Statistical Review and Evaluation

FDA Number:	99-2865 (STN 103979/0)
Task/Type:	BLA
Sponsor:	Genzyme Corporation
Product:	Human α-Galactosidase A (r-hαGAL)
Indication:	Fabry Disease
Date:	9/20/00
From:	Clara Chu, Ph.D./HFM-219
Through:	G. Gupta, Ph.D./HFM-219
To:	Dr. Jim Kaiser/HFM-573
CC:	original/DCC/HFM-99 ChronFile/HFM-210 Dr. S. Ellenberg/HFM-215 Dr. P. Lachenbruch/HFM-215

Background

Genzyme Corporation has submitted an application for the approval of recombinant human α-galactosidase (r-hαGAL) in patients diagnosed with Fabry disease. This submission consists of a Phase I/II study (FB9702-1) and a Phase III study (AGAL-1-002-98). This review will discuss only the Phase III study.

Design

This is a multi-center, placebo-controlled, double-blind, randomized study of patients diagnosed with Fabry disease with no prior treatment of r-hαGAL. This study has a 1:1 balanced design with up to 60 patients randomized to one of two treatment arms. Patients are randomized to receive 1mg/kg injections of either r-hαGAL or placebo every two weeks for a total of 11 infusions. The sponsor staff, investigator and staff and patients are blinded to what treatment the patient receives. Eight sites including three in the U.S. and five in Europe participated in this trial.

Primary endpoint

The primary efficacy endpoint is the morphologic assessment of GL-3 inclusions in the capillary endothelium (vasculature) of the kidney at 20 weeks (visit 11). All samples are scored by three independent pathologists, who are masked to which samples are pre- and to which are post- and to treatment groupings. The scores consist of 0 (none) through 3

(severe). At least 2 of the 3 pathologists must agree on the same score with the third pathologist having a difference of no more than 1. If there is a difference of more than 1, then the 3 pathologists must review the slide set together and attempt to decide on one score (adjudication). If this cannot be achieved, then a median score is used.

As per agreement with FDA on 3/24/00, all original (post-adjudicated) majority scores of 0 or 1 will be reread. The reread involves counting the total number of evaluable endothelial capillaries (vessels) that would be classified into each of the following scoring groups: 0 (number of clear vessels), trace (number of vessels scored as trace), 1 (number of vessels scored as mild), 2 (number of vessels scored as moderate), and 3 (number of vessels scored as severe). The total number of vessels (TOTAL COUNT) will be calculated by summing the total number of vessels in each of the 5 scoring groups. The scores from each vessel are ranked in ascending order (0 to 3). A maximum of 5% of the highest scores (involving vessels scored 1, 2, or 3) will be subtracted from the TOTAL COUNT. From the subset of remaining scores (REMAINDER), a score of 0 will be assigned if BOTH of the following criteria are met:

1. more than 50% clear vessels (100*(# of clear vessels/REMAINDER))
2. remaining vessels are clear or trace

If a slide was reread and was not assigned a score of 0, then it would be assigned a score of 1. Each pathologist will assign a score to the reread slide. From these three scores, a new majority score will be calculated. This new majority score will replace the previous majority score, if different. The original majority score will not change for any slide that was not reread.

Secondary endpoints

GL-3 accumulation of the heart, skin and kidney, as measured by light microscopy, will be scored by three independent pathologists. The majority score will be calculated for each organ and then summed across all organs to obtain a composite score. The composite score will range from 0 to 9. Change in vascular GL-3 composite score is defined as (visit 11 score – baseline score).

The percent change from baseline to visit 11 for urinary GL-3 and kidney GL-3 is defined as (visit 11 score-baseline score)/baseline score)*100. This change is ranked for each patient and the rank sum score is obtained by summing the ranks of these two parameters.

The short form McGill pain questionnaire consists of 15 questions pertaining to pain. Each question is rated a 0 (none), 1 (mild), 2 (moderate), or 3 (severe). The sum of questions 1-11 will define the sensory pain score. The sum of questions 12-15 will define the affective pain score. The sum of the sensory pain score and the affective pain score will define the total pain score. Also included are a question on present pain intensity (PPI) and a visual analog scale (VAS). The change in sensory, affective, total, PPI and VAS is defined as (visit 11 score-baseline score).

