345-5572253.

Copyright © 1997 by the American Association for the Study of Liver Diseases.

0270-9139/97/2606-0010$3.00/0

TABLE 1. Demographic and Clinical Data of the Patients

	Group A (Abstinent Patients)	Group B (Nonabstinent Patients)
n	15	17
Sex (m/f)	10/5	13/4
Age (yr)	57 (38-71)	55 (33-69)
Body weight (kg)	76 (59-95)	75 (50-100)
Child-Pugh score	7.0 (5-12)	6.5 (5-13)
Cirrhosis on liver biopsy	15	17
Signs of portal hypertension		
Esophageal varices	14	15
Upper gastrointestinal hemorrhage	11	12
Splenomegaly	9	12
Ascites	6	7
Jaundice	10	11
Number of return visits	5 (3-7)	5 (2-8)
Follow-up period (mo)	30 (12-39)	33 (24-42)

NOTE. Data are median (range). There are no statistically significant differences between the two groups.

have led to negative results.[28-32] Negative outcomes may be caused in part by uncertainty regarding the rate of change in liver function and by problems with the reproducibility of the tests.

Before a surrogate end-point may be used in clinical practice, its adequacy must be established for the clinical problem and interventions under study.[33,34] In particular, two standards should be met: first, the occurrence of the surrogate end-point is linked to the occurrence of the true end-point; second, the mechanism through which treatment influences the true end-point involves directly and at the same time the occurrence of the surrogate end-point.[34] The first Prentice's criterion is proven in several studies by showing a strict association of the decrease in quantitative liver function tests and survival.[35-37] The adherence to the second criterion is more difficult to prove, because every single treatment must be analyzed. Therefore, in the present study, we repeatedly applied quantitative liver function tests to patients with alcoholic liver disease. In view of the well-known effects of continued alcohol use and of abstinence on the liver,[38-40] abstinence from alcohol may be considered as a model for drug treatment in alcoholic cirrhosis. The data of our patients with alcoholic cirrhosis were analyzed with respect to the following questions: 1) What is the reproducibility of quantitative liver function tests and how large must differences between two treated groups be to become detectable by statistical methods? 2) If we assume that evaluation of a difference between a treated and a control group is similar to the course of alcoholics who stop and who continue to drink, how long a period of time is needed to detect such differences?

The goal of this retrospective analysis is to describe conditions for use of quantitative liver function tests as surrogate markers for hard end-points.

PATIENTS AND METHODS

Patients. From January 1986 to March 1991, 119 consecutive patients with alcoholic cirrhosis who agreed to undergo quantitative liver function tests were screened for inclusion into the study. Diagnosis was established by the following criteria: 1) history of alcohol consumption of more than 60 g/d for females and more than 80 g/d for males for at least 10 years; 2) routine laboratory tests compatible with alcoholic cirrhosis; 3) alcohol plasma concentrations >0.5‰ (10.9 mmol/L) at the beginning of the period of observation; 4) sonography with alterations indicative of cirrhosis without abnormalities of the extrahepatic bile ducts; and 5) liver biopsy (not necessarily current) showing cirrhosis compatible with chronic alcohol consumption in all patients.[42]

Thirty-two of these patients were followed for a minimum of 12 months. The follow-up was terminated in November 1993. Retrospectively, the patients were divided into two mutually exclusive groups (Table 1): group A: patients who stopped drinking (abstinent patients, n = 15) as reflected by history, including general well-being and social situation, and by negative plasma alcohol levels upon each return visit; group B: patients who continued to drink (nonabstinent patients, n = 17) as documented by alcohol plasma levels >0.3‰ (6.5 mmol/L) upon return visits. A single positive alcohol plasma level placed a patient in this group.

All patients were followed as cohorts of out-patients in the liver clinic of the Division of Gastroenterology and Endocrinology, University of Goettingen, and in the liver clinic of the Department of Medicine I, University of Erlangen-Nuernberg, by one of the authors (E. L.). They were regularly encouraged to stop drinking. Otherwise, clinical investigations and treatment modalities were adapted to the individual needs of each case. Median follow-up was 34 months (range, 12-42 months). The study was approved by the local ethics committees, and written informed consent was obtained from each participant in accordance with the Declaration of Helsinki, revised in Tokyo in 1975.

Investigations. All blood samples were taken in the fasting state. Routine biochemical and hematological examinations were performed with automated techniques, and immunoglobulins were assayed by radial immunodiffusion. Blood alcohol concentrations were determined enzymatically at every visit.

