B. Additional Statistical Analyses

The results of the Mayo Clinic Trial, for which data were trunkated at 2 years for patients maintained on double-blind for more than 2 years, were presented to the FDA during an end-of-phase II meeting held on April 18, 1995. [These are the data the Medical Officer has evaluated in detail, above.] At that meeting, the FDA recommended to extend the follow-up of patients participating in the long-term open label, extension phase of the Mayo trial [to report life-table analysis of the long-term data][37], to revise the definition of treatment failure by excluding reasons of doubling of bilirubin and voluntary withdrawal and to carry out additional analyses relating to the development of varices. Results of these statistical evaluations, carried out by Dr. J. Richard Trout (document dated December 20, 1995) were presented by the sponsor in vol. 39, pages 001 through 223). In addition to the FDA recommendations, submitted in Dr. Trout's report, an evaluation was performed including the Mayo Risk score as a possible covariate in the statistical modeling.

In May 1992, all patients consented to be switched to open-label treatment with UDCA. For the purposes of the analyses presented here, patients were left in the groups (UDCA or PL) to which they were originally assigned. [This is important as **this conservative approach may have biased results against UDCA**].

1. Patient Status in Extended Follow-up Period (Table 31)

NOTE: The data presented here include the grand total, that is patients that withdrew during the double-blind period of the trial plus those that withdrew during the follow-up extension period. As shown in Table 31, a total of 69 patients withdrew (UDCA, n=23; PL, n=46) for reasons specified in this Table. From this Table, the total number of deaths is 14 (UDCA=4; PL=10), but the correct number should be 16 (UDCA=6; PL=10). This is because two patients, #706 and #761, were not included in the initial analysis as treatment failures. Each of these patients died over one year following discontinuation of treatment (patient #706, 435 days; patient #761, 555 days). Therefore, in the statistical calculations, the total number of patients that died or were transplanted in the UDCA group is 13 (not 11 as shown in Table 31).

Table 31 also shows that in comparison to patients initially assigned to the PL group, UDCA-treated patients that withdrew stayed in the trial an average of 49 more days (903-854) than those treated with PL. If one considers the total number of patients initially randomized (UDCA=89; PL=91), UDCA-treated patients stayed in the trial an average of 253 days longer than those assigned to PL. This difference between the treatment groups is illustrated in Fig. 3.

[37] The reader is reminded that the Mayo Clinic trial was initiated in April 1988 and was designed so that the blind would be broken when the 132nd patient had completed 2 years of double-blind treatment. This occurred in May 1992 when all patients were offered and accepted participation in the L-T, open-label, UDCA treatment extension.

TABLE 31
Mayo Clinic Trial

Summary of Patient Status [Double-blind plus L-T Extension Periods]

Reason for Withdrawal	UDCA	PL
Death	4[a]	10
Transplant	7	11
Voluntary Withdraw	11	18
Other	1	7
Total	[n=23]	[n=46]
Time on Study (days) (Patients Who Withdrew)	[n=23]	[n=46]
Mean	903 (±444)[b]	854 (±497)
Range	144 to 1498	35 to 2133
Time on Study (days) (All Patients)	[n=89]	[n=91]
Mean	1543 (±552)	1290 (±653)
Range	144 to 2518	35 to 2511

a) To these 4, two additional patients (#706 and #761) each of which died over one year following D/C of treatment, should be added.
b) ± SD of the Mean.

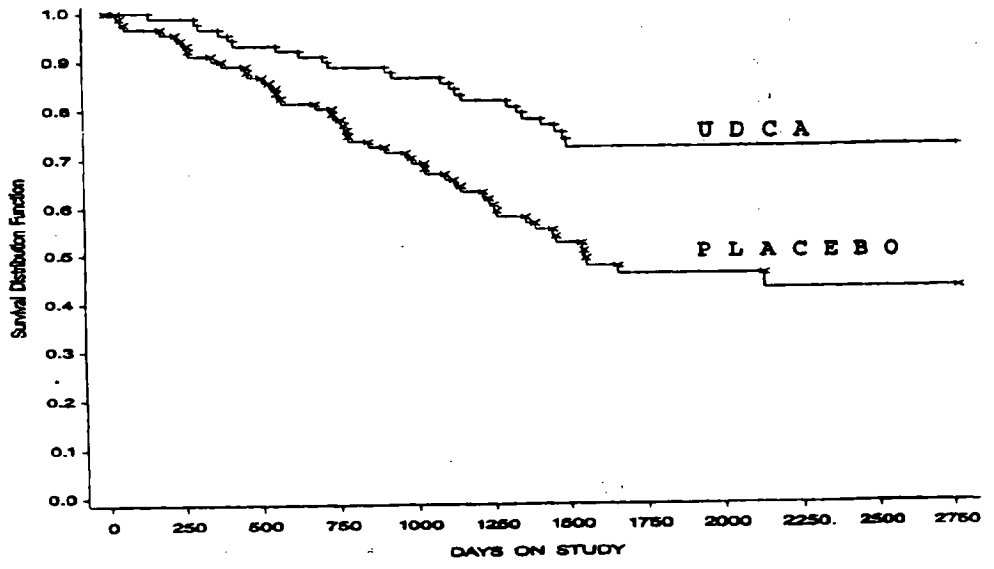

Fig. 3. - Mayo Clinic Trial Extended Follow-up Period: Time to Patient Withdrawal

2. Analysis of Death/Transplant (Table 32, Fig. 4)

In this Table, the results using statistical analysis specified in the original report (Log-rank and Wilcoxon tests) but applied to the L-T data, are summarized. For comparison purposes, also included in this Table are the results of analyzing death/transplant at the 2-year cutoff (double-blind portion of the trial). For the proportion of patients who died or were transplanted, the therapeutic gain was 8% and this is approaching statistical significance.