Statistical Analysis Plan

Summary statistics will be presented by treatment group for prospectively defined demographic (age, sex, ethnic background, weight, height), history of Fabry disease (family history of Fabry disease, gender of family members with Fabry disease, time since diagnosis, blood type, αGAL activity, leukocyte alpha-GAL concentration), and medical/surgical history variables (medical/surgical history of body system). Frequencies and percents will be presented for categorical variables, and descriptive statistics (N, mean, standard deviation/standard error, minimum, maximum) will be presented for continuous variables. The chi-square test will be used to test for significant differences between treatment groups for frequency variables, and the t-test will be used for continuous variables.

A chi-square test will be used to analyze treatment differences on the primary endpoint. If the expected cell frequency of any cell is less than 5, then Fisher's Exact Test will be performed. The Mantel-Haenszel test will be conducted on the same frequency tables stratified by site, age group and ethnicity. A second chi-square test will be used to analyze the same proportion of patients in each treatment group but at baseline. If there is an imbalance at baseline, then a logistic regression will be conducted using the proportion of patients with a score of '0' at visit 11 as the dependent variable and the proportion of patients with a score of '0' at baseline and treatment group as the independent variables.

Scoring of the primary endpoint will be performed by three independent pathologists. The Intra-class correlation between the three pathologists and the three tissue types will be calculated using the Fleiss statistic (Woolson, 1987). The Fleiss statistic will only be calculated for the post adjudicated majority score that is used in the primary analysis.

A sensitivity analysis will also be conducted to determine the effect of various cut-off percentages for excluding vessels classified as 1, 2, and 3 from the TOTAL COUNT variable. The current analysis plan uses a maximum of 5% as the cut-off percentage.

An analysis of variance will also be used to test for a significant difference in 20-week mean change scores between treatment groups and study sites. It is likely the number of patients per study may be too limited to test for the homogeneity of treatment effects across sites. Therefore, an interaction term will not be included in the model.

For the Intent-to-Treat (ITT), As-Treated (AT), and Per-Protocol (PP) populations, descriptive statistics (N, mean, median, standard deviation/error, minimum, maximum) will be presented by treatment group at baseline, at visit 11, and for changes from baseline to visit 11. Mean differences and proportions by treatment group will be presented with p-values and 95% confidence intervals.

Secondary endpoints will be analyzed and presented similarly to the primary endpoint. Hypothesis testing for the percent change in urinary and kidney GL-3 accumulation will be performed using a two-sample ---------- Sum test.

Missing values on the primary endpoint at visit 11 for the ITT and AT populations will be assigned the worst possible score (a non-zero score). For the PP population missing values will be excluded and any baseline values with a score of 0 will be excluded. For secondary endpoints, the Last Observation Carried Forward (LOCF) will replace missing values. This method will not favor rejection of the null hypothesis of no difference between treatment and control groups in favor of treatment. If the null hypothesis were true, then it would be expected that GL-3 score would change at the same rate in both treatment and control groups. By carrying the last score forward, the patient would have no change in score from baseline to final visit. Therefore, this method would not favor treatment group by being inappropriately declared significantly better than placebo. For the short form McGill questionnaire, sensory, affective and total scores will be summed from questions 1 through 15 even if there are missing values.

The ITT population consists of all patients randomized, where missing values at visit 11 were assigned the worst outcome. The AT population consists of all patients assigned to the treatment group representing the treatment they actually received. Missing values at visit 11 were assigned the worst outcome. The PP population consists of patients assigned to the treatment group representing the treatment they actually received. Missing values were excluded from the analysis, patients receiving both treatments during the course of treatment were excluded, and patients with a majority score of 0 at baseline were excluded.

Results

There were 58 patients in the ITT population with 29 randomized to r-hαGAL and 29 to placebo. The AT population consisted of 58 patients with 29 receiving r-hαGAL and 29 receiving placebo. However, there were 6 patients who did not receive the treatment they were originally randomized to receive. The per-protocol population consisted of 54 patients (26 in r- hαGAL and 28 in placebo). Since approximately 10% of the patients in the ITT population did not receive the treatment that they were assigned to receive, the clinical reviewer felt it was important to consider the AT population as the being the most clinically relevant population to analyze. Hence, all tables will show the results for the AT population.