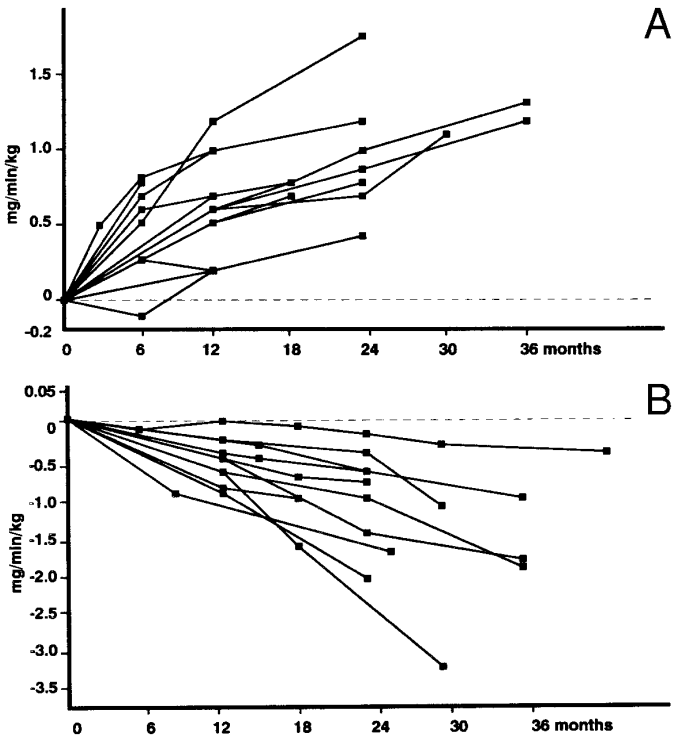

FIG. 1. Time course of differences of GEC between the initial value and the values at the time stated (AD) in (A) abstinent and (B) nonabstinent patients during the period of observation (individual polygon paths for all patients). Because of similar changes over time in a few patients, some points and lines represent more than one single patient.

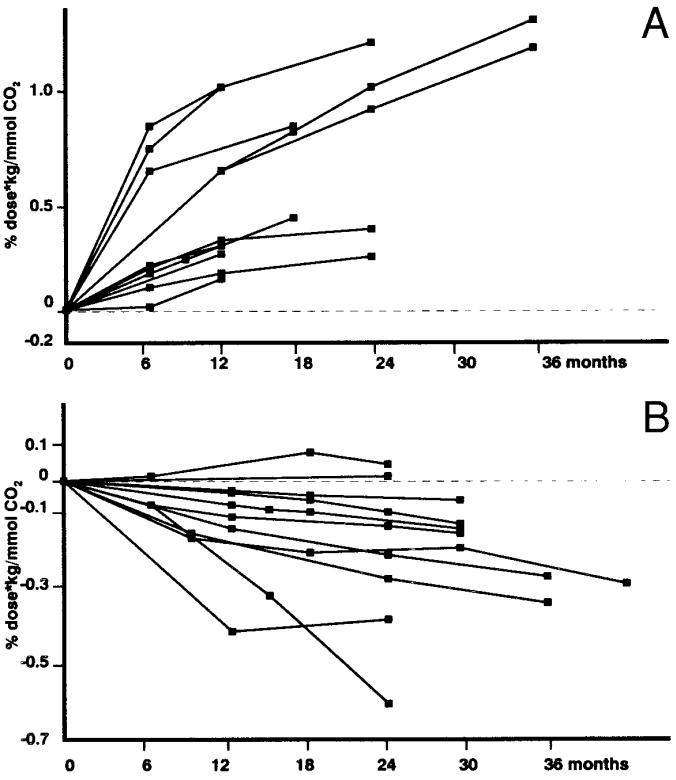

FIG. 2. Time course of differences of ABT between the initial value and the values at the time stated (AD) in (A) abstinent and (B) nonabstinent patients during the period of observation (individual polygone paths for all patients). Because of similar changes over time in a few patients, some points and lines represent more than one single patient.

Quantitative liver function tests were performed in each patient at the beginning of the period of observation and after irregular intervals lasting between 3 and 18 months.

The galactose elimination capacity (GEC) was calculated from venous samples taken every 5 minutes between 20 and 60 minutes after intravenous injection of 0.5 g/kg galactose as described by Tygstrup,[43] and a factor of 5 minutes was used for correction of uneven distribution of galactose in the body.[21] The normal value of this test is 7.0 ± 0.5 mg/min/kg body weight, with reproducibility tests in the same subjects giving a coefficient of variation of 4.1%. For the aminopyrine breath test (ABT), samples were examined 30 minutes after intravenous injection of 1 mg (1.5 μCi) dimethylamine-^{14}C-aminopyrine according to Miotti et al.[44] The range of normal is 0.65 to 1.00% dose · kg/mmol CO_2. Reproducibility was not tested for ethical reasons. Hepatic sorbitol clearance (SCl), which represents a measure of hepatic parenchymal plasma flow, was determined as described by Zeeh et al.[45] The normal value of SCl is 10.6 ± 2.1 mL/min/kg (mean ± 1 SD). Intraindividual retesting revealed a coefficient of variation of 7.6%. If alcohol plasma levels were above the limit of detection at the time of the quantitative liver function tests, the test results were rejected, because the presence of ethanol affects both GEC and ABT. If possible, the tests were repeated on a different day.

Statistical Analysis. For evaluation, the longitudinal data from each patient for GEC, ABT, and SCL were combined to a polygone path. The values for 12, 24, and 36 months were obtained by interpolation from the polygone paths.

Two outcome measures describing the progression of the disease were calculated from the data of the three quantitative liver function tests. Both measures are adjusted to the initial values of each patient at the beginning of the period of observation. First, the absolute difference (AD), which is defined as the difference of the test result at the stated time and the corresponding test result at the beginning of the period of observation ($t_x - t_0$), and secondly, the relative difference (RD), which is the AD in relation to the result at the beginning of the period of observation [$(t_x - t_0)/t_0$], were calculated.

Defining the corresponding variables, i.e., AD and RD for ABT, GEC, and SCl as trial end-points, a repeated-measurement ANOVA was separately applied to each of these variables. The factors of the ANOVA model and its levels were: 1) alcohol dependence in groups 1 and 2; 2) time points (i.e., after 12, 24, and 36 months); 3) interaction of alcohol dependence with time points; and 4) patients (1-32).