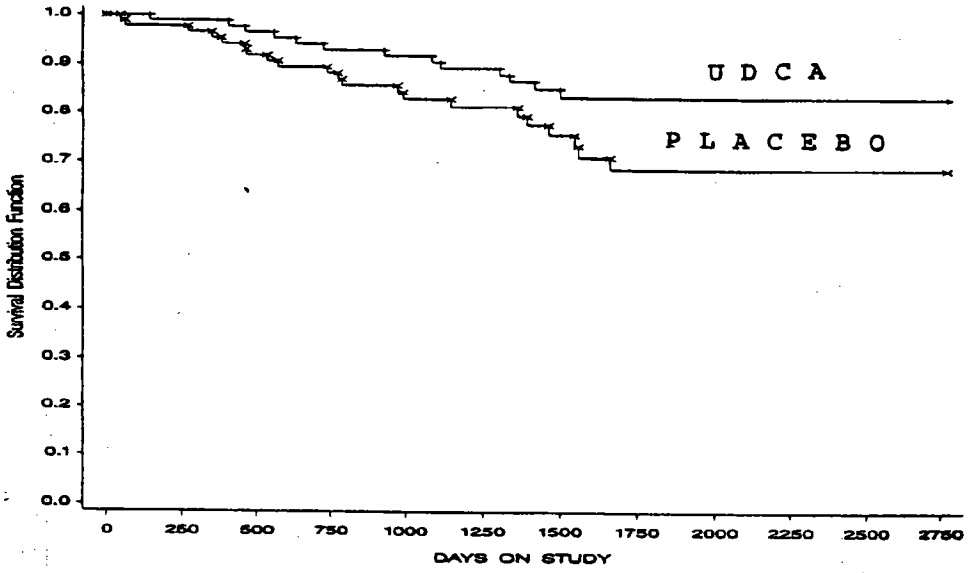

Fig. 4. - *Mayo Clinic Trial Extended Follow-up: Time to Death/Transplant*

3. Life Table Analysis Using Baseline Mayo Risk Score as a Covariate (Table 33)

- As previously shown (Table 10), after randomization, and based on a t-test, the treatment groups were well matched with regard to the mean Mayo Risk Score [UDCA (n=87)=5.1 (±1.1 SD); PL (n=90)=5.0 (±1.1); p-value=N.S. (0.60)].

- The sponsor used a Cox proportional hazards model to compare the treatment groups after adjusting for the Mayo Risk score. Summarized in Table 33 are results of analysis of both the double-blind portion (2 year cutoff) and the extended follow-up.

TABLE 32
Mayo Clinic Trial
Extended Follow-up

Time to Death/Transplant

	DOUBLE-BLIND PORTION (Two-year Cutoff)				EXTENDED FOLLOW-UP			
	UDCA [n=86]	PL [n=86]	Therapeutic Gain	p-value	UDCA [n=89]	PL [n=91]	Therapeutic Gain	p-value
Number (%) of Pts. who died or were transplanted	6 (7%)	11 (13%)	6%	0.11[a] 0.15[b]	13 (15%)	21 (23%)	8%	0.0597[a] 0.0778[b]
Mean Days to Death/ Transplant	707 (±10.1)	732 (±18.5)	NONE		1405 (±30.3)	1445 (±48.5)	NONE	

a) Log-rank Test
b) Wilcoxon Test

- For both analyses the treatment Mayo Risk score term in the model was found to be not statistically significant ($p>0.25$).

- After the interaction term was removed from the model the p-values for the comparison between the treatment groups were 0.07 for the 2-year data set and 0.007 for the extended follow-up.

- For all analyses the results favored the UDCA treatment group. The difference in Mayo Index was highly significant ($p=0.0001$ at both periods of observation). According to the sponsor the improvement in the sensitivity of the analysis is attributable to the reduction in the variability in the response due to the prognostic importance of the Mayo Risk score.

TABLE 33
Mayo Clinic Trial
Extended Follow-up

Time to Death/Transplant Using Cox Modeling[a]

	Double-blind Portion (Two-year Cutoff)	Extended Follow-up
	p-values[b]	
Treatment Group Comparison	0.07	0.007
Mayo Index	0.0001	0.0001
Treatment group*Mayo Index	N.S.[c]	N.S.[d]

a) The Mayo Risk score could not be determined for 3 patients (UDCA=2; PL=1). None of these 3 patients died or were transplanted. So, the corresponding n for UDCA and PL were 87 and 90.
b) All p-values are based on a Cox proportional hazards model, after adjusting for the Mayo Risk score.
c) p-value = 0.27
d) p-value = 0.28

4. Time to Treatment Failure Using a Revised Definition (Table 34, Fig. 5)

For both, the double-blind portion (2-year cutoff) and extended follow-up analysis, the exclusion of the doubling of BIL and voluntary withdrawal still lead to statistically significant differences between the two treatment groups. For both analyses the results were statistically significantly different. These analyses also showed a clinically important therapeutic gain (2-year cutoff = 10%; extended follow-up = 18%) in the proportion of patients that failed and in the mean days to failure (2-year cutoff period = 110 days and extended follow-up = 268; both favoring UDCA).

TABLE 34
Mayo Clinic Trial
Extended Follow-up

Time to Treatment Failure - Revised Definition

	DOUBLE-BLIND PORTION (Two-year Cutoff)				EXTENDED FOLLOW-UP			
	UDCA [n=86]	PL [n=86]	Therapeutic Gain	p-value	UDCA [n=89]	PL [n=91]	Therapeutic Gain	p-value
Number (proportion) of Pts. that Failed	17 (20%)	26 (30%)	10%	0.01[a] 0.03[b]	29 (33%)	46 (51%)	18%	0.001[a] 0.002[b]
Mean Days to Failure (S.E. Mean)	823 (±23)	713 (±20)	110		1497 (±55)	1229 (±64)	268	