The treatment arms are balanced with respect to the prospectively defined demographic, disease history and medical/surgical history variables in the ITT and AT populations. Table 1 shows a subset of the demographic and disease history variables. Baseline demographic, disease history and medical/surgical variables were also balanced in the ITT population.

Table 1. Subset of the Baseline Demographic and Disease History Variables

		Treatment		
Baseline variables	**Summary Statistics**	**r-ha GAL**	**Placebo**	**p-value**
Age	N	29	29	0.20
	Mean	32	28	
	Med	33	24	
	Std	9.38	11.04	
	Min/Max	16/48	17/61	
Sex	Male	27	29	0.49
	Female	2	0	
Ethnic background	Caucasian	27	26	1.0
	Hispanic	2	3	
Weight	N	29	29	0.46
	Mean	67.3	69.6	
	Med	68.0	68.6	
	Sd	9.87	13.42	
	Min/Max	49.5/85.9	46/96.1	
Height	N	29	29	0.96
	Mean	175.7	175.6	
	Med	175.6	173.5	
	Sd	8.26	8.23	
	Min/Max	157.5/196	156/203.2	
Blood type	A+,A-,O+,O-	26	28	0.61
	B+,B,AB+,AB-	3	1	
Family members diagnosed	Yes	27	28	1.0
	No	2	1	
Endogenous αGAL activity	N	29	29	0.30
	Mean	0.65	0.49	
	Med	0.40	0.30	
	Sd	0.62	0.55	
	Min/Max	0/1.5	0/1.5	
Yrs since onset of symptoms	N	29	28	0.20
	Mean	24.10	20.39	
	Med	23	17	
	Sd	10.51	10.92	
	Min/Max	8/44	8/55	
Yrs since diagnosis	N	28	28	0.89
	Mean	10.11	10.43	
	Med	6	9	
	Sd	8.26	9.22	
	Min/Max	0/28	0/46	

A chi-square test comparing the primary endpoint, proportion of patients with a score of 0 at visit 11, by treatment group, showed that there is a statistically significant difference in favor of r- hαGAL.

Table 2. Visit 11 Majority Scores by Treatment Group

	Treatment				
Visit 11 Majority Score	**r-ha GAL**	**Placebo**	**Odds Ratio**	**95% CI**	**p-value**
0	20	0	0.008	0.00,0.14	< 0.0001
1	9	29			

From Table 2, there is a statistically significant difference between treatment groups in the proportion of patients assigned a majority score of 0. This result was confirmed in the ITT (<0.0001) and the per-protocol (p< 0.001) populations.

The Mantel-Haenszel chi-square test stratifying on site, race and age showed that the statistically significant difference in favor of r-hαGAL remained for the ITT, AT, and PP populations. Tables 3 and 4 shown below are for the AT population.

Table 3. Visit 11 Majority Scores by Treatment Group and Site

		Majority score				
Site	**Treatment**	**0**	**1**	**OR**	**95% CI**	**p-value**
1	Placebo	0	10			
	r-hαGAL	8	2			
2	Placebo	0	1			
	r-hαGAL	2	0			
3	Placebo	0	3			
	r-hαGAL	1	3			
4	Placebo	0	1			
	r-hαGAL	1	0			
5	Placebo	0	4			
	r-hαGAL	0	2			
6	Placebo	0	2			
	r-hαGAL	3	0			
7	Placebo	0	5			
	r-hαGAL	3	1			
8	Placebo	0	3	0.0575	(0.0139,0.2378)	<0.0001
	r-hαGAL	2	1			

An sensitivity analysis revealed that the statistically significant difference in favor of r-hαGAL (p-0.0001) remained after excluding site 1. The odds ratio was 0.0820, and the 95% confidence interval was (0.168, 0.4011).

Table 4. Visit 11 Majority Scores by Treatment Group by Race and Age

		Majority Score		
Race	**Treatment**	**0**	**1**	**p-value**
Caucasian	Placebo	0	26	<0.0001
	r-hαGAL	19	8	
Hispanic	Placebo	0	3	
	r-hαGAL	1	1	
Age	**Treatment**	**0**	**1**	**p-value**
< 30	Placebo	0	18	<0.0001
	r-hαGAL	9	2	
≥ 30	Placebo	0	11	
	r-hαGAL	11	7	

A second chi-square test was performed to analyze the same proportions of patients in each treatment group only at baseline. In the analysis of majority scores, it was revealed that the two treatment groups did not differ. Therefore, a logistic regression was not performed.