The effects of the above factors were assessed in preliminary statistical tests. An adjustment for multiple testing was not performed. For each group, the mean values for the previously defined end-points were calculated for the time points 12, 24, and 36 months, estimating quantitatively the group effects over time. Additionally, the 95% confidence intervals (CI) for these means are presented.

In a second step of analysis, the residual standard deviation of the variables of interest, i.e., the SD of the data after elimination of their variability due to group and time effects, was determined together with 95% CIs. To assess reproducibility of the data, the residual variance (=SD^2) was deconvoluted into parts reflecting variability within persons over time and between persons, respectively. Test results are reproducible if variance within persons is small compared with the variance between persons.

For purposes of determining the sample sizes necessary for detecting certain differences fixed in advance between the outcome of two groups in controlled clinical trials (using levels of significance α = 0.01 and setting the power [1 − β] = 0.9), the problem was reduced to a two-sample test scenario. In this context, the upper

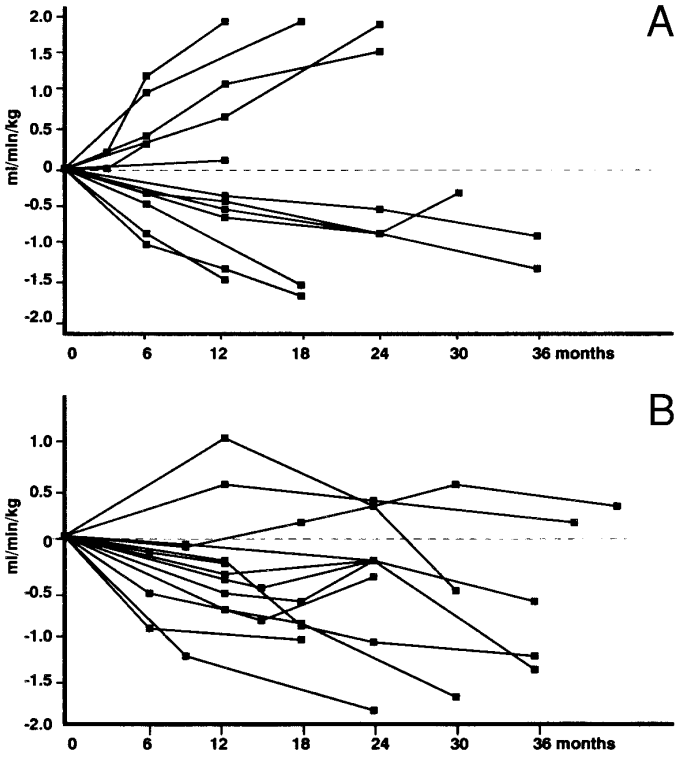

FIG. 3. Time course of differences of SCL between the initial value and the values at the time stated (AD) in abstinent (A) and nonabstinent patients (B) during the period of observation (individual polygone paths for all patients). Because of similar changes over time in a few patients, some points and lines represent more than one single patient.

TABLE 2. Routine Laboratory Tests and Quantitative Liver Function Tests at Entry Into the Study

	Group A (Abstinent Patients)	Group B (Nonabstinent Patients)
AST (<17 U/L)	28 (8-70)	32 (9-62)
ALT (<19 U/L)	27 (7-68)	25 (10-65)
GGT (<18 U/L)	68 (15-438)	85 (35-236)
Alkaline phosphatase (50-200 U/L)	212 (82-527)	178 (72-458)
Bilirubin (<1.2 mg/dL)	2.7 (0.5-27.2)	1.5 (0.6-6.3)
Prothrombin time (>70%)	68 (44-100)	71 (38-100)
Albumin (3.5-5.0 mg/dL)	3.9 (2.4-4.3)	3.7 (3.2-4.8)
γ-Globulins (11-20%)	28 (20-34)	28 (19-40)
Thrombocytes (150-300 $\times 10^3/\mu$L)	127 (57-400)	112 (52-234)
GEC (mg/min/kg)	3.0 (1.5-4.9)	4.8 (3.3-5.8)*
^{14}CABT (%dose·kg/mmol CO_2)	0.16 (0.01-0.49)	0.25 (0.04-0.76)†
SCI (mL/min/kg)	6.0 (3.1-9.9)	6.2 (3.9-8.8)

NOTE. Data are median (range).

Abbreviations: AST, aspartate transaminase; ALT, alanine transaminase; GGT, γ-glutamyl transferase.

Significant difference between the two groups: *$P < .005$, †$P < .01$.

bounds of the CIs of the residual SD served as the common SD of the two samples. This is a conservative procedure, i.e., oversized predictors of sample sizes are obtained.

For assessing the duration of a controlled clinical trial, the 95% CIs for the mean values in each group at each time point (obtained from ANOVA) were used. In group 2, showing on average higher values of the three quantitative liver function tests, the lower bounds of these intervals were taken as conservative substitutes for the means. In addition, the upper confidence bounds were taken in group 1. Proposals on trial duration were based on the difference of these estimates. The elapsed time from the beginning of the observation period to the point at which such deviations became significant gave us an estimate for the duration.

RESULTS

At the initial observation, the cohort of 15 patients who stopped drinking and the cohort of 17 patients who continued to drink were similar in regard to age, Pugh score, signs of portal hypertension, drug intake, number of return visits, follow-up period (Table 1), and routine laboratory tests (Table 2). At the end of the period of observation, no relevant differences were found between the two groups in regard to the routine laboratory tests (data not shown) and the Pugh score (median, 6.8; range, 5-14, in the nonabstinent group; median, 6.5; range, 5-12, in the abstinent group, respectively).