a) Log Rank Test
b) Wilcoxon Test

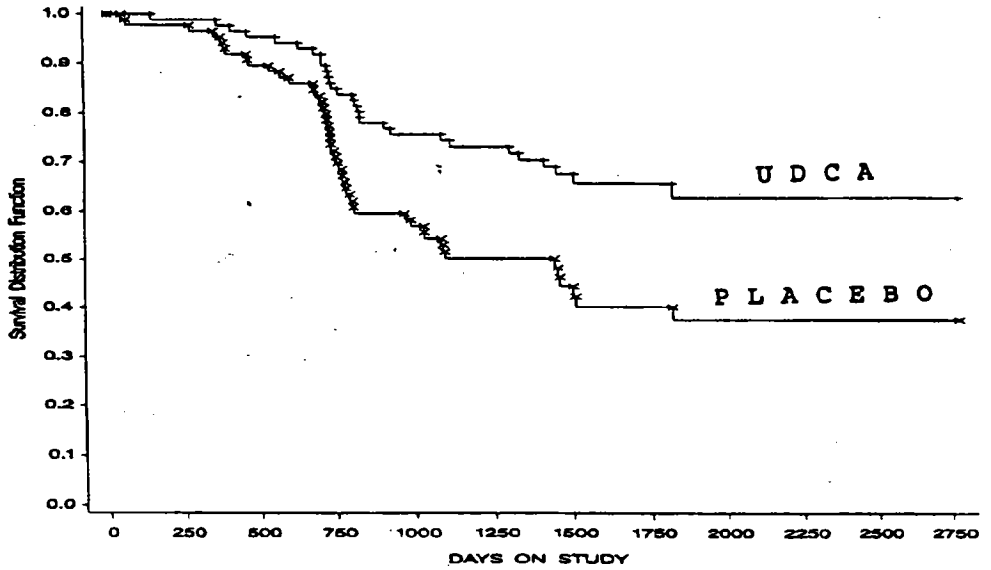

Fig. 5. - *Mayo Clinic Trial Extended Follow-up*
Time to Treatment Failure (Revised Definition)

5. Development of Varices (Table 35, Fig. 6)

- For the obvious reasons, these analyses included only those patients who presented at baseline without varices.
- The number of patients available for these analyses was, UDCA=70, PL=69.
- A summary of the life table analysis of the development of varices is given in Table 35 and Fig. 6.
- Analyses of both data at the 2-year cutoff period and extended follow-up period were carried out. The therapeutic gain at the 2-year cutoff period was 4% (p=0.04 by Log-rank Test) which increased considerable to 18% at the end of the extended follow-up period (p=0.003 by Log-rank Test and 0.004 by Wilcoxon Test). In both analyses, the results favor the patients initially assigned to UDCA treatment group. The superiority of UDCA over PL can be graphically seen in Fig. 6.

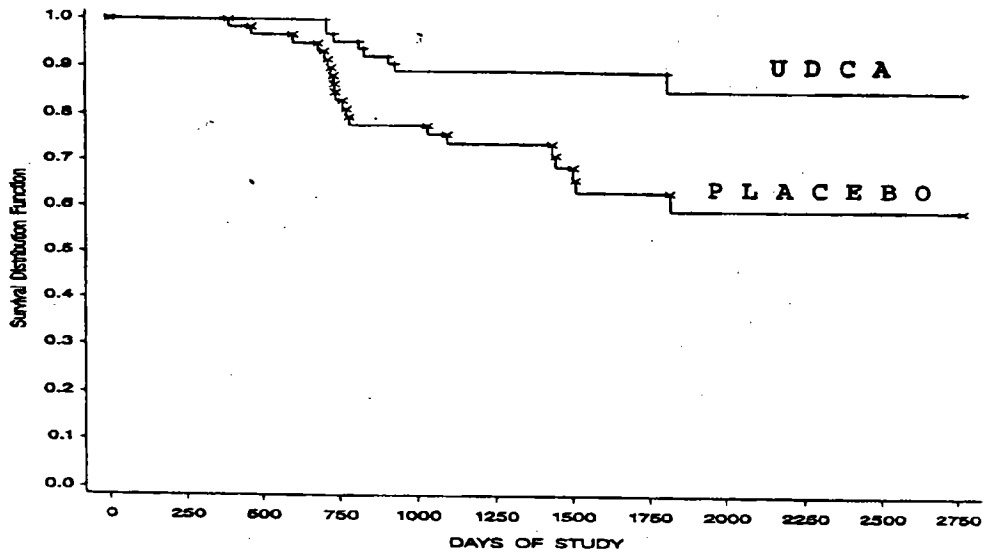

Fig. 6. - *Mayo Clinic Trial Extended Follow-up: Time to Development of Varices Among Patients Without Varices at Baseline*

APPEARS THIS WAY ON ORIGINAL

TABLE 35
Mayo Clinic Trial
Extended Follow-up

Proportion of Patients Developing Varices and Time to Development of Varices in Patients Without Varices at Baseline

	DOUBLE-BLIND PORTION (Two-year Cutoff)				EXTENDED FOLLOW-UP			
	UDCA [n=68]	PL [n=67]	Therapeutic Gain	p-value	UDCA [n=70]	PL [n=69]	Therapeutic Gain	p-value
Number (Proportion) of Pts. Developing Varices	6 (9%)	9 (13%)	4%	0.04 [a] 0.07 [b]	8 (11%)	20 (29%)	18%	0.003 [a] 0.004 [b]
Mean Days to Developing Varices (± S.E.)	885 (±22)	757 (±12)	128		1704 (±41)	1503 (±64)	201	

a) Log Rank Test
b) Wilcoxon Test

6. Sponsor's Conclusions from Additional Statistical Analyses

"The following conclusions can be drawn from the additional analyses performed on the long-term open label database of the Mayo PBC URSO study:

"1. Long-term treatment with URSO significantly improves survival or the need for transplant in patients with PBC. Overall, 15% of the URSO patients developed death/transplant compared to 23% for the placebo group.

"2. Long-term treatment with URSO significantly delays the time to treatment failure in patients with PBC. Overall, 33% of the URSO patients failed compared to 51% for the placebo group.

"3. Long-term treatment with URSO significantly reduces the risk of developing varices in patients with PBC. Overall, 11% of the URSO patients developed varices compared to 29% for the placebo group."

C. Reviewer's Comments on the Mayo Clinic Trial

The Mayo Clinic Trial, conducted under IND [redacted] is the adequate and well-controlled study the sponsor of NDA 20-675 has submitted in support of their request for approval of the marketing of UDCA (URSO™) film coated tablets for the treatment of patients with all stages of PBC. The sponsor presented very detailed data on the design, collection of clinically meaningful endpoints and other aspects of the execution of the trial and assessment of results. All of these submitted data and additional information requested during the course of this appraisal, were closely examined by the Medical Officer. It is clear that the main intention of this trial was to identify (rather ratify, since there are already several publications, suggesting that UDCA is efficacious in this patient population) a potentially useful therapy for PBC. One important aim of the MO's review was to determine if the data from the Mayo Clinic Trial could stand alone as a basis of approval.

