A LOG-LINEAR ANALYSIS

OF A SET OF MEDICAL DATA

A LOG-LINEAR ANALYSIS

OF A SET OF MEDICAL DATA

By

BARBARA MOOK-PIK KO, B.A.

A PROJECT

SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

McMASTER UNIVERSITY

February 1978

MASTER OF SCIENCE (Statistics)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE : A Log-Linear Analysis of a Set of Medical Data

AUTHOR : Barbara Mook-Pik KO, B.A.(Hons) (University of Hong Kong, Hong Kong)

SUPERVISOR : Professor C.W. Dunnett

NUMBER OF PAGES : v, 89.

ABSTRACT

Methotrexate had been suspected to be harmful to the liver in psoriatic patients. The data of a prospective study to find out whether the drug affected the acquisition and worsening of various liver pathology was analysed. Personal particulars which would have adverse effect on the liver were also investigated. Log-linear models were fitted to this set of categorical data in the form of multidimensional contingency tables. It was found that methotrexate would be hepatotoxic if the drug was taken over a prolonged period and/or if the cumulative dose taken was large. Otherwise, methotrexate could be administered to psoriatic patients without causing much harm to the liver.

ACKNOWLEDGEMENTS

The author would like to express her utmost gratitude to her supervisor Professor C.W. Dunnett. His guidance and suggestions throughout the preparation of the entire project have been most valuable and enlightening. The author wishes to thank the Department of Applied Math, McMaster University for granting the Teaching Assistantship which makes the pursuit of the degree possible.

Appreciation is also due to Dr. H. Roenigk of the Department of Dermatology, Cleveland Clinic, for permitting the use of the data for the analysis.

The author is much indebted to the staff of C.S.U. and in particular to Mrs. Maria Wong for her kind assistance in computational work.

Finally the author sends her most sincere thanks to her dear parents, for their unfailing support; and to her husband, Pius, whose constant encouragement and patience have been invaluable.

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
II.	DATA DESCRIPTION	3
III.	METHOD	8
IV.	RESULT (PRESENTATION & DISCUSSION)	16
	A. FATTY METAMORPHOSIS	17
	B. NUCLEAR VARIABILITY	22
	C. PERIPORTAL INFLAMMATION	28
	D. NECROSIS	33
	E. FIBROSIS	38
	F. CIRRHOSIS	42
V.	CONCLUSION	47
VI.	APPENDIX I : THEORY	52
VII.	APPENDIX II : TABLES	64
VIII.	APPENDIX III : DATA COLLECTION FORM	81
IX.	REFERENCES	83

I. INTRODUCTION

In the past twenty years, methotrexate has been used in the treatment of severe psoriasis. In recent years much concern was raised regarding the potential of methotrexate producing liver damages in some psoriatic patients. Studies [8, 22, 23] had been conducted to find the association of hepatic injury with the drug with respect to the method of administration, schedules, frequencies, and total dose. There were reports supporting both sides of the question. To clarify the situation, a large scale retrospective and prospective study [24] was organised involving clinical investigation in the United States and Europe. In the retrospective study, liver biopsies were carried out on psoriatic patients who had been taking methotrexate already. Detailed reports of this study have been published [24] suggesting the existence of correlation between increasing cummulative dose of methotrexate with liver damage, cause and effect relationship of daily oral methotrexate drug schedule to hepatic fibrosis. The association of liver disease with heavy alcohol intake, obesity and diabetes has been previously reported.

This project presents the results of an analysis of a prospective study of the effect of the drug on various

hepatic pathology, namely fatty metamorphosis, nuclear variability, periportal inflammation, necrosis, fibrosis and cirrhosis. Here, liver biopsies were carried out both prior to and after methotrexate had been administered. Morphologic findings of pre-MTX and post-MTX liver biopsies are compared and the changes noted. The object of this study is to find out whether the histologic changes of liver pathology is related to methotrexate, and if so, how. It is also of interest how these changes are affected by the patients' personal particulars.

Most of the information and results of the study are categorical data. Multivariate analysis is not applicable since both dependent and independent variables are not quantitative. To analyse this set of categorical data in the form of multidimensional contingency table, log-linear models are fitted. The independent variables are sorted out in a stepwise manner to explain the change of each of the six liver pathology. Those factors which are found to be related to the dependent variable are collapsed over those which have no effect and log-linear models are again fitted. The odds of the expected cell frequencies are found to pronounce the way in which those factors included affect histologic changes in liver pathology after methotrexate has been administered.