Consistency of Scoring

The sponsor analyzed the intra-class correlation between the three independent pathologists using the Fleiss statistic (Woolson, 1987), which is a generalized kappa statistic. This analysis was performed using a --------- macro, ----------, which was originally obtained from the ----- website and modified by Genzyme. This reviewer was unable to review this macro in detail due to time constraints.

To determine whether the scoring was done consistently for the pathologists, this reviewer analyzed the post-adjudicated, post-quantitated majority scores by treatment group for each pathologist using Fisher's exact test.

Table 5. Visit 11 Majority Scores by Treatment Group for Each Pathologist

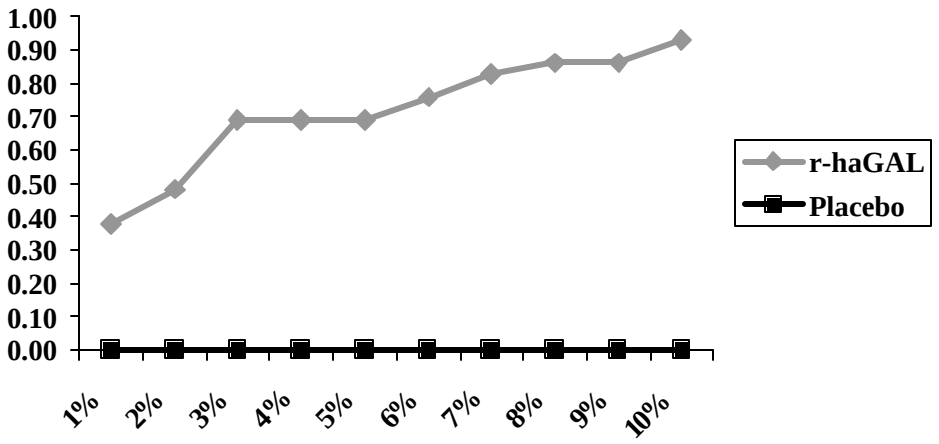
		Majority Score			
Pathologist	**Treatment**	**0**	**3 1**	**Proportion of 0's**	**p-value**
Rennke	Placebo	0	29	0/29	< 0.0001
	r-hαGAL	21	7	21/28	
Colvin	Placebo	0	29	0/29	< 0.0001
	r-hαGAL	12	16	12/28	
Dikman	Placebo	0	29	0/29	< 0.0001
	r-hαGAL	27	1	27/28	

Table 5 shows that the proportion of 0 majority scores in the r-hαGAL group is significantly higher than that of the placebo group. This result is consistent for all three pathologists and was confirmed in the ITT population.

Sensitivity analysis

A sensitivity analysis was performed that varied the cutoff percentage from 1% to 10%. New majority scores were generated for each cutoff value. The proportion of majority scores differed significantly between treatment groups for each cutoff value as shown in Figure 1. From the graph, the lower the cutoff percentage, the smaller the proportion of 0's and the higher the cutoff percentage, the higher the proportion of 0's. This analysis shows that the proportion of 0's in the r-hαGAL group remains significantly higher than the proportion of 0's in the placebo group regardless of the cutoff percentage. These results were also confirmed in the ITT population.

Figure 1. Cutoff percentages by proportion of 0 majority scores



Exploratory analyses on primary endpoint

At the request of the clinical reviewer, exploratory subgroup analyses were performed on the r-hαGAL group. The following baseline factors were examined to determine if there was any effect on the change in majority score from baseline to visit 11: age, plasma GL-3 level, urine GL-3 level (measured in nmol/filt), and kidney GL-3 level. The distributions of each of these variables was examined to determine cutpoints for the analysis using frequency tables. It was then decided to split age, plasma GL-3 level, and kidney GL-3 level at the quartiles and to split urine GL-3 level at the median. Frequency table analysis was used to examine the relationship between each variable on the change score. Since there was an ordering for both the row and column variables, the Jonckheere-Terpstra test (Lehmann, 1973) was used to look for any association. The results are displayed in Table 6.

