Analysis of Recurrent Events: Nonparametric Methods for Random-Interval Count Data

PETER F. THALL and JOHN M. LACHIN*

Many clinical trials require comparison among treatment groups of the rates over time of a particular recurrent event. The event typically reflects disease morbidity rather than mortality. Some examples are episodes of hypoglycemia in diabetics, angina pectoris in patients with heart disease, and seizures in epileptics. Trial protocol often requires that each patient report only the number of events occurring between clinic visits, so the exact times of the successive events are not available. In practice, patients are early, late, or miss scheduled visits, and follow-up may be censored. Thus each patient's data consist of a sequence of consecutive random intervals and corresponding event counts, some of which may be missing. In this fashion the National Cooperative Gallstone Study (NCGS) recorded the incidence of nausea of patients with gallstone disease treated with chenodiol or placebo. We describe an estimator of the continuous time-dependent rate function for such data. Wei and Lachin (1984) presented a nonparametric method for the analysis of repeated measures with missing observations. We describe estimators of group differences and additional tests of location-shift-type hypotheses based on the Wei–Lachin vector of Wilcoxon-like rank statistics. These methods are applied to compare the recurrence rates of two treatment groups over time, using random-interval count data, by representing each patient's empirical rate function as a vector corresponding to K fixed time intervals. Since this approach allows partially missing values, it uses the available data for patients lost to follow-up. We analyze the incidence of nausea over the first year of treatment of NCGS patients with gallstone disease. This analysis indicates that the recurrence rate of nausea for the placebo group was higher than the chenodiol-treated group for the first six months, but equal thereafter.

KEY WORDS: Repeated measures; Rate estimation; Multivariate nonparametric test; Clinical trial; Gallstone disease.

1. INTRODUCTION

In many clinical trials—especially of chronic diseases—protocol requires that each patient visit the clinical center at specified successive times. At each visit, the patient reports the number of events of some particular type that occurred since the previous visit. These events are recurrent (i.e., nonfatal) and are assumed to occur randomly. This article is concerned with recurrent episodes of nausea experienced by patients with gallstones. Analogous examples from clinical trials in other diseases are episodes of hypoglycemia in patients with diabetes, episodes of angina pectoris in patients with coronary disease, and seizures in epileptics. In all of these examples, the event suffered by the patient reflects disease morbidity rather than simple mortality. This type of data is medically significant both in terms of degradation of the patients' quality of life and the impact on the choice and timing of therapeutic intervention.

In most trials, a wide variety of signs, symptoms, and other events are monitored. It is not practical, however, for the patient to keep a daily diary of the occurrence time of each event. As a result, the exact times of the successive events are not known; instead the event counts corresponding to the intervals between clinic visits are available. Although the clinic visits are typically scheduled in advance, often the actual sequence of visits is random, because patients are often early, late, or entirely miss scheduled visits. A further complication is that each patient's time on study is subject to random censoring. This may be permanent—as in the cases of medical intervention, termination of the treatment protocol, or death—or temporary, if the patient misses a visit and then later reenters the trial protocol.

Thus the observations for each patient consist of a sequence of random intervals between clinic visits and corresponding counts, some of which may be partially censored (in that the count is missing), with a possible terminal time of permanent censoring. In this article we consider the problems of (a) estimating the underlying time-dependent rate function of events from such data and (b) testing the equality of two such rate functions based on data from two independent groups of patients. Although such statistical problems frequently arise in the analysis of clinical trial results, the potential range of application of our methods is much wider.

The special case in which the event occurs at most once for each patient (e.g., death) corresponds to the usual survival-analysis paradigm. In the present context, however, the specific occurrence time is not available. Rather, a random interval known to contain the time of the event or censoring is observed. Thus the grouped survival-time approach of Kalbfleisch and Prentice (1973) and Prentice and Gloeckler (1978) does not apply. Their methods require that the grouping intervals be identical for all patients, which is not the case here.

Similarly, when the events are recurrent it is important to note that we are dealing with observation intervals that are themselves random processes for each patient. This is quite different from the classical point-process analysis, where the intervals are taken to be nonrandom, identical

* Peter F. Thall is Associate Professor and John M. Lachin is Professor, Department of Statistics/Computer and Information Systems, Biostatistics Center, George Washington University, Rockville, MD 20852. Research was partially supported by Grant R01-AM-35952 and Diabetes Control and Complications Trial Contract N01-AM-22206, both from the National Institute of Arthritis, Diabetes, Digestive and Kidney Diseases. The authors thank the editors and a referee for their helpful comments.

© 1988 American Statistical Association
Journal of the American Statistical Association
June 1988, Vol. 83, No. 402, Applications and Case Studies

Copyright © 2001 All Rights Reserved

Table 1. Random-Interval Count Data for the Incidence of Nausea Among the 113 Patients With Floating Gallstones in High-Dose Cheno and Placebo Groups of the NCGS