The sponsor reported the results of a multicenter (primarily Mayo Clinic Rochester, MN), two-phase study on the effectiveness of UDCA, 15 mg/Kg, given q.i.d. in patients with PBC. The first phase was a stratified, randomized, double-blind, 2-arm, PL-controlled, 2-year trial. This was followed by a second, open-label up to 4 years phase. The observations at the end of the extended duration of the trial were essential to test the hypothesis that longer duration of therapy (4 years or more) may be required to obtain effects on important clinical endpoints. Included among the latter were delaying the time to orthotopic liver transplantation or death, significantly delaying the time to treatment failure and significantly decreasing the risk of developing esophageal varices. It was important to show that although already effective at 2 years, a 4+ year course is needed to markedly improve effectiveness. All of the above-mentioned clinically meaningful effects, needed to be demonstrated in addition to the expected significant effects on biochemical para-

meters of liver function such as bilirubin and AP, as well as IgM. The overall appraisal also explored subgroup hypothesis. The MO assessed the possibility that certain groups of PBC patients (i.e. those with "early" disease) may benefit the most. This made necessary a comparison of response in low vs high risk patients. Assessment of these subgroup hypothesis was made possible by the stratification of patients at entry on the basis of serum bilirubin level (≤1.8 vs >1.8 mg/dl), histological stage (early=I + II vs late=III + IV) and presence or absence of esophageal varices.

Certain requisites with regard to trial execution were met and this resulted in sound basis for conclusions. From the review of the evidence the MO wishes to mention that the objectives of the trial were clearly stated in the protocol and that patients for this study were selected in ways that provided adequate assurance that the patient population was suitable for the purposes of the trial. The explanations provided by the sponsor are reassuring that, during this trial's execution, steps were taken to minimize bias on part of the Principal Investigator, other observers of patients, examiners of histologic data and researchers carrying out the analyses of data.

The MO described in detail all aspects of the protocol, including specific aims (plural), study population, design, maintaining of blinding, clinical assessments (schedule of time and events), parameters of efficacy and statistical methodology. In short, the patients enrolled in the Mayo Clinic trial were typical of most series of patients with PBC with regards to gender (ca. 90% of the patients were women) and other demographic characteristics. Indeed, the differences between the two treatment groups (UDCA vs PL) are most likely due to the effect of the test medication because pre-drug (baseline) the groups were reasonably balanced with regards to demographic characteristics, associated diseases, pharmacologic treatment 3 months prior to randomization into trial as well as in disease baseline characteristics. The latter included proportion of patients with jaundice, pruritus, fatigue, hepatic biochemical markers, immunoglobulins (except as noted next), bile acids in duodenal contents, presence of esophageal varices, Mayo Risk score, hepatic histology, etiological factors and surgical history (Tables 9 and 10 in the text of the MO's review). The stratification/randomization process was therefore properly executed and resulted in the assignment of patients to each of the two treatment groups in a way that assured comparability of UDCA and control groups in all the variables examined. The exception was IgM. Baseline IgM antibody titer was significantly higher in the UDCA group than in the PL group. By this parameter, the UDCA group appear to have "more disease" but this does not seem tenable because - as already noted - the two experimental groups were well-balanced with regard to all the rest of baseline characteristics. From the practical viewpoint, the disbalance in IgM antibody prior to test medication is not of concern because it may, if anything, disadvantage the UDCA group.

As stated, demonstration of certain important clinical effect necessitated the administration of UDCA for extended periods (4 years). In the material that follows these effects are mentioned first. But many other effects were

demonstrated already after 2 years of double-blind therapy. These findings are addressed next. This approach was important to demonstrate efficacy with the two different formulations used in the trial: one formulation during the double-blind phase of the trial, the other during the 2-year extension period.

Results of the Mayo Clinic trial demonstrated superiority (therapeutic gain=8%) of UDCA, 13 to 15 mg/Kg, given in four divided doses, for 4 years or longer over PL in the development of death/transplant. The difference between the treatment groups was highly significant when using a Cox proportional hazards model after adjusting for the Mayo Risk Score. This statistical approach (introduction of the Mayo Risk Score) improved the sensitivity of the analysis since it resulted in a reduction in the variability in the response.

Evaluations of treatment failure were reassessed using an FDA recommended definition which excluded doubling of bilirubin at baseline and the inclusion of voluntary withdrawal in the analyses. Long-term administration of UDCA significantly decreased the proportion of patients failing (therapeutic gain=18%) and delayed significantly (therapeutic gain=268 more days) the mean days to failure, at a p-value of ≤ 0.002. Actually, using the revised definition, the superiority of UDCA over PL in time to treatment failure was already shown at the 2-year cutoff analyses. But, at this time and understandably so, the differences between the two groups were less marked than at the extended follow-up analysis (therapeutic gain in the proportion of patients that failed=10%, $p \leq 0.03$ and a therapeutic gain of 110 days in the mean days to failure).

Using the original, protocol-defined treatment failure, evaluations at 2 years of double-blind treatment showed the following. UDCA was superior to PL in the proportion of patients failing (therapeutic gain=24%, $p<0.01$), in the time to treatment failure (therapeutic gain=163 days, $p<0.0001$). These differences were convincingly shown among patients who, at baseline, had a serum bilirubin level of ≤ 1.8 mg/dl (therapeutic gain=166 days, $p<0.03$). Among those patients with a baseline screening bilirubin level of >1.8 mg/ml, the therapeutic gain was 124 more days to treatment failure. Although this difference was statistically significant by the log-rank test, it was N.S. ($p=0.06$) by the Wilcoxon test.

More importantly, at 2 years, using the protocol-stipulated definition of treatment failure, it did not matter whether the patients had grade I & III or grade III & IV liver histology at baseline. Among patients with early histologic stage at baseline, the therapeutic gain was 80 days ($p=0.02$). Among those with late histologic stage, the therapeutic gain was even greater, 180 days ($p=0.0003$). For the clinically important treatment failure parameter, both patients with early as well as those with advanced disease benefited from UDCA.

At 2 years, using the original definition of treatment failure included in the protocol, treatment failure was still statistically significant in favor of the UDCA-treated group when both baseline bilirubin and histologic stage were

entered into the Cox's proportional hazard model and logistic regression analyses were performed ($p<0.001$).