At study entry, values of the quantitative liver function tests were not comparable between the two patient groups (Table 2). Therefore, the previously defined outcome measures AD and RD were evaluated. Figures 1, 2, and 3 show the individual AD measures detected for each of the liver function tests over time for each patient.

Applying the ANOVA model, there was a general group-effect for each quantitative liver function test (i.e., different behavior regardless of time for the abstinent and nonabstinent group; $P < .001$ in ABT and GEC; $P < .05$ in SCl) for the AD and RD. A general time-effect could only be seen in the ABT ($P < .001$ for AD and RD) compared with the GEC and the SCl (not significant for AD and RD). Especially for the ABT and the GEC, significant interactions between group and time (ABT: $P < .001$, GEC: $P < .05$, for AD and RD, respectively) reflect the deterioration of liver function test results in the nonabstinent group and their amelioration in abstinent patients (Table 3).

The residual standard deviations and their 95% CIs after elimination of the influence of "group" and "time" effects on the outcome measures are listed in Table 4.

The reproducibility of the data can be assessed by the ratio of variance within patients over the whole residual variance. Good reproducibility is closely linked to a small ratio of these variances. Using AD as the outcome measure, values of 25% for GEC, 10% for ABT, and 12% for SCL were found, which are acceptable.

DISCUSSION

This retrospective investigation provides a realistic model for planning controlled clinical trials in hepatology, because

TABLE 3. Quantitative Liver Function Tests

		AD			RD		
Group	Time (mo)	GEC (mg/min/kg)	ABT (%dose·kg/mmol CO_2)	SCL (mL/min/kg)	GEC (mg/min/kg)	ABT (%dose·kg/mmol CO_2)	SCI (mL/min/kg)
A	12	0.64 (0.42; 0.86)	0.14 (0.09; 0.18)	0.02 (−0.42; 0.46)	0.25 (0.17; 0.33)	1.88 (0.64; 3.12)	0.04 (−0.05; 0.13)
B	12	−0.53 (−0.74; −0.32)	−0.09 (−0.13; −0.06)	−0.40 (−0.81; 0.01)	−0.12 (−0.19; −0.04)	−0.30 (−1.47; 0.86)	−0.06 (−0.15; 0.02)
A	24	0.97 (0.78; 1.21)	0.23 (0.19; 0.27)	0.10 (−0.36; 0.55)	0.37 (0.29; 0.45)	2.62 (1.37; 3.87)	0.06 (−0.03; 0.15)
B	24	−1.01 (−1.22; −0.80)	−0.14 (−0.17; −0.10)	−0.62 (−1.03; −0.20)	−0.22 (−0.29; −0.15)	−0.48 (−1.64; 0.68)	−0.10 (−0.19; −0.02)
A	36	1.29 (1.02; 1.55)	0.30 (0.25; 0.34)	0.06 (−0.42; 0.55)	0.49 (0.41; 0.58)	3.45 (2.18; 4.72)	0.07 (−0.03; 0.17)
B	36	−1.44 (−1.68; −1.20)	−0.16 (−0.20; −0.12)	−0.93 (−1.38; −0.49)	−0.30 (−0.38; −0.23)	−0.59 (−1.77; 0.60)	−0.16 (−0.25; −0.07)

NOTE. Data are means of AD and RD and their 95% CIs from ANOVA. In group A, follow-up was 12 months (n = 2), 18 months (n = 1), 21 months (n = 1), 24 months (n = 3), 30 months (n = 1), and 36 months (n = 7). In group B, follow-up was 24 months (n = 3), 30 months (n = 4), 32 months (n = 1), 36 months (n = 7), 39 months (n = 1), and 42 months (n = 1).

Abbreviations: A, abstinent group; B, nonabstinent group, AD, absolute difference between the test results at the indicated times and time zero ($t_x - t_0$); RD, AD of test results at indicated times and time zero related to the results at time zero $[(t_x - t_0]/t_0)$.

TABLE 4. Residual SDs From the ANOVA Model and the 95% CIs

Test	Outcome Measure	SD	95% CI
ABT	AD (%dose · kg/mmol CO_2)	0.08	(0.06; 0.10)
ABT	RD	2.41	(1.67; 2.96)
GEC	AD (mg/min/kg)	0.43	(0.32; 0.52)
GEC	RD	0.15	(0.11; 0.19)
SCI	AD (mL/min/kg)	0.85	(0.61; 1.04)
SCI	RD	0.17	(0.12; 0.21)

it involves actual patients with liver disease. It shows that, in alcoholic cirrhosis, GEC and ABT may be useful surrogate markers of treatment effects, and that trials based on these tests should be much more efficient and economical than those based on life-table methods with death as an end-point. To validate these conclusions, the limitations of this study must be discussed in detail.

The study subjects are representative of outpatients with moderately advanced alcoholic cirrhosis, exhibiting typical complications (Table 1). As one might expect, the 15 abstinent patients showed significant improvements in GEC and ABT. In contrast, those patients who continued to drink exhibited minor deteriorations. Obviously, complete abstinence cannot be assured in the cohort of patients who apparently stopped drinking. The extent of alcohol intake is unknown in the cohort of patients who continued to drink, but may have been no more than moderate because no patient developed alcoholic hepatitis or other complications of excessive alcohol intake.[38-40] Despite the limitations in the clinical assessment of drinking behavior, the two cohorts may be taken as representative of either successful medical treatment or continued chronic liver damage. In a true treatment trial, the differences between strictly separated placebo and effective drug groups could be expected to be even greater and the variability within groups smaller, even though estimating compliance with a drug regimen entails some of the same uncertainty as estimating alcohol intake.