	Study visit time (τ) and count (N)																				
Patient	τ_1	N_1	τ_2	N_2	τ_3	N_3	τ_4	N_4	τ_5	N_5	τ_6	N_6	τ_7	N_7	τ_8	N_8	τ_9	N_9	τ_{10}	N_{10}	Exit	Week
High-dose cheno group																						
1	4	0	8	0	13	0	26	0	38	0	51	0	69	0	.	.	.	.	.	.		
2	4	0	9	3	13	0	26	0	39	0	51	0	68	0	.	.	.	.	.	.		
3	4	0	8	0	12	0	24	0	38	0	51	0	69	0	.	.	.	.	.	.		
4	4	0	8	0	12	0	26	0	38	0	51	0	69	0	.	.	.	.	.	.		
5	4	0	8	0	13	0	26	0	38	0	52	0	70	0	.	.	.	.	.	.		
6	4	0	8	0	12	0	25	0	39	0	51	0	68	0	.	.	.	.	.	.		
7	4	0	9	0	14	0	26	0	39	0	52	0	70	0	.	.	.	.	.	.		
8	4	0	9	0	14	0	28	0	39	0	53	0	69	0	.	.	.	.	.	.		
9	4	0	9	1	14	0	27	1	38	1	54	4	71	0	.	.	.	.	.	.		
10	4	0	9	0	13	0	17	0	22	0	26	0	38	0	43	0	62	0	.	.		
11	3	0	8	0	13	0	26	0	40	4	53	2	68	0	.	.	.	.	.	.		
12	4	0	8	0	13	1	27	0	39	0	52	0	70	0	.	.	.	.	.	.		
13	4	20	10	2	14	2	17	10	28	0	41	0	54	6	71	0	.	.	.	.		
14	5	1	9	0	13	0	26	0	38	0	52	0	69	0	.	.	.	.	.	.		
15	5	0	9	0	15	0	27	0	39	0	51	0	70	0	.	.	.	.	.	.		
16	4	0	9	0	13	0	26	0	38	0	52	0	69	0	.	.	.	.	.	.		
17	4	0	8	0	12	0	27	0	39	0	51	0	68	0	.	.	.	.	.	.		
18	4	0	8	0	12	0	26	0	37	0	48	0	70	0	.	.	.	.	.	.		
19	4	0	9	0	14	0	28	0	38	0	52	0	71	0	.	.	.	.	.	.		
20	9	0	22	0	31	0	38	0	55	0	68	0	.	.	.	.	.	.	.	.		
21	5	0	10	0	13	0	25	0	50	2	81	0	.	.	.	.	.	.	.	.		
22	4	0	9	0	12	0	25	0	39	0	50	0	59	0	81	0	.	.	.	.		
23	5	0	8	0	13	0	25	0	40	0	60	0	.	.	.	.	.	.	.	.		
24	4	0	9	0	13	0	26	0	38	0	51	0	69	0	.	.	.	.	.	.		
25	4	0	9	0	13	0	26	0	38	0	52	99	84	0	.	.	.	.	.	.		
26	4	0	9	1	13	0	26	0	39	0	53	5	68	2	.	.	.	.	.	.		
27	3	0	8	0	13	1	25	0	40	0	51	0	61	0	.	.	.	.	.	.		
28	4	0	8	0	13	0	24	0	38	0	52	0	68	0	.	.	.	.	.	.		
29	3	0	9	0	12	5	26	0	38	0	50	0	60	0	.	.	.	.	.	.		
30	4	0	10	0	15	1	28	0	41	0	55	3	72	0	.	.	.	.	.	.		
31	3	0	8	0	13	0	26	0	39	0	52	0	93	0	.	.	.	.	.	.		
32	3	1	9	3	13	0	26	0	38	0	52	0	70	0	.	.	.	.	.	.		
33	4	0	10	0	16	0	29	0	41	0	54	6	72	0	.	.	.	.	.	.		
34	3	0	7	0	12	0	25	0	38	0	51	0	71	0	.	.	.	.	.	.		
35	4	0	9	0	13	0	26	0	39	0	51	0	69	0	.	.	.	.	.	.		
36	5	0	9	2	13	0	26	0	39	0	51	0	68	0	.	.	.	.	.	.		
37	6	0	12	6	16	0	28	0	41	0	63	0	.	.	.	.	.	.	.	.		
38	4	0	9	0	13	0	25	0	38	0	51	0	70	0	.	.	.	.	.	.		
39	4	0	8	0	12	0	26	0	40	0	53	0	71	0	.	.	.	.	.	.		
40	4	0	8	0	12	10	26	0	39	0	52	0	71	5	.	.	.	.	.	.		
41	5	0	9	0	14	0	27	0	39	0	52	0	72	3	.	.	.	.	.	.		
42	5	0	9	0	13	0	26	0	36	2	38	0	51	0	67	0	.	.	.	.		
43	4	0	10	0	14	0	26	0	39	0	53	0	71	0	.	.	.	.	.	.		
44	4	0	9	0	16	2	28	4	39	0	51	0	69	0	.	.	.	.	.	.		
45	5	0	10	0	15	0	29	0	40	0	55	0	71	0	.	.	.	.	.	.		
46	4	0	9	0	13	0	26	0	37	0	51	0	70	0	.	.	.	.	.	.		
47	4	0	8	0	13	0	26	0	38	0	51	0	69	0	.	.	.	.	.	.		
48	5	0	10	0	13	0	25	0	39	0	53	0	69	0	.	.	.	.	.	.		
49	3	0	7	0	13	2	25	0	36	5	49	3	68	7	.	.	.	.	.	.		
50	3	0	8	0	13	0	25	8	37	20	53	0	73	0	.	.	.	.	.	.		
51	6	0	9	0	13	0	26	0	40	0	51	0	72	0	.	.	.	.	.	.		
52	5	0	8	0	12	0	25	0	38	0	51	0	69	0	.	.	.	.	.	.		
53	4	0	8	0	13	0	25	0	41	0	53	0	71	1	.	.	.	.	.	.		
54	4	0	8	0	15	0	27	0	40	0	51	10	68	0	.	.	.	.	.	.		
55	4	0	8	1	12	0	27	0	41	0	53	2	56	0	62	0	.	.	.	.		
56	5	0	12	0	16	0	29	0	41	0	52	0	71	0	.	.	.	.	.	.		
57	5	0	11	4	16	0	30	5	44	24	51	40	82	30	.	.	.	.	.	.		
58	.	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Dropout	4
59	.	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Dropout	4
60	3	0	9	0	14	0	26	0	.	.	.	.	.	.	.	.	.	.	.	.	Dropout	38
61	4	0	9	0	13	0	25	0	38	0	.	.	.	.	.	.	.	.	.	.	Dropout	47
62	4	0	8	0	14	0	18	0	20	0	.	.	.	.	.	.	.	.	.	.	Hepatotoxicity	20
63	4	0	8	0	13	0	17	0	23	0	27	0	32	0	.	.	.	.	.	.	Hepatotoxicity	37
64	3	0	10	0	26	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Withdrawn	44
65	8	5	19	0	28	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Withdrawn	50

Copyright © 2001 All Rights Reserved

Table 1 (continued)

Patient	Study visit time (τ) and count (N)																			Exit	Week
	τ_1	N_1	τ_2	N_2	τ_3	N_3	τ_4	N_4	τ_5	N_5	τ_6	N_6	τ_7	N_7	τ_8	N_8	τ_9	N_9	τ_{10}	N_{10}	
Placebo group																					
66	4	0	8	0	12	0	25	0	38	0	52	0	68	0	.	.	.	.	.	.	
67	4	0	8	0	13	0	27	0	40	0	44	0	53	0	69	0	.	.	.	.	
68	4	0	11	0	14	0	26	0	39	0	52	0	69	0	.	.	.	.	.	.	
69	4	0	9	0	12	0	25	0	40	0	52	0	70	0	.	.	.	.	.	.	
70	4	0	8	0	14	0	27	0	40	0	52	0	69	0	.	.	.	.	.	.	
71	5	1	9	0	13	0	26	1	40	0	53	0	69	0	.	.	.	.	.	.	
72	4	0	8	0	13	0	24	0	37	0	50	0	67	0	.	.	.	.	.	.	
73	4	1	9	0	14	4	28	3	41	1	54	1	71	2	.	.	.	.	.	.	
74	3	0	9	0	13	0	25	0	38	0	50	0	67	1	.	.	.	.	.	.	
75	5	0	9	0	13	0	27	0	38	0	51	0	69	0	.	.	.	.	.	.	
76	4	0	8	0	13	0	27	0	38	0	51	0	67	0	.	.	.	.	.	.	
77	4	3	9	0	14	0	25	0	39	0	51	0	69	1	.	.	.	.	.	.	
78	3	8	8	0	11	1	17	4	24	0	38	2	42	0	46	0	51	20	61	1	
79	4	0	9	0	13	0	25	0	39	0	51	0	68	0	.	.	.	.	.	.	
80	4	0	8	0	13	0	24	0	38	0	51	0	69	0	.	.	.	.	.	.	
81	4	0	9	0	13	0	26	0	40	0	51	0	68	0	.	.	.	.	.	.	
82	4	0	9	0	14	0	28	0	40	0	51	0	71	0	.	.	.	.	.	.	
83	5	0	8	0	16	0	28	0	36	0	55	0	81	0	.	.	.	.	.	.	
84	5	0	7	0	12	0	25	2	38	0	53	1	72	0	.	.	.	.	.	.	
85	5	0	10	0	15	0	29	0	41	0	55	0	69	0	.	.	.	.	.	.	
86	4	0	9	0	13	0	25	0	35	0	56	0	74	0	.	.	.	.	.	.	
87	4	0	9	0	13	0	28	0	39	0	59	0	70	0	.	.	.	.	.	.	
88	4	0	9	3	12	0	24	0	37	0	51	0	68	0	.	.	.	.	.	.	
89	4	0	8	60	13	0	24	0	40	1	55	0	74	4	.	.	.	.	.	.	
90	3	0	8	1	14	0	26	0	38	0	53	0	70	0	.	.	.	.	.	.	
91	5	0	9	0	13	0	27	0	40	0	54	0	73	0	.	.	.	.	.	.	
92	3	0	8	0	11	0	25	0	37	0	51	0	68	0	.	.	.	.	.	.	
93	3	1	7	4	11	0	24	0	38	0	54	9	.	.	.	.	.	.	.	.	
94	3	5	8	0	13	0	25	0	38	0	52	0	68	0	.	.	.	.	.	.	
95	4	0	9	0	13	0	26	3	39	0	52	0	70	0	.	.	.	.	.	.	
96	4	0	9	0	14	0	26	0	39	0	52	0	68	0	.	.	.	.	.	.	
97	4	6	9	0	18	1	28	0	39	0	54	0	74	10	.	.	.	.	.	.	
98	5	0	9	0	15	0	27	0	39	0	53	0	69	0	.	.	.	.	.	.	
99	4	0	9	0	13	2	25	0	38	0	50	0	68	0	.	.	.	.	.	.	
100	3	3	7	0	12	0	25	6	38	0	52	0	69	0	.	.	.	.	.	.	
101	4	0	7	0	12	0	25	0	38	1	53	0	69	0	.	.	.	.	.	.	
102	4	0	8	0	13	0	26	0	39	0	51	0	70	0	.	.	.	.	.	.	
103	4	0	8	0	13	0	26	0	40	0	52	0	78	10	.	.	.	.	.	.	
104	4	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Cholecystectomy	6
105	4	0	8	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Cholecystectomy	8
106	5	0	9	0	13	0	17	0	21	0	28	1	39	1	.	.	.	.	.	.	Cholecystectomy	50
107	3	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Dropout	3
108	6	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Dropout	13
109	3	25	8	30	14	20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Dropout	15
110	4	0	9	0	13	12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Dropout	16
111	4	0	9	0	13	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Dropout	23
112	5	0	9	0	14	0	26	0	.	.	.	.	.	.	.	.	.	.	.	.	Dropout	33
113	4	0	9	0	14	0	25	0	.	.	.	.	.	.	.	.	.	.	.	.	Dropout	33