Life table analyses of patients who did not have esophageal varices at baseline showed already a therapeutic gain of UDCA over PL of 4% at the 2-year cutoff ($p=0.04$ by Log-rank test; N.S. by Wilcoxon test). This therapeutic gain, favoring those patients initially assigned to UDCA, increased to a clinically meaningful 18% at the end of the extended follow-up period ($p=0.003$ by Log-rank test and 0.004 by Wilcoxon test). With regard to the mean days to developing varices, the therapeutic gain of 128 days at the 2-year cutoff was even longer at the end of the extended follow-up period (201 days).

As expected, a 2-year course of UDCA therapy showed highly significant improvement in the change from baseline for practically all hepatic biochemical markers, except albumin. As a rule, the changes in the two groups were divergent in that, with the progress of time, the abnormally high values at baseline decreased (or got better) among the patient treated with UDCA but increased (or got worse or did not change as much) among those randomized to the PL group.

Long considered a very important parameter of liver function, total serum bilirubin - which, at entry, was similar in the two experimental groups - showed statistically significant changes to the 2-year endpoint from baseline. But whereas bilirubin decreased by a mean of 0.63 mg/dl among those patients treated with UDCA, those treated with PL experienced an increase (they got worse) in bilirubin of 0.80 mg/dl. Also impressive was the effect on AP. Whereas this parameter decreased by an average of 708 IU/l in the UDCA-treated group, it increased by an average of 15 IU/l in those patients randomized to PL. An additional example of divergent trends was seen from analysis of IgM antibody results. We have already commented that, at baseline, IgM was significantly higher among those patients randomized to UDCA in comparison to those randomized to PL. In spite of this disadvantageous start, a 2-year course of UDCA therapy resulted in a mean decrease in IgM of 152 from baseline which contrasted with the 32 increase seen in those patients randomized to PL. This marked fall in IgM (together with less marked changes in IgA, IgG and γ-globulin) supports the proposal that UDCA may at least in part, exert an immune effect [Y. Calmus et al., Hepatology 1:12-15 (1990); Y. Calmus et al., Gastroenterology 103:617-621 (1992)]. The differences between the two treatment groups for IgM, bilirubin, AP and SGOT were all highly significant ($p<0.0001$). It is also important to note that several groups of investigators consider serum bilirubin levels the most important variable in estimating survival in patients with PBC [J. Roll et al., NEJM 308:1-7 (1983); E. Christensen et al., Gastroenterology 89:1084-1091 (1985); E.R. Dickson et al., Hepatology 10:1-7 (1989)].

Since data on laboratory variables at the end of the extended follow-up period and upon discontinuation of the treatment were not presented by the sponsor, it is not known if further improvement in laboratory markers of liver disease

is obtained with longer period of treatment and whether, as reviewed by others [R.A. Rubin et al., Ann. Intern, Med. 121:207-218 (1994)], the improvements in laboratory variables are limited to the treatment period.

After a 2-year course of test medication, there was a statistically significant difference in the Mayo Risk score in both treatment groups. Again, a decrease in the UDCA-treated group contrasted with an increase in the PL-treated group. The difference in change in Mayo Risk score at endpoint from baseline between the two treatment groups was highly significant (p=0.0001) in favor of the UDCA-treated group. To a certain extent, these findings are not unexpected, as the Mayo Risk score depends heavily on parameters that changed - individually - in favor of UDCA.

On overall analyses, the Mayo Clinic trial did not show clinically important or statistically significant differences between UDCA and PL in pruritus or fatigue, two important symptoms of PBC. Although it is possible that pruritus may be markedly relieved in some individual patients treated with UDCA, while in other, this symptom may be actually worsened, it is important to consider recent publications on why cholestatic patients itch. It is worth mentioning that this subjective complaint is hard to quantify. This difficulty has hindered rigorous analysis of the effect of medical therapy on pruritus. The symptom occurs and/or intensifies at certain, so far unpredictable, intermittent periods of time and there is a differential patient threshold of symptom recognition. It is also of interest to mention that the mechanism mediating pruritus in patients with PBC has not been conclusively elucidated. Invoked theories include bile salt deposition in the skin [L.J. Schoenfield et al., Nature 213:93-94 (1967)], release of a pruritogen as a result of the detergent action of certain bile salts on the hepatocyte membrane [C.N. Ghent, Amer. J. Gastroenterol. 82:117-118 (1987)] and the more recent theory that endogenous opiates, which accumulate in PBC and other chronic cholestatic liver diseases, may mediate pruritus through a central origin mechanism [E.A. Jones and N.V. Bergasa, Hepatology 11:884-887 (1990)]. In a recent review on the subject matter, evidence is presented that increased opioidergic neurotransmission/neuromodulation (tone) in the CNS seems to contribute to the pruritus of cholestasis [E.A. Jones and N.V. Bergasa, Gut 38:644-645 (1996)]. These new findings and proposals provide a rationale for administering orally bioavailable opiate antagonists to obtain long-term relief from the cholestatic form of pruritus. Furthermore, it is conceivable that, as proposed by H. Schworer et al. [Z. Gastroenterol. 33:265-274 (1993)], altered serotoninergic neurotransmission may contribute to the pruritus of cholestasis. Preliminary data have been published suggesting that ondansetron, a 5-HT$_3$ serotonin receptor antagonist, may ameliorate the pruritus of cholestasis [H. Schworer et al., Pain 61:33-37 (1995)]. These considerations suggest that in PBC patients with pruritus, therapeutic modalities in addition to UDCA may be needed to exert significant and consistent antipruritus effects.

The Mayo Clinic trial made use of a well-established criteria for staging of chronic nonsuppurative destructive cholangitis (syndrome of PBC), developed at

APPEARS THIS WAY
ON ORIGINAL

and non-compliance. The FDA suggestion of not including voluntary withdrawals in the assessment of treatment failure was a sound recommendation since indeed, in only some of these patients the reason for voluntary withdrawals may have been insufficient therapeutic effect. Although this approach did not change the overall conclusion on efficacy, it improved the quality of the data used to draw sound conclusions.

The minor changes in vital signs and except as noted next, in laboratory parameters, do not appear to be of clinical importance. The changes in the two treatment groups in bone mineral density and chest x-rays and endoscopy findings were not statistically significant. Total fasting serum cholesterol levels decreased significantly in the UDCA in comparison to less marked decreases seen in the PL group. This fall in cholesterol in PBC patients treated with UDCA was also seen in the Canadian Multicenter trial (reviewed in the next section) and was previously reported by Poupon et al [NEJM 324:1548-1554 (1991)].