Of note, the differences between the two cohorts were best reflected by GEC and ABT. Given the differences in GEC and ABT between the abstinent and the nonabstinent group at study entry, with the drinking group having the better values (Table 2), one might argue that the two groups might show similar liver function at the end of the observation period. This is an intrinsic problem of this type of study, in which allocation to different groups can only be performed retrospectively, rather than by initial randomization. Patients with more advanced liver disease were more likely to abstain from drinking. However, including these differences in time-zero GEC and ABT in the ANOVA as an additional covariable has only a slight effect on the AD and RD values (Table 3). The interactions between group and time still remain statistically significant for the ABT ($P = .012$), but fall short of statistical significance in the case of GEC ($P = .24$). Thus, the conclusion derived from this model study that "effective treatment" of chronic liver disease can be detected through measurable improvements in GEC and ABT remains undisputed. In a randomized drug trial, similar differences at study entry should not exist.

Improvement in ABT upon cessation of drinking and deterioration of GEC and ABT in chronic cholestatic liver disease are well documented.[26,46-48] In contrast, neither the time course of various routine laboratory tests nor the Child-Pugh score parallel alcohol use in the investigated patients. SCL, which is thought to reflect parenchymal liver plasma flow, also does not distinguish the two cohorts (Table 3). Such observations and our findings support the use of GEC and ABT as surrogate markers for end-points in controlled clinical trials.

Our data were analyzed with regard to the numbers of patients and the duration of time required to detect a significant difference between two groups. One may calculate that mean differences in GEC and ABT must be $\geq$0.6 mg/min/kg and $\geq$0.11 %dose · kg/mmol CO_2, respectively, to reach statistical significance at the 5% level in a study with two groups of 20 patients each (Table 5). Mean differences for smaller and larger patient groups are also given in Table 5. Because residual variances observed in an age- and sex-matched group of patients with stable disease (data not shown) were appreciably smaller than those in the two study cohorts, in such a situation, the numbers in Table 5 are likely to be overestimated, i.e., the data are on the conservative side.

Assuming that a treated and an untreated group of patients would follow an evolution similar to the two cohorts shown in Figs. 1 and 2, one may also calculate periods of observation necessary to detect the significant differences presented in Table 5. In a study with two groups of 20 patients each, treatment-related differences may be detected at the 5% level of significance after 11 months using the GEC and after 9 months with the ABT. Periods of observation for smaller or

TABLE 5. Least Treatment Effects (absolute differences of means) in GEC, ABT, and SCI Between Two Groups of Patients Detectable With a Power of 90% for Levels of Significance $\alpha = 0.05$ or $\alpha = 0.01$

No. of Patients Per Group	GEC (mg/min/kg)		ABT (%dose · kg/mmol CO_2)		SCI (mL/min/kg)	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
10	0.8	0.9	0.15	0.18	1.6	1.9
15	0.7	0.8	0.12	0.15	1.3	1.5
20	0.6	0.7	0.11	0.13	1.1	1.3
25	0.5	0.6	0.10	0.11	1.0	1.2
30	0.5	0.6	0.09	0.10	0.9	1.1

NOTE. Data based on a two-sided test. Based on the differences of the two cohorts (A and B) for quantitative liver function tests and according to the "group-effect" and "time-effect" applying the ANOVA model, mean differences in the corresponding tests were calculated.

TABLE 6. Periods of Observation Necessary for Patients Per Group in a Controlled Clinical Trial to Detect the Treatment Effects on GEC, ABT, and SCI Given in Table 5 With a Power of 90% for $\alpha = 0.05$ or $\alpha = 0.01$

No. of Patients Per Group	GEC (mo)		ABT (mo)		SCI (mo)	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
10	14	16	12	15	NE	NE
15	12	14	10	12	NE	NE
20	11	12	9	11	NE	NE
25	9	11	8	9	NE	NE
30	9	11	8	8	NE	NE

NOTE. Data based on a two-sided test. Based on the differences on the two cohorts (A and B) for quantitative liver function tests and according to the "group-effect" and "time-effect" applying the ANOVA model, mean periods of observation for the corresponding tests were calculated to determine significant differences in these tests. Treatment effects were only available at 12, 24, and 36 months. Other values were obtained by linear interpolation.
Abbreviation: NE, not evaluated.

larger patient groups are summarized in Table 6. Because the data of an age- and sex-matched group of patients (data not shown) with stable disease exhibit residual variances smaller than those used to construct Tables 5 and 6, the periods of observation necessary to detect the above differences may be shorter in reality.