NOTE: The table shows all visit times (in weeks) within the first 58 weeks of study and the corresponding counts. The reason and week of exit from the study are shown if such occurred during the first 58 weeks of study.

for all observations, and usually of equal length. Indeed, grouping events in this manner and dealing with the resulting counts rather than event times is a common technique for analyzing point-process data (e.g., see Cox 1955; Cox and Lewis 1966; Lewis 1972).

The case in which the actual successive event times are available has been treated by many authors using a wide variety of parametric and nonparametric models. In particular, there has been a rapid development in recent years of statistical methods based on the martingale theory of counting processes. Following the seminal paper of Aalen (1978), Gill (1980) derived methods for survival analysis and Andersen, Borgan, Gill, and Kieding (1982) derived methods for K-sample tests. Andersen and Gill (1982) gave a rigorous treatment in the context of Cox's proportional-hazards regression model, and Gill (1984) presented an informal account of the connection between martingales and survival models. Note that Andersen and Gill (1982) also derived a two-sample test for the equality of recurrent event rates. Finally, Karr (1986) gave a comprehensive account of modern inferential methods for point processes, including those already cited.

None of these methods are applicable to the present problem, because they require that the time of each successive event be known. Thall (1988) proposed a parametric regression approach to interval count data when

Copyright © 2001 All Rights Reserved

baseline covariates are available: In this article we take an alternative approach, considering the problem in the context of repeated measurements. Koch, Amara, Stokes, and Gillings (1980) produced a review and bibliography of the statistical analysis of split-plot and repeated-measures data. Multivariate nonparametric tests were discussed by Chatterjee and Sen (1964, 1970), Puri and Sen (1966), and Koch (1969, sec. 3), among others. Nonparametric methods accommodating partially missing or incomplete multivariate data were given by Koziol, Maxwell, Fukushima, Colmerauer, and Pilch (1981), Wei and Lachin (1984), and Wei and Johnson (1985), among others.

In this article we compare the recurrence rates of two treatment groups over time based on random-interval count data. We use the methods of Wei and Lachin (1984); their nonparametric procedures, applied in the present context, may be regarded as tests of treatment effect, time effect, and treatment-time interaction, where data may be partially missing.

In Section 2 we describe the recurrent biliary-symptom data from the National Cooperative Gallstone Study (NCGS), which motivated this work. A formal description of the data structure and notation is given in Section 3. Section 4 presents statistical methods for rate estimation and application of the Wei and Lachin (1984) multivariate rank test, and related normal-theory linear model methods. An analysis of the incidence of nausea in the NCGS using these techniques is presented in Section 5.

2. BILIARY SYMPTOMS IN THE NCGS

The NCGS was a 10-year, multicenter, double-masked, placebo-controlled clinical trial of the use of the natural bile acid chenodeoxycholic acid (chenodiol) for the dissolution of cholesterol gallstones (Schoenfield et al. 1981). A total of 916 patients were treated for up to two years each with a high dose (750 mg per day), low dose (375 mg per day), or placebo, randomly assigned. To avoid side effects, patients were initially administered $\frac{1}{3}$, then $\frac{2}{3}$, then all of the nominal dose during the first, second, and third months of treatment, respectively. The principal objectives of the trial were (a) to assess the effectiveness of chenodiol for the dissolution of gallstones, thus possibly avoiding the need for surgical removal of the gallbladder; (b) to assess its potential toxicity, especially on the liver; and (c) to assess the impact of treatment on the incidence of biliary (digestive) symptoms commonly associated with gallstone disease. These symptoms range from milder episodes of nausea/vomiting, dyspepsia, and diarrhea to more severe episodes of biliary colic (severe pain) and cholecystitis (biliary obstruction). During the study patients were asked to report the total number of each type of symptom that had occurred during the interval preceding each successive clinic visit.

In this article we analyze the incidence of nausea during the first year of follow-up in the NCGS. The more severe events occurred infrequently during the study (Thistle et al. 1984). Among the less severe symptoms, nausea is very commonly associated with gallstone disease. *Nausea* is an unpleasant sensation vaguely referred to the epigastrium and abdomen, often culminating in vomiting. In the NCGS, it was hypothesized that chenodiol might affect the incidence of biliary symptoms, either through the restoration of metabolic homeostasis or reduction of the aggravating effect of gallstones on the gallbladder itself. Thus it is important to determine whether the incidence of nausea differed significantly between the chenodiol and placebo groups over the first year of the study. Only the first year of observation is considered, because it was hypothesized that any treatment effect should be observed shortly after patients achieved maximal dose (usually by three months), and the effect might later begin to dissipate.

During the first year of follow-up, patients were scheduled to return for clinic visits at approximately 4.3, 8.7, 13, 26, 39, and 52 weeks (equivalently, at 1, 2, 3, 6, 9, and 12 months). Table 1 presents the successive visit-times in study weeks and the associated counts of episodes of nausea for the subset of 113 patients with floating gallstones in the high-dose chenodiol and placebo groups. Data from the low-dose group are not presented, because this dose proved to be ineffective. Also, only the data from patients with floating stones are considered, because such stones are almost pure cholesterol (as opposed to pigment) and thus are maximally susceptible to dissolution with a bile acid such as chenodiol.