There are a number of proposed mechanisms by which the reduction in serum cholesterol level may occur. These include decreased cholesterol absorption from the gut, because of diminished micelle formation and increased cholesterol secretion from cell to lumen, with subsequent desquamation and loss in the feces. There may be an effect of UDCA on lipoprotein metabolism. UDCA apparently directly stimulates receptor-dependent LDL uptake in the liver and decreases HMG-CoA reductase [A. Lanzini and T.C. Northfield, Gastroenterology 95:408-816 (1985); H. Fromm, Dig. Dis. Sci. 34:21S-23S (1989)]. These effects of UDCA in lowering serum cholesterol levels in PBC patients have also been reported in hypercholesterolemic children with intrahepatic cholestasis [W.F. Balistreri, in: Bile Acids as Therapeutic Agents; G. Paumgartner et al., editors, Kluwer Academic Publishers, London, Chapter 40, pp. 323-333 (1991)].

In summary, the Mayo Clinic trial made use of a well-designed protocol with an appropriate control (placebo) and was apparently well-executed. The MO documented, in detail, adherence to all aspects of the protocol, including stratification/randomization, and procedures to maintain the double-blind nature of the study and other methods to minimize bias. The in-depth review of the detailed submitted data resulted in sound basis for conclusions. The trial was conducted in two phases. The first phase (2 years) consisted of observations under double-blind conditions. At the end of this phase, UDCA was superior to PL in the proportion of patients failing and in the number of days to treatment failure, in the development of (significantly less) varices and in the improvement in practically all hepatic biochemical markers (including bilirubin, AP, SGOT and IgM antibody titer) and in decreasing the Mayo Risk score, an important predictor of survival in PBC. The double-blind phase was followed by a long-term phase of 4 years consisting of open-label observations. Observations at the end of this open-label extension of the trial were required to demonstrate UDCA effects on important clinical endpoints. In patients with PBC, long-term treatment with UDCA significantly improved survival or the need for liver transplant, delayed the time to

treatment failure (even further than at the 2-year cutoff period) and reduced even further the risk of developing esophageal varices in those patients who, at randomization, had no varices.

UDCA was in general well-tolerated:

In conclusion, the Mayo Clinic trial results demonstrate efficacy of UDCA in important clinical parameters over and beyond the effects in hepatic biochemical markers. Based on results of this study, the reviewer identified an effective dose of 13 to 15 mg/Kg. This should be given in the same manner as in the Mayo Clinic trial, that is, with each of the three main meals and at bedtime [whether higher doses of this enterohepatic drug are more effective remains unanswered]. From the detailed review of the evidence on efficacy, the MO concludes that results from the Mayo Clinic trial could stand alone as a basis of approval.

D. The Canadian Multicenter Trial

"A double-blind, randomized, placebo-controlled, multicenter trial of ursodeoxycholic acid in primary biliary cirrhosis"

This trial, sponsored by the Medical Research Council of Canada (Grant #MA10305), was carried out at eleven centers throughout Canada. The principal investigator and clinical contact was E.J. Heathcote, M.D., from the Toronto Hospital Western Division. The test medication was provided by Axcan Pharma (Canada) Inc. [redacted]

The study was initiated on April 1, 1988 and completed on July 1992. The initial centers were Toronto Western Hospital, Toronto General Hospital, Sunnybrook Medical Center (Toronto) and the University Hospital in London. Other centers were added from October 1988 on.

1. Hypothesis to be Tested

This trial was set to test the hypothesis that UDCA, a choleretic drug, will prevent the hepatic damage secondary to cholestasis in PBC and slow the progression of this disease as judged by clinical and laboratory markers of cholestasis. Clinically, cholestasis may be manifested by the presence of pruritus and/or jaundice. The most reliable laboratory marker of deterioration of PBC is cholestasis. Cholestasis is manifested by a steady rising in serum bilirubin, which, in turn, provides a suitable marker for patient assessment. Additional markers of cholestasis include bile acids, γ-glutamyl-transpeptidase, AP and serum transaminases.

2. Objectives

The primary objective was to:

- Compare the effects of UDCA versus placebo on the percentage rise in serum bilirubin at two years.

The secondary objectives were to:

- Compare the effects of UDCA vs PL on total serum BAs, AP, AST, ALT, γ-GT, total serum CHOL, ALB and Ig levels, Hb, platelet count and PTs.
- Evaluate the effects of UDCA on signs and symptoms of PBC including fatigue, pruritus, ascites, xanthelasma and encephalopathy.
- Determine the effects of UDCA on histologic stage at two years.
- Determine whether UDCA favorably affects survival or the need for transplantation.
- Assess toxicity and determine the safety and tolerability of UDCA.

3. Trial Design

This trial was conducted in a randomized, stratified, double-blind, multicenter manner. This 2-arm study was placebo-controlled. The protocol specified that 202 PBC patients were to be enrolled over a 3-y period. Previous to randomization to either UDCA or PL, the patients were stratified into two groups, symptomatic and asymptomatic. Each center received approval from their respective Ethics Review Boards. Written IC was obtained from each patient prior to randomization into the trial.

The randomization within each stratified group was done by a designated outpatient pharmacist at each trial center, according to randomization tables supplied by the McMaster University trial statistician. Each trial center was randomized separately. All others involved in the trial (except the pharmacist) did not know the identity of the test medication.

4. Study Population (Table 36)

As summarized in this Table, the inclusion-exclusion criteria were adequate for this type of study. To be admitted into the trial, patients were required to have biopsy-proven PBC, cholestatic disease, increased AP and a positive AMA_s test. Equally adequate were the exclusion criteria.

Each site investigator was asked to complete a Pre-Trial Eligibility form for all PBC patients at their respective center. The original copy was filed and kept with the trial chart. Copies were sent to the Clinical Research Coordinator (Toronto) who decided eligibility and assigned an ID number according to the randomization scheme at the study center.