Based on these extrapolations, the following design would be appropriate for a controlled clinical trial aiming at an improvement of the diseased liver itself: two groups of 15 to 30 patients each must be observed for approximately 1 year to detect, at the 5% level of significance, a difference that corresponds to the difference between the cohorts continuing or stopping alcohol consumption in this study. For instance, if 20 patients are included per group and GEC and ABT are used as outcome measures, only 37 or 30 patient-years (α = 5%) and 40 or 37 patient-years (α = 1%) would be required, respectively (Table 7). This contrasts favorably with controlled trials in hepatology including survival analysis and published in major journals between 1986 and 1994. These trials use median observation times of 444 patient-years (range, 48-2,094 patient-years) (Table 7).[6-20] The impressive economy resulting from the use of GEC and ABT may be due to several circumstances. First, the severity of the disease at entry into the trial can be characterized exactly using quantitative liver function tests,[2,21,25-27,32,47-53] and the outcome can then be measured as the difference in the GEC or ABT. In contrast, the severity of disease at onset is poorly defined in conventional trials using death as the only endpoint.[6-20] Second, survival analysis may be appropriate for end-stage disease but will fail to detect effects of treatment early in the course of the disease. In contrast, the use of quantitative liver function tests permits the investigation of

TABLE 7. Patient-Years in Controlled Clinical Trials Published Between 1986 and 1994 Compared With Tables 5 and 6

Disease	Treatment	Total Number of Cases	Duration of Observation (mo)	Patient-Years	Reference
Survival analysis					
PSC	Methotrexate	24	24	48	18
AC	Nutrition	51	12	51	16
ALD	Silymarin	72	15	90	13
PBC	Colchicine	64	18	96	8
PBC	Prednisolone	36	36	108	12
PBC	Malotilate	101	28	236	14
PSC	Penicillamine	70	47	274	10
PBC	UDCA	222	24	444	19
CI	Colchicine	100	56	467	9
AC	Testosterone	221	28	516	6
PBC	UDCA	145	48	580	17
CI	Silymarin	170	41	581	11
ALD	Malotilate	174	42	609	20
ALD	Propylthiouracil	310	24	620	7
PBC	Ciclosporin A	349	72	2094	15
QLFT analysis*					
AC	GEC ($\alpha = 0.05$)	40	11	37	
	($\alpha = 0.01$)	40	12	40	
AC	ABT ($\alpha = 0.05$)	40	9	30	
	($\alpha = 0.01$)	40	11	37	

Abbreviations: AC, alcoholic cirrhosis; ALD, alcoholic liver disease; CI, cirrhosis; PBC, primary biliary cirrhosis; PSC, primary sclerosing cholangitis; QLFT, quantitative liver function tests; UDCA, ursodeoxycholic acid.
* Results taken from Table 6.

therapeutic interventions at any stage of the disease. Finally, quantitative liver function tests used as surrogate markers in controlled clinical trials measure treatment effects on liver function itself.[22-27] This is a more direct way of determining treatment effects on the liver than the indirect estimation of liver function based on changes in clinical symptoms, routine laboratory tests, and histological features as described in some clinical trials in patients with chronic cholestatic liver disease.[17,54]

Strictly speaking, the time course of changes in GEC or ABT described in this study applies only to patients with alcoholic cirrhosis who are similar to the investigated cohorts and who stopped or continued to drink. Should controlled clinical trials be performed for diseases with a different natural history, the calculated periods of observation may vary. Primary biliary cirrhosis progresses more slowly than alcoholic liver disease,[17,24,25,54] and autoimmune or viral chronic active hepatitis progresses more rapidly.[1,21,27] However, the least differences in treatment effects needed for statistical significance will not be affected (Table 5), if the variability within and between patients is comparable with ours (Table 4). Whether or not the cost or precision of clinical trials may be further improved by the use of other quantitative liver function tests such as antipyrine,[55] caffeine,[56] or indocyanine green[57] clearance, or monoethylglycinexylidide formation,[58] remains to be shown.

In conclusion, our data suggest that quantitative liver function tests may reduce sample sizes and observation times in controlled clinical trials in hepatology. Using this methodology, trials could be performed in less prevalent diseases or in more homogenous patient groups. The number of dropouts will be reduced, the risk of noncompletion of trials would be small, and, thus, such trials would be more economical than large studies depending on long-term observation of patients. Finally, earlier stages of liver disease may be investigated. This is crucial to the development of treatment strategies for early liver disease.

Acknowledgment: The authors gratefully acknowledge the sympathetic care for the patients and the excellent technical assistance of Mrs. Christine Hofmann.