From Table 1 note that during the first year of follow-up some patients left the study because of (a) their own choice (dropout), (b) liver toxicity (hepatotoxicity), (c) other clinical reasons (withdrawal), or (d) surgical removal of the gallbladder (cholecystectomy). Some patients had no visits, some the scheduled 6 visits, and some as many as 10 visits. First visit times ranged from 3 to 9 weeks, with corresponding event counts ranging from 0 to 25.

From these random-interval count data, we wish to estimate the underlying rate function for each treatment group and test the equality of the rates over time for the two groups. To do this we require a more precise description of the available information.

3. RANDOM-INTERVAL COUNT DATA

First, consider the data of a single patient. In the following we define time to be *study time*—that is, the elapsed time since the patient's randomization into the trial. Let $0 = \tau_0 < \tau_1 < \cdots < \tau_m$ be the sequence of the patient's distinct visit times, denoting the jth interval between visits by $E_j = (\tau_{j-1}, \tau_j]$. Let N_j be the number of events reported (at τ_j) to have occurred during E_j.

To incorporate both transient and right censoring, we define $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m)$ as follows: Suppose a patient leaves the trial protocol at a time s subsequent to visit time τ_{j-1} and later reenters protocol, that is, begins counting events, at time t. If no count is available for the interval $(\tau_{j-1}, s]$ or if s is not known, then s is irrelevant, $\tau_j = t$, and E_j is a *censored interval.* If a count is available for the interval $(\tau_{j-1}, s]$, then $\tau_j = s$, $\tau_{j+1} = t$, and $E_{j+1} = (s, t]$ is censored. In any case, the last interval E_m is not cen-

Copyright © 2001 All Rights Reserved

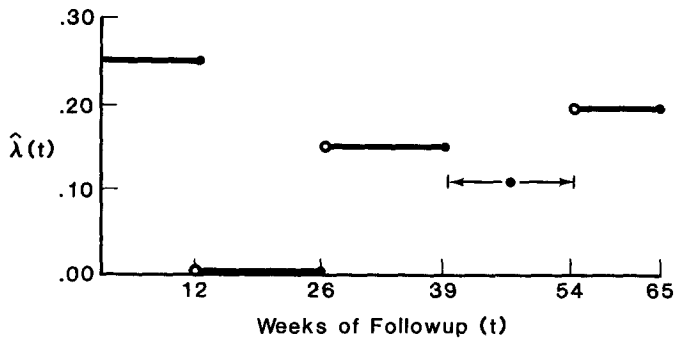

Figure 1. Observed Rate Function Corresponding to the Sequence of Visit Times τ = (12, 26, 39, 54, 65) and Corresponding Counts $\mathbf{N}$ = (3, 0, 2, ·, 2).

sored, and we write $N_j = \cdot$ when E_j is censored. For convenience write $\mathbf{N} = (N_1, \ldots, N_m)$ and $\mathbf{E} = (E_1, \ldots, E_m)$.

This data structure may be expressed equivalently by the observed rate function

$$\hat{\lambda}(t) = \sum_{j=1}^{m} \frac{N_j}{\tau_j - \tau_{j-1}} I(t \in E_j), \qquad (3.1)$$

when $\hat{\lambda}(t) = \cdot$ if t is in a censored interval. For example, the graph corresponding to τ = (12, 26, 39, 54, 65) and $\mathbf{N}$ = (3, 0, 2, ·, 2) is given in Figure 1. Since the raw data τ and $\mathbf{N}$ can be completely reconstructed from the graph of (3.1), $\hat{\lambda}$ contains all of the available information for the patient. In this sense, it is reasonable to use $\hat{\lambda}$ as the basis for analyzing these data.

Note that although N_j becomes available at τ_j, it corresponds to the preceding interval E_j. In fact, the observed counts $\mathbf{N}$ arise from an underlying event process $B(t)$ = number of events occurring up to study time t. The essential feature of such data is that $B(t)$ is not observed for all t. Rather, $B(t)$ is observed only at the successive visit times through the counts $N_j = B(\tau_j) - B(\tau_{j-1})$ $(j = 1, \ldots, m)$. Moreover, the sequence of visit times $\tau_1, \ldots, \tau_m$ is itself a random process.

4. STATISTICAL METHODS

4.1 Rate Estimation

Consider a sample of size n, with the subscript i denoting patient. Thus τ_{ij} is the jth clinic visit time of the ith patient, and so on. We assume the underlying event processes to have some common rate function

$$\lambda(t) = \lim_{\delta \downarrow 0} \delta^{-1} \Pr(B_i(t + \delta) - B_i(t) = 1), \qquad (4.1)$$

under the usual assumption that no patient can have more than one event at any instant in time. An obvious estimator of λ, based on a single patient's data, is $\hat{\lambda}_i$ [given by (3.1)]. For a sample estimator, we define

$$\bar{\lambda}(t) = \sum_{i=1}^{n} \hat{\lambda}_i(t) \Big/ \sum_{i=1}^{n} \sum_{j=1}^{m_i} I(t \in E_{ij}), \qquad (4.2)$$

with the provisions that (a) any i for which $\hat{\lambda}_i(t) = \cdot$ is omitted from the sums in both numerator and denominator, and (b) if all $\hat{\lambda}_i(t) = \cdot$, then $\bar{\lambda}(t) = \cdot$. In particular, $\bar{\lambda}$ is defined only for values of t ranging over the actual study time of the trial. Although other estimators may certainly be used, this estimator is easy to compute and has the unambiguous model-free interpretation that it is essentially the sample mean of the patients' individual observed rates. A $100(1 - p)\%$ confidence band for λ may be computed by standard methods.

4.2 Two-Sample Tests

To test the hypothesis H_λ: $\lambda^{(1)} = \lambda^{(2)}$ that the event rate functions corresponding to two different treatment regimens are equal, we apply the procedure of Wei and Lachin (1984) (hereafter referred to as WL). Their test is based on two samples $\mathbf{X}_1^{(1)}, \ldots, \mathbf{X}_{n_1}^{(1)}$ and $\mathbf{X}_1^{(2)}, \ldots, \mathbf{X}_{n_2}^{(2)}$ of K-variate nonnegative-valued random vectors with respective distributions G_1 and G_2 on $[0, \infty)^K$. If the kth entry of $\mathbf{X}_i$ is missing, we write $X_{ik} = \cdot$. Using a device similar to that used with censored survival data, we define the corresponding censoring variable $U_{ik} = \infty$ if X_{ik} is observed, -1 if $X_{ik} = \cdot$, with $\tilde{X}_{ik} = \min(X_{ik}, U_{ik})$. The corresponding indicator $\Delta_{ik} = 1$ if $\tilde{X}_{ik} = X_{ik}$, 0 if not. The entries of each $\mathbf{X}_i$ are assumed to be missing at random; that is, we assume the $\mathbf{U}_i$ are mutually independent and independent of the data vectors $\mathbf{X}_i$. An important feature of the WL procedure is that the patterns of missing data need not be the same in the two groups. That is, the K-dimensional distributions of the censoring vectors may be different.