TABLE 36
The Canadian Multi-Center Trial

Characteristics of the Study Population

INCLUSION CRITERIA[a]	REASONS FOR EXCLUSION
• M and F patients over the age of 18y who had been diagnosed with PBC. • PBC was defined as follows: - Confirmed cholestatic liver disease - Elevated serum AP levels (defined as being greater than the ULN for that specific laboratory) - Positive AMA_s test (AMA_s titer ≥1:20) - Liver Bx in the previous 12 months diagnostic of or compatible with the diagnosis of PBC. • In addition, patients had to agree to attend a trial clinic every 3 months for the duration of the study and to agree to stop taking concurrent medication with the exception of cholestyramine for PBC.	• Taking enzyme inducers that could not be discontinued (e.g., antiepileptics). [Patients who were taking enzyme-inducing drugs that could be exchanged for an equally effective non-enzyme inducing drug, were allowed to enter the study.] • Active transplant list. • Taking any other trial drug (e.g., colchicine). • Pregnancy. • Presence of any other serious, co-morbid condition(s). [Each case was decided on an individual basis. A patient was excluded following a diagnosis of cancer if the disease-free interval was less than 5 years.]

a) Patients were considered symptomatic if they had (a) any pruritus (b) any jaundice (even if not evident clinically), e.g. serum BIL greater than 20 μmol/L, (c) fatigue combined with either pruritus or jaundice, or (d) xanthelasma combined with either pruritus or jaundice. Asymptomatic patients were recruited only until July 1989.

5. Randomization Procedures, Blinding, Test Medication(s), Dosage Regimen and Concomitant Medications

- Patients were first stratified into two groups, symptomatic and asymptomatic and then randomized (at each center trial separately)[38] into UDCA or PL.
- Blinding was maintained as mentioned above. The contents of the PL capsules were manufactured to be equally bitter-tasting to help prevent unblinding. A blind taste-test conducted by the research coordinator confirmed that this was true.
- Study medication (UDCA and PL) was manufactured by Beecham Canada and Schering Canada for Axcan Pharma [redacted] and distributed by the Canadian licensee, [redacted]
- The following CTM[39] was used in the trial.

[38] The randomization schedule for this trial was provided in sponsor's Appendix II.

[39] A sufficient number of capsules to treat 10 patients, together with quality control certifications, were retained by the Investigator. A four-month supply of capsules was given to the patient during each visit.

UDCA	PL
Hard gelatin capsules (Size 0) that contained 250 mg UDCA and excipients. Excipients: cornstarch, colloidal silicon dioxide, magnesium stearate, Windsor, salt, polysorbate 80 and alcohol.	Hard gelatin capsules (Size 0) identical to UDCA capsules in appearance, taste and weight. Excipients: cornstarch, lactose, colloidal silicon dioxide, magnesium stearate, Windsor, salt, polysorbate 80, alcohol, potassium chloride.

- Patients receiving cholestyramine to control pruritus were asked to take the drug with breakfast and lunch, or at least 4 hours prior to or after ingesting the test medication, to minimize binding to the UDCA. Patients were otherwise not permitted to receive any associated concomitant medications for the treatment of PBC (e.g., colchicine, corticosteroids, azathioprine, cyclosporin, chlorambucil or D-penicillamine were proscribed). Information regarding concomitant medications was collected at each three month follow-up visit.

6. Clinical Assessments

The study consisted of three phases: screening/eligibility phase, baseline phase and treatment phase.

a. Screening/Eligibility Phase

During this phase, investigators completed pre-trial eligibility (see Table 36) on all PBC patients referred to them to determine if the patient was eligible for inclusion into the trial. Liver biochemistries and a pre-trial liver Bx was assessed to confirm the diagnosis of PBC and the presence of cholestasis.

b. Baseline Phase

During this Phase, patients with well-defined PBC (as confirmed by liver Bx) who were AMA_s positive and had ↑ levels of serum AP were enrolled. Complete clinical and laboratory assessments, including abdominal U/S (if required to confirm the presence of ascites), were performed. Based upon the results of BL assessments, patients were stratified according to whether they were symptomatic or asymptomatic.

c. Treatment Phase

- During this 2-year Phase, patients were treated with UDCA or PL in a double-blind fashion, and treatment effects were assessed.
- A physical exam, hematology, immunology and biochemistry parameters, and a measurement of serum BAs were repeated every three months.

- In addition, patients were asked to complete weekly diaries that included records of their pruritus (if any); their need for antipruritic medication, including the number of cholestyramine (Questran) packets; their energy level (Visual Analog Scale); and any other untoward complaints.

- Each patient was queried about any new treatments or medications to assess the degree of co-intervention.

Subjects took two to five 250 mg capsules of clinical trial material (CTM) daily with evening meals. The individual dose depended upon their body weight. The randomized treatments (14 mg/Kg/day of UDCA or PL) were to be maintained from Day 1 to the end of the study two years later.

d. Schedule of Time and Events (Table 37)/Monitoring

As shown in this Table, patients were seen by the investigator at the initial BL visit and at 3-month intervals for the duration of the 24-month trial. Details of events and time of events are given in Table 37.

Most of the liver Bxs were performed within one year before randomization and again at 24 months and were reviewed without knowledge of treatment allocation. If a pt. was transplanted before completing 24 months, transplantation tissue was obtained and assessed as a post-trial specimen.

All Bxs were staged I through IV based on the degree of fibrosis:

I	II	III	IV
No fibrosis	Periportal Fibrosis	Fibrosis with Septa	Cirrhosis

According to the Clinical Report, this system produces, in practice, gradings very similar to those of the Ludwig system [Virchows Arch. A. Pathol. Anat. Histopathol. 379:103-111 (1978)]. Use of the latter system, which combines fibrosis and necroinflammatory activity, however, would not be appropriate when treatment effects are expected to change these parameters independently [I.R. Wanless et al., Hepatology 18:A174 (1993)].