II. DATA DESCRIPTION

Out of over one thousand patients from twelve participating clinics of the co-operative study, only 102 patients were studied prospectively. On these 102 psoriatic patients, 142 liver biopsies were performed. Since whether change occurred or not is the prime interest, for sequential post-MTX biopsies on the same patient, the result of the last biopsy was used. When more than one pathologist reviewed a liver biopsy specimen, the average of their scores carried to the nearest integer was analysed.

In the original questionaire of the co-operative study, part A consists of 10 questions on the patients' history. Part B consists of 12 questions listing laboratory values, physical examinations and methotrexate drug history. The last part is the interpretion of the pathologist. A copy of this data collection form is given in Appendix III.

Owing to the fact that there are only 102 patients being studied, the inclusion of all 22 independent variables plus the dependent variables would render a contingency table so huge that the total number of cells would well exceed the total entries. Hence some of these independent variables are eliminated from the present study after a preliminary examination. The following makes up the seven

independent variables of the analysis :

(1) Age at which MTX therapy started (AGE)

1. Less than 25 years old
2. 25 to 50 years old
3. 50 years and older

(2) Sex (SEX)

1. Male
2. Female

(3) Alcohol intake (ALCOHOL)

1. Non-drinker
2. 1 to 3 drinks per week
3. 1 to 3 drinks per day
4. 4 or more drinks per day

(4) Presence of any one of the following factors (FACTORS) - previous liver disease, diabetes, obesity, blood transfusion, hepatotoxin and physical findings of liver disease (significant hepatomegaly, jaundice, ascites, spider angioma or esophageal varices)

1. No
2. Yes

(5) Cumulative duration of MTX therapy (DURATION)

1. Less than 1 year
2. 1 to 2 years
3. Over 2 years

(6) Predominant dose schedule (SCHEDULE)

1. Daily oral
2. Weekly oral single dose
3. Weekly intramuscular or intravenous of a single dose
4. Weekly oral of divided dosage (4 doses over a 36 hour period weekly)

(7) Total cumulative dose of MTX taken (DOSE)

1. Less than 500 mg.
2. 500 to 1500 mg.
3. More than 1500 mg.

Previous studies had shown that daily oral schedule was associated with significant increase in nuclear variability, necrosis and fibrosis. Hence the schedule was not advised by most doctors and frequently dismissed by the clinics. For the study population of 102 patients, there are just two who were under Schedule 1. Only the remaining three schedules are considered in subsequent analysis, so there are three categories for variable 6.

In the questionaire besides cirrhosis which is classified as present or absent (2 categories), each of the other five liver pathology is graded according to the degree of severeness into four asymmetric categories :

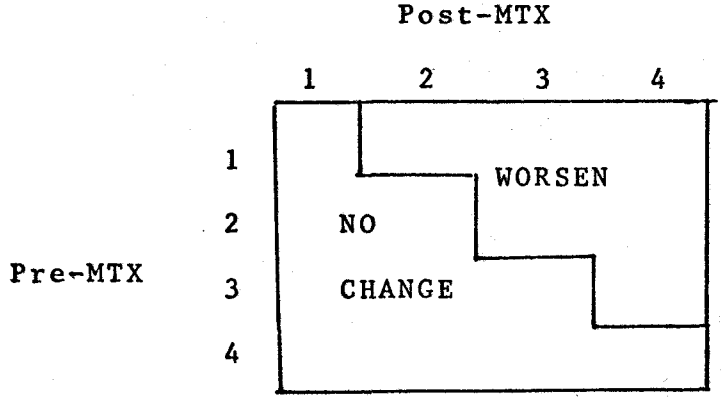
1. None
2. Slight

3. Moderate

4. Marked

For each of these five findings, if the "score" of the post-MTX biopsy is GREATER than the pre MTX one, it is termed to have "WORSEN". Otherwise, it is termed to have "NO CHANGE". This includes the cases when patients whose pre-MTX scores were higher than 1 had smaller post-MTX scores, which could mean that the patients improved after taking methotrexate, or it was because different pathologists have different interpretations of the degree of the histologic finding.

There was only 1 patient who had cirrhosis before taking methotrexate, and so just the 101 patients who started with none are used in the analysis. It is again categorized into "NO CHANGE" and "WORSEN" according to whether the post-MTX biopsy showed absence or presence of cirrhosis.