We now transform each $\hat{\lambda}_i$ into a K-vector $\mathbf{X}_i$ and then apply the WL procedure to the statistics so obtained. We first partition the entire study-time period into K fixed, consecutive intervals $A_1, \ldots, A_K$. For each interval $k = 1, \ldots, K$, the rate $\hat{\lambda}_i$ of patient i is a step function on A_k. The corresponding entry of $\mathbf{X}_i$ is defined to be the weighted average

$$X_{ik} = \sum_{j=1}^{m_i} \frac{N_{ij}}{\tau_{ij} - \tau_{i,j-1}} \frac{\text{length}(E_{ij} \cap A_k)}{\text{length}(A_k)}, \qquad (4.3)$$

provided that $\hat{\lambda}_i(t)$ is not missing for any $t \in A_k$. If $\hat{\lambda}_i$ is missing on any subinterval of A_k, we define $X_{ik} = \cdot$. Given that the $(\mathbf{N}_i, \mathbf{E}_i)$ are iid within each sample, it follows that the $\mathbf{X}_i$ are iid. A test of H_λ may now be carried out by applying the WL test of $H_G: G_{1k} = G_{2k}$ $(1 \le k \le K)$, based on the constructed K-vectors. Using Wilcoxon scores, the kth entry of the two-sample K-variate statistic $\mathbf{T} = (T_1, \ldots, T_K)$ can be written as

$$T_k = n^{-3/2} \times \sum_{i=1}^{n_1} \sum_{i'=1}^{n_2} [\Delta_{ik}^{(1)} I(\tilde{X}_{i'k}^{(2)} \ge \tilde{X}_{ik}^{(1)}) - \Delta_{i'k}^{(2)} I(\tilde{X}_{ik}^{(1)} \ge \tilde{X}_{i'k}^{(2)})], \qquad (4.4)$$

where $n = n_1 + n_2$. Under the hypothesis $G_1 = G_2$, $\mathbf{T}$ is asymptotically normal, with mean $\mathbf{0}$ and covariance matrix Σ, which can be consistently estimated by $\hat{\Sigma} = [\hat{\sigma}_{kl}]$ (say). This result and a formula for $\hat{\Sigma}$ are given in theorem 1 of WL.

Copyright © 2001 All Rights Reserved

To apply the WL test as described, we must assume that (a) any censoring mechanisms act independently of the underlying event process and (b) the clinic visit processes are identically distributed in the two samples. Assumption (a) is required to ensure that U_{ik} is independent of X_{ik} for each i and $k = 1, \ldots, K$; that is, unobserved values are missing at random. Also, G_1 and G_2 must be truly K-dimensional; that is, their common support may not lie in a $K - 1$ or lower dimensional space. In the unlikely event that all $X_{ik} = \cdot$ for some k, the interval A_k would simply be excluded from the analysis. This does not occur in practice, however, and transient censoring rarely occurs. Our primary concern here is loss to follow-up, which leads to right-truncation of the vector $\mathbf{X}_i$.

This provides a basis for testing the equality of G_1 and G_2. Subject to the representation of each observed $\hat{\lambda}_i$ by the vector $\mathbf{X}_i$, these serve as tests of the equality of the two underlying rates on the K intervals, taken either separately or together. For each of the K fixed intervals, an approximate 1 df chi-squared test statistic for H_{0k}: $G_{1k} = G_{2k}$ versus H_{ak}: $G_{1k} \neq G_{2k}$ is given by $T_k^2/\hat{\sigma}_{kk}$. If K separate tests are performed, an adjustment of individual significance levels is necessary to control the overall Type I error rate. An alternative approach here is to conduct a single test of the joint hypothesis H_G. Under the global assumption that $G_1 = G_2$, the quadratic form $\mathbf{T}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{T}'$ is asymptotically chi-square on K df. This provides a test of H_G against the alternative H_G': $G_{1k} \neq G_{2k}$ for some k, under the assumption that the covariance structures of the two populations are identical. As is the case in the general multivariate location-scale setting, however, the test has no power against the global alternative $G_1 \neq G_2$ if the marginals are the same but the covariance matrices are different.

These are tests of rather general hypotheses. If we consider a more restrictive family of location-shift–type hypotheses, however, a more powerful family of tests is available. Let n_{1k} and n_{2k} denote the respective numbers of nonmissing values in the two samples corresponding to the fixed interval A_k. Let $X_k^{(1)}$ and $X_k^{(2)}$ be independent variables following the kth marginals G_{1k} and G_{2k} of G_1 and G_2, respectively. The parameter $\theta_k = \Pr[X_k^{(2)} \geq X_k^{(1)}] - \Pr[X_k^{(1)} \geq X_k^{(2)}]$ is an index of the degree of overlap of the marginals G_{1k} and G_{2k}, so $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$ may be regarded as a K-dimensional similarity index between G_1 and G_2. Note that the hypothesis H_0: $\boldsymbol{\theta} = \mathbf{0}$ holds under H_G.

Since the statistic (4.4) has the interpretation $n^{3/2}T_k/n_{1k}n_{2k} = \hat{\theta}_k$, when Wilcoxon scores are used specific hypotheses in terms of $\boldsymbol{\theta}$ may be tested based on $\hat{\boldsymbol{\theta}} = D\mathbf{T}'$, where $D = \operatorname{diag}(n^{3/2}/n_{1k}n_{2k})$. For any K-vector $\mathbf{w}$ of weights summing to unity, a test of H_0:$\boldsymbol{\theta} = \mathbf{0}$ versus H_1:$\theta_1 = \cdots = \theta_K = \theta_* \neq 0$ (i.e., a common nonzero value) may be carried out using the statistic $(\mathbf{w}\hat{\boldsymbol{\theta}}')^2/(\mathbf{w}\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}})\mathbf{w}')$, which is asymptotically chi-square on 1 df, with $V(\hat{\boldsymbol{\theta}}) = D\boldsymbol{\Sigma}D = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. A locally most powerful test is obtained by using the optimal weights $\mathbf{w}_* = \mathbf{e}\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}})^{-1}(\mathbf{e}\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}})^{-1}\mathbf{e}')$, where $\mathbf{e} = (1, \ldots, 1)$ of dimension K, because $\hat{\theta}_* = \mathbf{w}_*\hat{\boldsymbol{\theta}}'$ is the minimum variance unbiased estimator of θ_* under H_1 (e.g., see Rao 1972, p. 60). This procedure is asymptotically equivalent to the optimally weighted linear combination of U-statistics derived by Wei and Johnson (1985), since both tests are fully efficient.

A test for heterogeneous overlap between groups takes H_1 as the null hypothesis versus the alternative H_2:$\theta_k \neq \theta_{k'}$ for some $k \neq k'$. A test statistic for H_1 versus H_2 is given by the simple quadratic form

$$C\hat{\boldsymbol{\theta}}'(C\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}})C')^{-1}\hat{\boldsymbol{\theta}}C', \tag{4.5}$$

where

$$C = \begin{bmatrix} 1 & -1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & -1 & & & & \cdot \\ \cdot & & & \cdot & & & \cdot \\ \cdot & & & & \cdot & & \cdot \\ \cdot & & & & & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 & -1 \end{bmatrix}$$

of dimension $(K - 1) \times K$, since H_1 may be equivalently expressed as $C\boldsymbol{\theta}' = 0$. By the usual theory of quadratic forms for normal random vectors, the statistic (4.5) is approximately chi-square on $K - 1$ df under H_1. Note that any $q \leq K$ linearly independent linear combinations of $\boldsymbol{\theta}$ may be treated analogously.