Tables 6 and 7 show that there is no distinct pattern in the distribution of the counts of patients across age group, plasma GL-3 level and urine GL-3 level vs. change scores. Hence, the p-values from the Jonckheere-Terpstra test are not statistically significant. For kidney GL-3 level there appears to be a trend, where those patients with a higher baseline GL-3 level tend to have better change scores. However, due to the limited number of patients in this analysis, the result is not statistically significant.

Table 6. Change Score by Age in r-hα GAL

	Change Score					
Age	**-3**	**-2**	**-1**	**0**	**2**	**p-value**
≤ 21.5	0	3	1	1	0	0.94
(21.5, 29.5]	3	1	1	0	1	
(29.5, 40]	1	5	2	2	0	
> 40	1	4	2	1	0	

Table 7. Change Score by Plasma, Kidney, and Urine GL-3 in r-hα GAL

	Change Score					
Plasma GL-3	**-3**	**-2**	**-1**	**0**	**2**	**p-value**
≤ 6.3	0	3	3	2	0	0.22
(6.3, 13.55]	0	5	1	0	0	
(13.55, 20.6]	2	4	1	0	0	
> 20.6	3	1	1	2	1	
Kidney GL-3	**-3**	**-2**	**-1**	**0**	**2**	**p-value**
≤ 3026	0	3	3	2	0	0.10
(3026, 5848]	1	5	1	1	0	
(5848, 9624]	2	1	2	1	0	
> 9624	2	4	0	0	1	
Urine GL-3	**-3**	**-2**	**-1**	**0**	**2**	**p-value**
≤ 1396.1977	2	8	3	3	0	0.79
> 1396.1977	3	5	3	1	1	

ANOVA

An analysis of variance was performed on the change scores, calculated as the score obtained at visit 11 minus the baseline score. The model included both treatment group and site as independent variables. In both the ITT and the AT populations, treatment group was statistically significant in favor of r-hαGAL. These results were also confirmed in the ITT population.

Table 8. Mean Change by Treatment Group and Site

	Change Score				
Treatment	**N**	**Mean**	**St. Dev**		**p-value**
Placebo	29	-0.07	1.13		< 0.0001
r-hαGAL	29	-1.55	1.15		
Site					
1	20	-1.1	1.37		0.3078
2	3	-1.0	2.0		
3	7	0.0	1.53		
4	2	-2.0	1.41		
5	6	-0.33	0.82		
6	5	-1.0	0.71		
7	9	-0.78	1.48		
8	6	-0.67	1.51		

Secondary endpoints

Kidney, heart, skin and composite (sum of all three) change scores were analyzed using t-tests to determine if there were differences between treatment groups.

Table 9. Kidney, Heart, Skin and Composite Mean Change by Treatment Group

Endpoint	Treatment	N	Mean change	Std Dev	Diff in Means (r-pl)	95% CI	p-value
Kidney	Placebo	29	-0.07	1.13	-1.48	(-2.08, -0.88)	< 0.0001
	r- hαGAL	29	-1.55	1.15			
Skin	Placebo	29	-0.10	0.98	-2.03	(-2.49, -1.58)	< 0.0001
	r- hαGAL	29	-2.14	0.74			
Heart	Placebo	29	0.24	0.79	-0.79	(-1.19, -0.39)	< 0.0002
	r- hαGAL	29	-0.55	0.74			
Composite	Placebo	29	0.07	2.05	-4.31	(-5.33, -3.30)	< 0.0001
	r- hαGAL	29	-4.24	1.81			

Table 9 shows there is a statistically significant difference in favor of r-hαGAL. The means in the r- hαGAL tend to have larger negative differences, which suggests patients in the treated group obtain better scores at visit 11. Similar results were obtained in the ITT population.

An analysis on percent change in urine GL-3 and kidney tissue GL-3 data was performed using the ---------- sum test. According to the sponsor's analysis, the urine GL-3 analysis consists of only 42 patients (21 per group). Patient 307 was excluded due to an inability to obtain a kidney sample. The clinical report states that patients in sites 7 and 8 were excluded because at site 7, urine was not collected at baseline and at site 8, a centrifugation method used rendered the samples unevaluable. This reviewer found a discrepancy in what was stated in the clinical report regarding site 7 as there was baseline

urine data collected at site 7. Genzyme's explanation was that handling issues related to the filtration procedures compromised these samples.

Table 10 shows the sponsor's results of the percent change in urine GL-3 and kidney tissue GL-3 levels.