Each of the following six dependent variables form a eight- way contingency table with the seven independent

variables :

(8) FATTY METAMORPHOSIS

1. No change
2. Worsen

(9) NUCLEAR VARIABILITY

1. No change
2. Worsen

(10) PERIPORTAL INFLAMMATION

1. No change
2. Worsen

(11) NECROSIS

1. No change
2. Worsen

(12) FIBROSIS

1. No change
2. Worsen

(13) CIRRHOSIS

1. No change
2. Worsen

Table 1 presents the raw data of the analysis. The scores of each of the thirteen variables pertinent to every one of the 102 patients are listed under their respective headings. For the 13 variables, the frequencies of their various levels are tabulated in Tables 2 to 14 respectively.

III. METHOD

The aim of the study is to pick out the independent factor variables which affect each of the six response variables, fatty metamorphosis, nuclear variability, periportal inflammation, necrosis, fibrosis and cirrhosis. There are thus six 8-way contingency tables to be studied. Each of these tables consists of seven independent variables and one response variable. Using the SPSS[21], the response variables are classified into either one of the categories "NO CHANGE" or "WORSEN". Throughout the analysis, the Multivariate Contingency Tables program BMDP3F [9] is used to fit log-linear models. A discussion on Log-linear Models is given in Appendix I.

Analogous to the stepwise multiple regression procedure discussed by Draper and Smith [10], similar technique is applied to each of the six 8-way tables. However this stepwise procedure as applied to categorical data is not a standard method, and is a special technique developed and employed to analyse the present set of medical data. Either "step-up" or "step-down" fitting procedures may be employed to select the explaining variables sequentially. In the "step-down" process, the highest order log-linear model is fitted to the full contingency table. The factor which has the smallest interaction with the response is eliminated and the table is collapsed over that factor. The procedure is repeated for the collapsed table and factors which do

not affect the response are excluded one by one in order of their insignificance.

This method is not applicable to the 8-way tables under investigation. The 3 x 2 x 4 x 2 x 3 x 3 x 3 x 2 table has 2592 cells but only 102 entries. There will be too many marginal zeroes for adequate models to be fitted. Hence the "step-up" procedure is adopted instead.

Consider as an example the dependent variable fatty metamorphosis, i.e., variable 8. In the first step, two-way tables are formed with each of the factor variables 1 to 7 by collapsing over the remaining six independent variables. Log-linear model under the hypothesis of independence between factor and response is fitted. Models (1/8), (2/8), ... , (7/8) are fitted to the seven two-way tables respectively. The factor, say 7, which has the most significant chi-square value associated with the independent model, (7/8) is picked out. In other words, out of the seven factors, 7 is the one which is least independent of 8.

Next, six three-way tables are formed using variables 8, 7 and one of the remaining six factor variables, say Y. Since both 7 and Y are factors, the marginal totals C_{7Y} are prefixed by the design. In order that the fitted marginals would be equal to the observed ones, the interaction (7Y) must be included in the log-linear model for these 2-factors, 1-response contingency tables. The inter-

action (78) has been found to be significant in the first step, so the models (7Y/78) are fitted to the six three-way tables.

After the inclusion of the first explaining variable, the independence of variable 8 with the other factors (Y = 1, ... , 6), as well as the joint effect of 7 and Y on 8, which is the second order interaction (78Y) are tested, i.e.

$$H^* : u_{8Y} = u_{78Y} = 0$$

To test whether the interaction (8Y) is zero, assuming that (78Y) is zero, i.e.

$$H' : u_{8Y} = 0 \mid u_{78Y} = 0$$

find the difference between the tests-of-fit of the models differing only in the effect (8Y), namely (7Y/78) and (78/7Y/8Y).

$$X^2(8Y = 0 \mid 78Y = 0) = X^2(78/7Y) - X^2(78/7Y/8Y)$$

If this test statistic is significant, then reject the null hypothesis H' that there is no interaction between 8 and Y, assuming that no second-order interaction exists.

As an option of the program BMDP3F, the partial association and marginal association of all effects in the saturated model, 7, 8, Y, 78, 7Y, 8Y and 78Y are given. Judging from the chi-square values associated with these effects, their significance can easily be seen. The factor, say 1, which gives the most significant chi-square test is picked out as the second explaining variable. For the table

formed by variables 1, 7 and 8, the null hypothesis H* that (18) = (178) = 0 is rejected. Using the model (17/18/78), the significance of the three-factor interaction (178) can be tested. If (178) is significant, then factors 1 and 7 have joint effect on the response 8. By hierarchy, (18) must also be included in this case. Otherwise, accept the hypothesis that (18) $\neq$ 0 but (178) = 0. Then only the interaction (18) should be included in the model.