It is important to note that if H_2 is true, then the test of H_0 versus H_1 is inappropriate, since it is directed at a local alternative. Simply put, under H_2 there is no single parameter θ_* on which to base the locally most powerful test. In general, the inferential process should proceed based on prior knowledge of the phenomenon under study. In many applications, tests must be performed as preliminary procedures to determine a model for subsequent tests or parameter estimation. Bancroft (1972) gave a general discussion of preliminary tests of significance. In the present context, if it is reasonable to assume a common overlap parameter θ_*, then it is appropriate simply to test H_0 versus H_1. Alternatively, one might begin with a conditional specification in the sense of Bancroft and Han (1977), where the separate estimates of $\theta_1, \ldots, \theta_K$ are used in a preliminary test of H_1 versus H_2. If H_1 is accepted, the estimate of the common θ_* is used to test H_0 versus H_1. Here the sizes of the individual tests should be adjusted to control Type I error. For example, the Bonferroni inequality yields an overall level of $1 - (1 - \alpha_1)(1 - \alpha_2)$ if the two tests are performed at levels α_1 and α_2.

The analogous approach based on $T\boldsymbol{\Sigma}^{-1}T'$ is to conduct a preliminary test of H_G versus H_G'. The K separate tests of H_{0k} versus H_{ak} would then be performed only if H_G were rejected, with significance levels (or interpretation of p-values) adjusted accordingly. This is directly analogous to the problem of controlling Type I error rate when univariate t-tests are performed after obtaining a significant Hotelling T^2 statistic.

5. STATISTICAL ANALYSIS

We now present an analysis of the incidence of nausea over the first 12 months of study for those 113 NCGS patients with floating gallstones in the high-dose chenodiol ($n_1 = 65$) and placebo ($n_2 = 48$) groups. The reporting

Copyright © 2001 All Rights Reserved

times in study weeks and associated event counts are presented in Table 1. Of the 113 patients, 8 high-dose and 10 placebo patients exited from the trial during the first year. Apart from dropout, the reasons for exit differed between the two groups. In the high-dose group, eight patients had dissolution of gallstones observed at month 9 and confirmed at month 12. Nevertheless, all of these patients continued follow-up for the entire 12 months. Thus there is no evidence within either group that the missing observations are in any way associated with the incidence of nausea. Therefore, it is reasonable to assume that observations are missing at random (as required by the WL test).

During the first year, patients were scheduled to return for follow-up visits at 1, 2, 3, 6, 9, and 12 months. The distributions of differences between the scheduled and actual visit times were not significantly different between the two groups. The actual visit times were symmetrically distributed about the scheduled times, with 32.4% of all visits occurring at least one week prior to or after the scheduled visit week.

Figure 2 presents the empirical estimates (4.2) of the rate functions for the two treatment groups. In the placebo group, the mean event rate is initially rather high, near .5 events per week during the first eight weeks, but drops sharply to .028 by week 19, with a temporary rise to .124 late in the year. In contrast, the high-dose chenodiol group showed a low mean rate ($\leq$.120 events per week) during the first 38 weeks of treatment, with a temporary rise to the range .282–.284 events per week during weeks 45–51. Aside from the initially higher rate in the placebo group, it is notable that both groups show a temporary increase in mean incidence during the same time period late in the year.

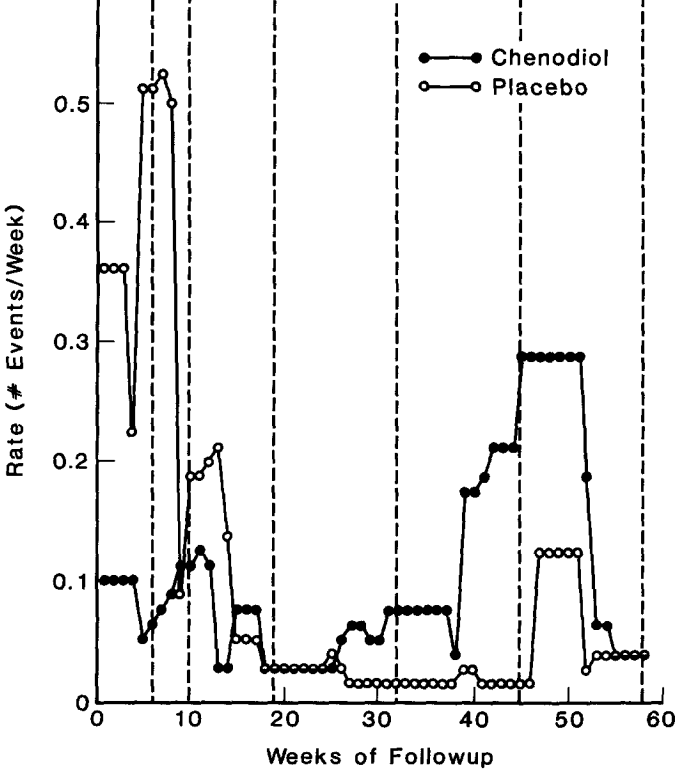

Figure 2. Mean Observed Rate Functions for Nausea Experienced by 65 High-Dose Chenodiol-Treated and 48 Placebo-Treated Patients With Floating Stones During the First Year of the NCGS.

One of the salient characteristics of these data is that most of the counts are 0 (Table 1), reflecting low underlying event rates in both groups. Of the 63 high-dose patients with at least one clinic visit, 36 (57.1%) had *no* reported episodes throughout their periods of study. Similarly, 25 of 48 placebo patients (52.1%) had no reported episodes. It is also notable that high-dose patient 25 reported $N_6 = 99$ events for the (39–52)-week interval, giving an individual observed rate of $\hat{\lambda}_{25}(t) = 7.071$ for that period. This patient contributed 27%–57% of the sample empirical rate during this time period, so the late temporary rise in $\bar{\lambda}$ for the high-dose group may in part be attributed to this single patient.

For application of the WL test, six standard intervals were constructed using the midpoints between the prescheduled visit times (i.e., intervals ending at 6, 10, 19, 32, 45, and 58 weeks; these boundaries also appear in Fig. 2). The individual mean rate estimators $\mathbf{X}_i$ defined by (4.3) for each patient are presented in Table 2. Recall that $X_{ik} = \cdot$ whenever the patient did not have an observed rate that spanned the entire kth interval A_k. The WL procedure was applied using Wilcoxon scores. The statistics $\mathbf{T} = (T_1, \ldots, T_6)$ [defined by (4.4)] are presented in Table 3, along with the estimated covariance matrix $\hat{\Sigma}$. This table also contains the estimated overlap parameters and the 1 df chi-square statistics for each of the six intervals. Note that $\hat{\Pr}(X^{(2)} \geq X^{(1)})$ and $\hat{\Pr}(X^{(1)} \geq X^{(2)})$ do not sum to a value anywhere near 1, since a substantial number of X_{ik} values are tied at 0.

The preliminary multivariate test of H_G versus H_G' yields an observed $\chi^2 = 12.72$ on 6 df, $p = .048$. The empirical mean rate plots show a higher level in the placebo group over the first eight weeks and a higher level in the high-dose chenodiol group over weeks 39–52. But for the $K = 6$ intervals considered separately, none of the 1 df chi-square tests of H_{0k}: $G_{1k} = G_{2k}$ were significant at the adjusted level $p = .0085 = 1 - (.95)^{1/6}$. Thus the multivariate test indicates a significant difference between groups in the joint marginal distributions, even though the univariate tests fail to detect a difference. This is due in part to moderate correlations (ranging from .18 to .63) among the six statistics $T_1, \ldots, T_6$.