TABLE 37
The Canadian Multicenter Trial

Schedule of Time and Events

Event	Screening/ Eligibility Phase	BL Phase	Treatment Phase (months)							
			3	6	9	12	15	18	21	24
History[a]	x									
P.E.[b]	x	x	x	x	x	x	x	x	x	x
AMA_o Test[c]	x	x				x				x
Immunoglobulins,[d] IgM, IgG, IgA		x	x	x	x	x	x	x	x	x
Hematology	x	x	x	x	x	x	x	x	x	x
Clinical Chemistry	x	x	x	x	x	x	x	x	x	x
PT		x	x	x	x	x	x	x	x	x
U/S[e]		x								x
Liver Bx[f]	x									x
Stored Serum for BA[g] Determination		x	x	x	x	x	x	x	x	x

a,b) The clinical assessment checked for signs of hepatic decompensation, AEs and symptoms, including presence or absence of fatigue and severity of pruritus.
c) Measured by indirect immunofluorescence using rat kidney and stomach as substrate.
d) Measured by standard techniques [I.J. Check et al.; sponsor's Ref. 15, vol. 40, p.136].
e) As required to assess ascites.
f) In addition to the parameters listed below (I=no fibrosis; II=periportal fibrosis; III-fibrosis with septa and IV=cirrhosis),

- 5 further parameters were graded: 0 (none), 1 (mild), 2 (moderate) or 3 (severe), including lobular lymphoid inflammation, portal lymphoid inflammation, duct paucity, ductular proliferation and periportal (or paraseptal) hepatocellular ballooning (feathery degeneration).
- Duct paucity was graded: 0 (all ducts present), 1 (occasional ducts present), 2 (up to half ducts absent) and 3 (more than half ducts absent).
- Pre- and post-treatment liver biopsies were paired at the end of the trial and a comparison analysis was performed.
- For each parameter, the difference in pre- and post-trial scores was tabulated as improvement (decrease of one or more grades), no change, or progression (increase of one or more grades).

g) Total serum BA were measured by a standard method [F. Mashige et al., Clin. Chem. Acta 70:79-86 (1976)] using a 3α-hydroxysteroid dehydrogenase. Serum UDCA was measured by a fluorescence spectrophotometry method using 7β-hydroxysteroid dehydrogenase [I.A. MacDonald et al., Anal. Biochem. 135:349-354 (1983)] adapted to measuring UDCA with excellent specificity and reproducibility, using the method of Mashige [(locus cited) (1976)].

7. Criteria for Patient Withdrawal

These were all adequate.

8. Data Collection and Entry/Validation

- All data were collected and entered at the coordinating center (Toronto Western Hospital) study site by the Clinical Research Coordinator. The procedures for data collection and entry, as described by the sponsor in the Clinical Report (vol. 40, p. 032) were all adequate.[40]

- With regard to validation, the sponsor states the following.

> Within one week of the double data entry, the Research Coordinator completed SAS® comparison procedure of the entire DATABASE1 versus DATABASE2, and data entry mismatches were identified. The comparison procedure was rerun the same day until both databases were identical. The databases were backed-up on diskette until the next batch entry. At the completion of the study, when all data entries and the final comparison procedure were completed, the Research Coordinator ran the SAS univariate procedure to check for outliers. Outlying data were checked against the CRFs, and the data listings were compared against source documents at the study center. Finally, the databases were backed up and locked for data analysis.

9. Efficacy Parameters

a. Primary Efficacy Parameter

The study protocol defined the primary outcome measure as the percentage change in serum BIL, assessed after 2 years.[41]

b. Secondary Efficacy Parameters

Secondary outcome measures of efficacy, as listed in the protocol, included:

I) change in pruritus, indicated by the patient weekly diary;
ii) change in well-being, also indicated by the weekly diary;
iii) serum laboratory measurements, i.e., AP, total serum CHOL, ASAT and ALAT, serum albumin and immunoglobulin levels.
iv) AMA_s titers;
v) liver Bx, and;
vi) death and/or liver transplantation.

NOTE: Treatment Failure was defined post-hoc (see Results).

---

[40] Photocopies of the CRFs were received by courier from each study site within one month of the patient visits. All original CRFs and clinical laboratory reports were retained and filed by the original clinics. Prior to data entry, the CRFs were logged in and reviewed for visual clarity, missing pages, missing data, and any data outside the appropriate laboratory range. If any problems were found, the Research Coordinator immediately contacted the study center coordinator until the situation was resolved. Any authorized corrections were made to the CRF and were initialed and dated. Completed CRFs were then initialed and batched for data entry. The batched CRFs were then entered by a Research Assistant into DATABASE1. There was a double data entry of the batched CRFs by the Research Coordinator into DATABASE2.

[41] Analysis of serum BIL levels in 30 symptomatic PBC patients followed for 20 to 28 months by the principal investigator had shown that, while on no treatment, the rise in their serum BIL was >50% in 14 patients.

10. Safety Evaluations

The methods to collect, record, assess relationship to drug and report AEs were all adequate.

11. Statistical Methodology

a. Power and Justification of Sample Size

- Based on an outcome of percentage rise in BIL levels over a two-year period, the sample size was estimated from preliminary estimates of effect from patients in the practice of the principal investigator. From prior experience, 47% of eligible patients would go on to have a minimum of a 50% increase in BIL level over 2 years. For UDCA to be considered clinically effective, this porportion of 47% would need to be reduced by half.

- A risk reduction from 47% to 23.5% would require 85 patients per group (total 170, $\alpha = 0.05$, power = 90%). To take into account that an interim analysis was proposed at 33 months, this necessitated two tests of significance on the primary outcome. Thus, the sample size was increased accordingly, $\alpha=0.05/2=0.025$. The final estimate was 101 patients per group or 202 in total.[42] The sponsor noted that while it would have been possible to anticipate a smaller sample size by testing the difference between the means for BIL, the variation would be so great that it was not feasible.

b. Interim Analysis

According to the protocol, the interim analysis was to be conducted after 33 months or when at least 50% of the required sample had completed 2 years, whichever occurred first. The statistician provided "unidentifiable" data to the Data Monitoring Committee to be set up at the time of interim analysis. This Committee was to decide if the trial should be stopped because of side-effects or lack of efficacy of the drug, or if the drug was so effective that administration of PL was no longer justified. This information was then given to the PI.

In summary, the sample size calculation included a provision of one interim analysis. In May 1991, the interim data were presented to a monitoring committee composed of three physicians not associated with any of the trial centers. The data were reviewed while remaining blinded to treatment allocation and using Fisher's exact method. The committee recommended the study continue to completion.

---

[42] A slightly different version was given in the Clinical Report. In this document it was stated that the required sample size was calculated using data on the change in total serum BIL observed over two years in a population of 100 PBC patients living in Ontario. By assuming that an effect of UDCA treatment similar to that reported by Poupon et al. (1991) would be observed, it was determined that 204 patients would be required to demonstrate a 50% reduction in total serum BIL in the UDCA group with 90% power ($\alpha = 0.05$, 2-tailed, $\beta=0.20$).