Table 10. Percent change in Urine GL-3 and Kidney Tissue GL-3 by Treatment Group

Secondary Endpoint	Statistics	Placebo	r- ha GAL	p-value
% Change Urine GL-3	N	21	21	0.005
	Median	42.80	-23.31	
	Range	-79.67/1558.99	-83.91/3491.57	
% Change Kidney Tissue GL-3	N	28	27	0.26
	Median	-6.20	-34.13	
	Range	-89.89/1710.43	-97.32/205.32	
Rank Sum Score	N	21	20	0.004
	Median	48	32.5	
	Range	26/73	4/67	

This analysis shows that the median is significantly lower for the r- hαGAL group for the Percent change in urine GL-3 and the rank sum scores of both these parameters, but not percent change in kiney tissue GL-3. These results were also confirmed in the ITT population. This reviewer also performed the ---------- Sum test on the AT population without excluding individuals from sites 7 and 8. The results are shown in Table 11.

Table 11. Percent change in Urine GL-3 and Kidney Tissue GL-3 by Treatment Group

Secondary Endpoint	Statistics	Placebo	r- ha GAL	p-value
% Change Urine GL-3	N	29	28	0.12
	Median	15.88	-8.49	
	Range	-80.30/1558.99	-83.91/3491.57	
% Change Kidney Tissue GL-3	N	28	27	0.26
	Median	-6.20	-34.13	
	Range	-89.89/1710.43	-97.32/205.32	
Rank Sum Score	N	28	27	0.07
	Median	60.5	49.0	
	Range	30/98	5/98	

This analysis shows that the percent change in urine GL-3 and rank sum score parameters are no longer significantly different.

An analysis of the short form McGill pain scores was performed, where differences were examined using t-tests.

Table 12. Mean change in McGill Pain Parameters by Treatment Group

Endpoint	Treatment	N	Mean change	Std Dev	Diff in Means (r-pl)	95% CI	p-value
Sensory Pain	Placebo	29	-3.21	6.52	0.31	(-3.18, 3.80)	0.86
	r-hαGAL	29	-2.90	6.75			
Affective Pain	Placebo	29	-1.45	2.56	0.41	(-0.79, 1.62)	0.49
	r-hαGAL	29	-1.03	1.99			
Total Pain	Placebo	29	-4.66	8.63	0.72	(-3.74, 5.19)	0.75
	r-hαGAL	29	-3.93	8.35			
PPI	Placebo	29	-0.48	1.27	-0.17	(-0.96, 0.62)	0.66
	r-hαGAL	29	-0.66	1.70			
VAS	Placebo	29	-1.42	2.42	0.44	(-0.91, 1.79)	0.51
	r-hαGAL	29	-0.98	2.69			

Table 12 shows that there are no statistically significant differences between treatment groups for any of the pain parameters. This result was also confirmed in the ITT population.

Conclusions

1. There is a statistically significant treatment difference in favor of r-hαGAL with respect to the proportion of patients having a majority score of 0. This difference remained after stratifying on site, age, and race.
2. The scoring was shown to be consistent in that the proportion of 0's was shown to be statistically significantly different between treatment groups (in favor of r-hαGAL) for all three independent pathologists.
3. The sensitivity analysis showed that there is a statistically significant difference in favor of r-hαGAL when using different cutoff percentages to determine the proportion of 0 majority scores.
4. The exploratory analysis on the r-hαGAL group did not reveal any significant covariate effects (age group, plasma GL-3 level, kidney GL-3 level, and urine GL-3 level).
5. The analysis of variance performed on the change scores showed a statistically significant difference in favor of r-hαGAL after adjusting for treatment group and site.
6. The analysis of mean change scores for kidney, heart, skin and the composite score shows a statistically significant difference in favor of r-hαGAL.
7. From the sponsor's analysis of percent change in urine GL-3 and rank sum score, there is a statistically significant difference in favor of r-hαGAL. However, from this reviewer's analysis, this difference disappears after including patients in sites 7 and 8.
8. There was no statistically significant difference between treatment groups with respect to the McGill pain parameters.

References

Woolson, R.F. *Statistical Methods for the Analysis of Biomedical Data.* John-Wiley & Sons. 1987.

Lehmann, E.L. *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day. 1973.