Alternatively, it was reasonable to assume a priori that a location-shift family might apply in this situation. The preliminary test of H_1 versus H_2, that is, heterogeneous overlaps in the fixed intervals, yields an observed $\chi^2 = 12.40$ on 5 df, $p = .030$. The estimated overlap parameters $\hat{\theta}_1, \ldots, \hat{\theta}_6$ differ not only quantitatively but qualitatively, as indicated by a sign change over time.

Omitting the one patient (No. 25) who reported 99 events for the interval 39–52 weeks did not substantively alter these results, because the overlap parameter estimators and tests are based on rank statistics and are thus robust against extreme values. Moreover, elimination of this pa-

Copyright © 2001 All Rights Reserved

Table 2. Mean Rate X_k Over Each Fixed Interval for the 113 Patients With Floating Gallstones in the NCGS

Patient	X_1	X_2	X_3	X_4	X_5	X_6	Patient	X_1	X_2	X_3	X_4	X_5	X_6	Exit	Week
	High-dose cheno group						58	.	.	.	.	.	.	Dropout	4
1	.00	.00	.00	.00	.00	.00	59	.	.	.	.	.	.	Dropout	4
2	.20	.45	.00	.00	.00	.00	60	.00	.00	.00	.	.	.	Dropout	38
3	.00	.00	.00	.00	.00	.00	61	.00	.00	.00	.00	.	.	Dropout	47
4	.00	.00	.00	.00	.00	.00	62	.00	.00	.00	.	.	.	Hepatotoxicity	20
5	.00	.00	.00	.00	.00	.00	63	.00	.00	.00	.00	.	.	Hepatotoxicity	37
6	.00	.00	.00	.00	.00	.00	64	.00	.00	.00	.	.	.	Withdrawn	44
7	.00	.00	.00	.00	.00	.00	65	.63	.31	.00	.	.	.	Withdrawn	50
8	.00	.00	.00	.00	.00	.00									
9	.07	.15	.04	.08	.18	.17								*Placebo group*	
10	.00	.00	.00	.00	.00	.00	66	.00	.00	.00	.00	.00	.00		
11	.00	.00	.00	.13	.23	.09	67	.00	.00	.00	.00	.00	.00		
12	.00	.10	.07	.00	.00	.00	68	.00	.00	.00	.00	.00	.00		
13	3.44	.33	1.33	.00	.14	.32	69	.00	.00	.00	.00	.00	.00		
14	.17	.00	.00	.00	.00	.00	70	.00	.00	.00	.00	.00	.00		
15	.00	.00	.00	.00	.00	.00	71	.17	.00	.05	.04	.00	.00		
16	.00	.00	.00	.00	.00	.00	72	.00	.00	.00	.00	.00	.00		
17	.00	.00	.00	.00	.00	.00	73	.17	.20	.47	.17	.08	.09		
18	.00	.00	.00	.00	.00	.00	74	.00	.00	.00	.00	.00	.04		
19	.00	.00	.00	.00	.00	.00	75	.00	.00	.00	.00	.00	.00		
20	.00	.00	.00	.00	.00	.00	76	.00	.00	.00	.00	.00	.00		
21	.00	.00	.00	.04	.08	.03	77	.50	.00	.00	.00	.00	.03		
22	.00	.00	.00	.00	.00	.00	78	1.33	.17	.48	.09	.07	1.59		
23	.00	.00	.00	.00	.00	.00	79	.00	.00	.00	.00	.00	.00		
24	.00	.00	.00	.00	.00	.00	80	.00	.00	.00	.00	.00	.00		
25	.00	.00	.00	.00	3.81	3.81	81	.00	.00	.00	.00	.00	.00		
26	.07	.15	.00	.00	.16	.27	82	.00	.00	.00	.00	.00	.00		
27	.00	.10	.07	.00	.00	.00	83	.00	.00	.00	.00	.00	.00		
28	.00	.00	.00	.00	.00	.00	84	.00	.00	.12	.07	.04	.04		
29	.00	.42	.37	.00	.00	.00	85	.00	.00	.00	.00	.00	.00		
30	.00	.00	.11	.00	.07	.16	86	.00	.00	.00	.00	.00	.00		
31	.00	.00	.00	.00	.00	.00	87	.00	.00	.00	.00	.00	.00		
32	.42	.38	.00	.00	.00	.00	88	.20	.45	.00	.00	.00	.00		
33	.00	.00	.00	.00	.14	.32	89	5.00	7.50	.00	.04	.04	.05		
34	.00	.00	.00	.00	.00	.00	90	.10	.10	.00	.00	.00	.00		
35	.00	.00	.00	.00	.00	.00	91	.00	.00	.00	.00	.00	.00		
36	.08	.38	.00	.00	.00	.00	92	.00	.00	.00	.00	.00	.00		
37	.00	1.00	.22	.00	.00	.00	93	.67	.25	.00	.00	.30	.		
38	.00	.00	.00	.00	.00	.00	94	.83	.00	.00	.00	.00	.00		
39	.00	.00	.00	.00	.00	.00	95	.00	.00	.15	.12	.00	.00		
40	.00	1.25	.56	.00	.00	.12	96	.00	.00	.00	.00	.00	.00		
41	.00	.00	.00	.00	.00	.07	97	1.00	.03	.10	.00	.00	.15		
42	.00	.00	.00	.09	.06	.00	98	.00	.00	.00	.00	.00	.00		
43	.00	.00	.00	.00	.00	.00	99	.00	.13	.17	.00	.00	.00		
44	.00	.07	.30	.23	.00	.00	100	.50	.00	.36	.21	.00	.00		
45	.00	.00	.00	.00	.00	.00	101	.00	.00	.00	.04	.04	.00		
46	.00	.00	.00	.00	.00	.00	102	.00	.00	.00	.00	.00	.00		
47	.00	.00	.00	.00	.00	.00	103	.00	.00	.00	.00	.00	.18		
48	.00	.00	.00	.00	.00	.00	104	.	.	.	.	.	.	Cholecystectomy	6
49	.00	.25	.11	.24	.30	.33	105	.17	.	.	.	.	.	Cholecystectomy	8
50	.00	.00	.44	1.21	.64	.00	106	.00	.00	.00	.10	.	.	Cholecystectomy	50
51	.00	.00	.00	.00	.00	.00	107	.	.	.	.	.	.	Dropout	3
52	.00	.00	.00	.00	.00	.00	108	.00	.	.	.	.	.	Dropout	13
53	.00	.00	.00	.00	.00	.02	109	7.17	4.67	.	.	.	.	Dropout	15
54	.00	.00	.00	.00	.35	.42	110	.00	.75	.	.	.	.	Dropout	16
55	.08	.13	.00	.00	.05	.10	111	.00	.06	.	.	.	.	Dropout	23
56	.00	.00	.00	.00	.00	.00	112	.00	.00	.00	.	.	.	Dropout	33
57	.11	.67	.19	.57	2.02	3.16	113	.00	.00	.00	.	.	.	Dropout	33

tient would be inappropriate, since 6 of all 1,044 patients in the NCGS had a maximum reported count ≥90. Also, it is interesting to note that the locally most powerful 1 df test of H_0 versus H_1 yields an observed $\chi^2 = .33$, with $\hat{\theta}_* = .033$. This test is inappropriate and essentially misleading, given the heterogeneity of the overlaps seen before.