REFERENCES

1. Tygstrup N, Juhl E. Dilemmas of controlled clinical trials in hepatology. In: Schaffner F, Sherlock S, Leevy CM, eds. The Liver and Its Diseases. Stuttgart, Thieme, 1974.
2. Villeneuve JP, Infante-Rivard C, Ampelas M, Pomier-Layrargues G, Huet PM, Marleau D. Prognostic value of the aminopyrine breath test in cirrhotic patients. HEPATOLOGY 1986;6:928-931.
3. Infante-Rivard C, Esnaola S, Villneuve JP. Clinical and statistical validity of conventional prognostic factors in predicting short-term survival among cirrhotics. HEPATOLOGY 1987;7:660-664.
4. Hay JE, Czaja AL, Rakela J, Ludwig J. The nature of unexplained chronic aminotransferase elevations of a mild to moderate degree in asymptomatic patients. HEPATOLOGY 1989;9:193-197.
5. Saunders JB, Walters JRF, Davies P, Paton A. A 20-year prospective study of cirrhosis. Br Med J 1981;282:263-266.
6. The Copenhagen study group for liver diseases. Testosterone treatment of men with alcoholic cirrhosis: a double-blind study. HEPATOLOGY 1986;6:807-813.
7. Orrego H, Blake JE, Blendis LM, Compton KV, Israel Y. Long-term treatment of alcoholic liver disease with propylthiouracil. N Engl J Med 1987;317:1421-1427.
8. Warnes TW, Smith A, Lee FI, Haboubi NY, Johnson PJ, Hunt L. A controlled trial of colchicine in primary biliary cirrhosis. Trial design and preliminary report. J Hepatol 1987;5:1-7.
9. Kershenobich D, Vargas F, Garcia-Tsao G, Tamayo RP, Gent M, Rojkind M. Colchicine in the treatment of cirrhosis of the liver. N Engl J Med 1988;318:1709-1713.
10. LaRusso NF, Wiesner RH, Ludwig J, MacCarty RL, Beaver SJ, Zinsmeister AR. Prospective trial of penicillamine in primary sclerosing cholangitis. Gastroenterology 1988;95:1036-1042.
11. Ferenci P, Dragosics B, Dittrich H, Frank H, Benda L, Lochs H, Meryn S, et al. Randomized controlled trial of silymarin treatment in patients with cirrhosis of the liver. J Hepatol 1989;9:105-113.
12. Mitchison HC, Palmer JM, Bassendine MF, Watson AJ, Record CO, James OFW. A controlled trial of prednisolone treatment in primary biliary cirrhosis. Three-year results. J Hepatol 1992;15:336-344.
13. Bunout D, Hirsch S, Petermann M, de la Maza MP, Silva G, Kelly M, Ugarte G, et al. Controlled study of the effect of silymarin on alcoholic liver disease. Rev Med Chil 1992;120:1370-1375.
14. A European multicentre study group. The results of a randomized double blind controlled trial evaluating malotilate in primary biliary cirrhosis. J Hepatol 1993;17:227-235.
15. Lombard M, Portmann B, Neuberger J, Williams R, Tygstrup N, Ranek L, Ring-Larsen H, et al. Cyclosporin: a treatment in primary biliary cirrhosis: results of a long-term placebo controlled trial. Gastroenterology 1993;104:519-526.
16. Hirsch S, Bunout D, de la Maza P, Iturriaga H, Petermann M, Icazar G, Gattas V, et al. Controlled trial on nutrition supplementation in outpatients with symptomatic alcoholic cirrhosis. J Parenter Enteral Nutr 1993;17:119-124.
17. Poupon RE, Poupon R, Balkau B, and the UDCA-PBC Study Group. Ursodiol for the long-term treatment of primary biliary cirrhosis. N Engl J Med 1994;330:1342-1347.
18. Knox TA, Kaplan MM. A double-blind controlled trial of oral-pulse methotrexate therapy in the treatment of primary sclerosing cholangitis. Gastroenterology 1994;106:494-499.
19. Heathcote EJ, Cauch-Dudek K, Walker V, Bailey RJ, Blendis LM, Ghent CN, Michieletti P, et al. The Canadian multicenter double-blind randomized controlled trial of ursodeoxycholic acid in primary biliary cirrhosis. HEPATOLOGY 1994;19:1149-1156.
20. Keiding S, Badsberg JH, Becker U, Bentsen KD, Bonnevie O, Caballeria J, Eriksen J, et al. The prognosis of patients with alcoholic disease. An international randomized, placebo-controlled trial on the effect of malotilate on survival. J Hepatol 1994;20:454-460.
21. Bircher J. Quantitative assessment of deranged hepatic function: a missed opportunity? Semin Liver Dis 1983;3:275-284.
22. Meister F, Reichen J. Quantitative Leberfunktionstests bei Porphyria cutanea tarda: Verbesserung unter Therapie mit Phlebotomie. Schweiz Med Wschr 1986;116:341-345.
23. Rikkers LF, Burnett DA, Volentine GD, Buchi KN, Cormier RA. Shunt surgery versus endoscopic sclerotherapy for long-term treatment of variceal bleeding. Ann Surg 1987;206:261-271.
24. Warnes TW, Rowan BP, Smith A, Babbs C. Colchicine improves hepatic excretory failure in primary biliary cirrhosis. HEPATOLOGY 1988;8:1221.
25. Lotterer E, Stiehl A, Raedsch R, Foelsch UR, Bircher J. Ursodeoxycholic acid in primary biliary cirrhosis: no evidence for toxicity in the stages I to III. J Hepatol 1990;10:284-290.
26. Cotting J, Lentze MJ, Reichen J. Effects of ursodeoxycholic acid treatment on nutrition and liver function in patients with cystic fibrosis and longstanding cholestasis. Gut 1990;31:918-921.
27. Reichen J, Bianchi L, Frei PC, Malé P-J, Lavanchy D, Schmid M, and the Swiss Association for the Study of the Liver. Efficacy of steroid withdrawal and low-dose interferon treatment in chronic active hepatitis B. J Hepatol 1994;20:168-174.
28. Gluud C, Henriksen JH, and the Copenhagen study group for liver diseases. Liver haemodynamics and function in alcoholic cirrhosis—relation to testosterone treatment and ethanol consumption. J Hepatol 1987;4:168-173.
29. Achord JL. A prospective randomized clinical trial of peripheral amino acid–glucose supplementation in acute alcoholic hepatitis. Am J Gastroenterol 1987;82:871-875.
30. Simon D, Galambos JT. A randomized controlled study of peripheral parenteral nutrition in moderate and severe alcoholic hepatitis. J Hepatol 1988;7:200-207.
31. Vinel JP, Caucanas JP, Calès P, Suduca JM, Voigt JJ, Pascal JP. Effect of propranolol on metabolic activity of the liver in patients with alcoholic cirrhosis. J Hepatol 1988;7:186-192.
32. Reichen J, Solioz M, Bühler H, Gonvers JJ, Knoblauch M, Lavanchy D, Malé P-J, et al. Low-dose interferon in chronic hepatitis non-A/non-B:

effects on quantitative liver function and structure in a randomized, controlled multicenter trial. Clin Invest 1993;71:888-893.
33. Herson J. The use of surrogate endpoints in clinical trials (an introduction to a series of four papers). Stat Med 1989;8:403-404.
34. Prentice RL. Surrogate endpoints in clinical trials: definitions and operational criteria. Stat Med 1989;8:431-440.
35. Merkel C, Bolognesi M, Bellon S, Bianco S, Honisch B, Lampe H, Angeli P, et al. Aminopyrine breath test in the prognostic evaluation of patients with cirrhosis. Gut 1992;33:836-842.
36. Merkel C, Gatta A, Zoli M, Bolognesi M, Angeli P, Iervese T, Marchesini G, et al. Prognostic value of galactose elimination capacity, aminopyrine breath test, and ICG clearance in patients with cirrhosis. Comparison with the Pugh score. Dig Dis Sci 1991;36:1197-1203.
37. Reichen J, Widmer T, Cotting J. Accurate prediction of death by serial determinations of galactose elimination capacity in primary biliary cirrhosis: a comparison with the Mayo model. HEPATOLOGY 1991;14:504-510.
38. Parés A, Caballeria J, Bruguera M, Torres M, Rodés J. Histological course of alcoholic hepatitis—influence of abstinence, sex and extent of hepatic damage. J Hepatol 1986;2:33-42.
39. Marbet UA, Bianchi L, Meury U, Stalder GA. Long-term histological evaluation of the natural history and prognostic factors of alcoholic liver disease. J Hepatol 1987;4:364-372.
40. Lieber CS. Alcohol and the liver: 1994 update. Gastroenterology 1994; 106:1085-1105.
41. Bloor JH, Mapoles JE, Simon FR. Alcoholic liver disease: new concepts of pathogenesis and treatment. Adv Intern Med 1994;39:49-92.
42. Scheuer PJ. Liver Biopsy Interpretation. 4th ed. London: Baillière Tindall, 1988.
43. Tygstrup N. Determination of the hepatic elimination capacity (Lm) of galactose by single injection. Scand J Clin Lab Invest 1966;18(Suppl): 118-125.
44. Miotti T, Bircher J, Preisig R. The 30-minute aminopyrine breath test: optimization of sampling times after intravenous administration of ^{14}C-aminopyrine. Digestion 1988;39:241-250.
45. Zeeh J, Lange H, Bosch J, Pohl S, Loesgen H, Eggers R, Navasa M, et al. Steady-state extrarenal sorbitol clearance as a measure of hepatic plasma flow. Gastroenterology 1988;95:749-759.
46. Saunders JB, Lewis KO, Paton A. Early diagnosis of alcoholic cirrhosis by the aminopyrine breath test. Gastroenterology 1980;79:112-114.
47. Gross JB, Reichen J, Zeltner T, Zimmermann A. The evolution of changes in quantitative liver function tests in a rat model of cirrhosis. HEPATOLOGY 1987;7:457-463.
48. Reichen J, Widmer T, Cotting J. Accurate prediction of death by serial determination of galactose elimination capacity in primary biliary cirrhosis: a comparison with the Mayo model. HEPATOLOGY 1991;14:504-510.
49. Bircher J, Blankart R, Halpern A, Häcki W, Laissue J, Preisig R. Criteria for assessment of functional impairment in patients with cirrhosis of the liver. Eur J Clin Invest 1973;3:72-85.
50. Schneider JF, Baker AL, Haines NW, Hatfield G, Boyer JL. Aminopyrine N-demethylation: a prognostic test of liver function in patients with alcoholic liver disease. Gastroenterology 1980;79:1145-1150.
51. Monroe PS, Baker AL, Schneider JF, Krager PS, Klein PD, Schoeller D. The aminopyrine breath test and serum bile acids reflect histologic severity in chronic hepatitis. HEPATOLOGY 1982;2:317-322.
52. Orrego H, Israel Y, Blake JE, Medline A. Assessment of prognostic factors in alcoholic liver disease: toward a global quantitative expression of severity. HEPATOLOGY 1983;3:896-905.
53. Oakland CDH, Hickman R, Terblanche J. The aminopyrine breath test predicts the outcome of hepatic transplantation in pigs. HEPATOLOGY 1989;9:602-605.
54. Wiesner RH, Ludwig J, Lindor KD, Jorgensen RA, Baldus WP, Homburger HA, Dickson ER. A controlled trial of cyclosporine in the treatment of primary biliary cirrhosis. N Engl J Med 1990;322:1419-1424.
55. Andreasen PB, Ranek L, Statland BE, Tygstrup N. Clearance of antipyrine-dependence of quantitative liver function. Eur J Clin Invest 1974; 4:129-134.
56. Renner E, Wietholtz H, Huguenin D, Arnaud M, Preisig R. Caffeine: a model compound for measuring liver function. HEPATOLOGY 1984;4: 38-46.
57. Merkel C, Bolognesi M, Finucci GF, Angeli P, Caregaro L, Rondana M, Gatta A. Indocyanine green intrinsic hepatic clearance as a prognostic index of survival in patients with cirrhosis. J Hepatol 1989;9: 16-22.
58. Oellerich M, Raude E, Burdelski M, Schulz M, Schmidt FW, Ringe B, Lamesch P, et al. Monoethylglycine-xylidide formation kinetics: a novel approach to assessment of liver function. J Clin Chem Clin Biochem 1987;25:845-853.