Another biliary symptom generally associated with gallstone disease is *dyspepsia,* a nonspecific digestive disturbance characterized by nausea, belching, and flatus or abdominal discomfort, possibly with persistent but poorly defined pain. An identical analysis of the incidence of dyspepsia showed no significant difference between the two groups according to any of the tests described in Section 4.

In the original description of the principal results of the NCGS (based on simpler methods of analysis), it was concluded that "no deleterious or salutary effect on nausea, dyspepsia, or biliary pain was observed with either dose of chenodiol" (Schoenfield et al. 1981, p. 273). Despite this conclusion, it has been conjectured that patients who

Copyright © 2001 All Rights Reserved

Table 3. Multivariate Rank Analysis Using Wilcoxon Scores for the Rate of Nausea in the NCGS Over the Subintervals Corresponding to the Six Scheduled Reporting Times During the First Year of Follow-up

Statistic	Subinterval A_k (in study weeks) (k = 1, 6)					
	0–6	7–10	11–19	20–32	33–45	46–58
n_{1k} (cheno)	63	63	63	59	57	57
n_{2k} (placebo)	46	44	41	39	38	37
$\hat{\Pr}(X_k^{(2)} \geq X_k^{(1)})$	.876	.777	.828	.872	.762	.754
$\hat{\Pr}(X_k^{(1)} \geq X_k^{(2)})$	.729	.785	.824	.793	.874	.823
T_k	.375	−.019	.009	.159	−.214	−.127
$\hat{\Sigma}$	.0431	.0263	.0108	.0079	.0108	.0119
		.0444	.0168	.0063	.0118	.0118
			.0327	.0171	.0090	.0110
				.0270	.0144	.0091
					.0249	.0168
						.0281
χ_1^2	3.26	.008	.002	.937	1.83	.577
p value	.071	.927	.961	.333	.176	.448
$\hat{\theta}_k$	.147	−.008	.004	.079	−.112	−.069
SE ($\hat{\theta}_k$)	.082	.086	.080	.081	.083	.091

NOTE: SE represents standard error.

receive bile acid treatment for gallstones generally "feel better." In this article we have demonstrated that the incidence of nausea in the high-dose chenodiol group differs significantly from that of the placebo group over the first 58 weeks of treatment. Further, there is significant heterogeneity in the degree of overlap between groups in the event rates over time. The statistics $\hat{\boldsymbol{\theta}}$ in Table 3 indicate that the recurrence rate of nausea for the placebo group is generally greater than or equal to that of the high-dose group over the first six months of follow-up, with no substantial difference thereafter.

[*Received January 1986. Revised August 1987.*]

REFERENCES

Aalen, O. O. (1978), "Nonparametric Inference for a Family of Counting Processes," *The Annals of Statistics*, 6, 701–726.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1982), "Linear Nonparametric Tests for Comparison of Counting Processes, With Applications to Censored Survival Data" (with discussion), *International Statistical Review*, 50, 219–258.

Andersen, P. K., and Gill, R. D. (1982), "Cox's Regression Model for Counting Processes: A Large Sample Study," *The Annals of Statistics*, 10, 1100–1120.

Bancroft, T. A. (1972), "Some Recent Advances in Inference Procedures Using Preliminary Tests of Significance," in *Statistical Papers in Honor of George W. Snedecor*, Ames: Iowa State University Press, pp. 19–32.

Bancroft, T. A., and Han, C. P. (1977), "Inference Based on Conditional Specification: A Note and a Bibliography," *International Statistical Review*, 45, 117–127.

Chatterjee, S. K., and Sen, P. K. (1964), "Non-Parametric Tests for the Bivariate Two-Sample Location Problem," *Calcutta Statistical Association Bulletin*, 13, 18–58.

——— (1970), "Nonparametric Tests for the Multisample Multivariate Location Problem," in *Essays in Probability and Statistics in Memory of S. N. Roy*, eds. R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao, and K. J. C. Smith, Chapel Hill: University of North Carolina Press, pp. 197–228.

Cox, D. R. (1955), "Some Statistical Methods Connected With Series of Events" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 17, 129–164.

Cox, D. R., and Lewis, P. A. W. (1966), *The Statistical Analysis of Series of Events*, London: Methuen.

Gill, R. D. (1980), *Censoring and Stochastic Integrals* (Tract 124), Amsterdam: The Mathematical Centre.

——— (1984), "Understanding Cox's Regression Model: A Martingale Approach," *Journal of the American Statistical Association*, 79, 441–447.

Kalbfleisch, J. D., and Prentice, R. L. (1973), "Marginal Likelihoods Based on Cox's Regression and Life Model," *Biometrika*, 60, 267–278.

Karr, A. F. (1986), *Point Processes and Their Statistical Inference*, New York: Marcel Dekker.

Koch, G. G. (1969), "Some Aspects of the Statistical Analysis of 'Split Plot' Experiments in Completely Randomized Layouts," *Journal of the American Statistical Association*, 64, 485–505.

Koch, G. G., Amara, I. A., Stokes, M. E., and Gillings, D. B. (1980), "Some Views on Parametric and Non-Parametric Analysis for Repeated Measurements and Selected Bibliography," *International Statistical Review*, 48, 249–265.

Koziol, J. A., Maxwell, D. A., Fukushima, M., Colmerauer, M. E. M., and Pilch, Y. H. (1981), "A Distribution-Free Test for Tumor-Growth Curve Analyses With Application to an Animal Tumor Immunotherapy Experiment," *Biometrics*, 37, 383–390.

Lewis, P. A. W. (1972), "Recent Results in the Statistical Analysis of Univariate Point Processes," in *Stochastic Point Processes*, ed. P. A. W. Lewis, New York: John Wiley, pp. 1–54.

Prentice, R. L., and Gloeckler, L. A. (1978), "Regression Analysis of Grouped Survival Data With Application to Breast Cancer Data," *Biometrics*, 34, 57–67.

Puri, M. L., and Sen, P. K. (1966), "On a Class of Multivariate Multisample Rank-Order Tests," *Sankhya*, Ser. A, 28, 353–376.

Rao, C. R. (1972), *Linear Statistical Inference and Its Applications* (2nd ed.), New York: John Wiley.

Schoenfield, L. J., Lachin, J. M., et al. (1981), "National Cooperative Gallstone Study: A Controlled Trial of the Efficacy and Safety of Chenodeoxycholic Acid for Dissoluation of Gallstones," *Annals of Internal Medicine*, 95, No. 3, 257–282.

Thall, P. F. (1988), "Mixed Poisson Likelihood Regression Models for Longitudinal Interval Count Data," *Biometrics*, 44, 197–209.

Thistle, J. L., Cleary, P. A., Lachin, J. M., Tyor, M. P., Hersh, T., and the NCGS Group (1984), "The Natural History of Untreated Cholelithiasis During the National Cooperative Gallstone Study (NCGS)," *Annals of Internal Medicine*, 100, 171–175.

Wei, L. J., and Johnson, W. E. (1985), "Combining Dependent Tests With Incomplete Repeated Measurements," *Biometrika*, 72, 359–364.

Wei, L. J., and Lachin, J. M. (1984), "Two-Sample Asymptotically Distribution-Free Tests for Incomplete Multivariate Observations," *Journal of the American Statistical Association*, 79, 653–661.

Copyright © 2001 All Rights Reserved