Four-way tables are formed using variables 1, 7, 8 and one of the five factors (2, .. , 6) not yet chosen. Interaction between response and factors which have been found to be significant in previous steps, and the interactions between factor variables are included in the log-linear models for these tables. Say if (178) has been found to be insignificant, then models (17Z/17/78) are fitted for Z = 2, ... , 6. The hypothesis tested is

$$H^{**} : \quad u_{8Z} = u_{1Z8} = u_{7Z8} = u_{178} = u_{17Z8} = 0$$

If (178) is found to be significant, models (178/17Z) are fitted. The corresponding hypothesis is

$$H^{**} : \quad u_{8Z} = u_{1Z8} = u_{7Z8} = u_{17Z8} = 0$$

The significance of the 3-factor interactions between the response and two independent variables, and the highest order interaction can be judged from their

partial association and marginal association. Generally no interactions of order higher than two are included in log-linear models fitted.

This selection process is stopped if the hypothesized models for every remaining variables are adequate and no other effects not yet included in the model indicate significance. Otherwise a third explaining variable is picked, and five-way tables are formed. The iteration terminates when all significant effects have been picked out. This iterative process is performed on every response variable, and factors which affect them are selected in a stepwise manner.

The criterion of selecting "factor" variables is the comparison of the chi-square statistic associated with the log-linear models fitted. In all log-linear models only interactions that include the response variable are assumed to be zero. The most significant chi-square value indicate the greatest lack-of-fit. In that case, some or all of the response-factors interactions that have been hypothesized to be zero should be included. This step-up procedure enables the inclusion of the explaining variables in order of their significance.

Some of the factor variables are correlated, for example, cumulative duration of methotrexate therapy (5) and cumulative dose of the drug (7). The related factors

may both appear to have association with the response in the first step. After the inclusion of the most significant one of these factors and the interaction between them, the other variable(s) may have no more significant effect on the response. The step-up process ensures that no such spurious factors are included.

Since the seven independent variables have different numbers of categories, the different models associated with them have unequal degrees of freedom. Therefore the magnitude of their chi-square statistics cannot be used to rank their importance. Two selection criteria are used to select the explaining factors. The tail probability associated with each of the chi-square value is given by the computer program. The one with the smallest probability is thus most significant. The chi-square statistic when divided by its own degree of freedom, under the null hypothesis is a "Chi-square over D.F."† which can be treated as if it has one degree of freedom. When all the statistics are standardized in such a way, they become comparable since the degrees of freedom are the same. Then the factor associated with the model giving the largest chi-square over d.f. value and the smallest tail probability is chosen to be included in subsequent analysis.

After all factors that affect the respone have been picked out, and the significant effects noted, the

† This term is used here in the sense referred to by Dixon & Massey [25] (p. 48, 102) as "Chi-square over degrees of freedom", i.e. $\chi^2/d.f.$

"simplest adequate" model is found for the dependent variable. Expected values for each cell and the parameters under that model is estimated. These are also optional output of the BMDP3F program.

Since the response variable is dichotomous, a linear logit model can be formed. For example for Fatty Metamorphosis (8), factors 1 and 7 are found to affect it. An adequate model is (17/18/78) and the log-linear model fitted is

$$e_{178(ijk)} = \log m_{178(ijk)}$$
$$= u + u_{1(i)} + u_{7(j)} + u_{8(k)} + u_{17(ij)} + u_{18(ik)} + u_{78(jk)}$$

then $\text{logit } M^{8}_{17(ij)} = \log \frac{m_{178(ij2)}}{m_{178(ij1)}}$

$$= \log m_{178(ij2)} - \log m_{178(ij1)}$$
$$= (u_{8(2)} - u_{8(1)}) + (u_{18(i2)} - u_{18(i1)}) + (u_{78(j2)} - u_{78(j1)})$$
$$= 2 (u_{8(2)} + u_{18(i2)} + u_{78(j2)})$$
$$= w_8 + w_{1(i)} + w_{7(j)}$$

and $M^{8}_{17(ij)} = \frac{m_{178(ij2)}}{m_{178(ij1)}}$

$$= \exp [w_8 + w_{1(i)} + w_{7(j)}]$$

In the logit model, the u-terms that do not involve the response variable are discarded. Those u-terms that are retained appear only in pairs whose members sum to zero. The w-terms have the same additive property as the u-terms. At the ith level of 1 and the jth level of 7, $M^{8}_{17(ij)}$ is the odds that fatty metamorphosis will worsen rather than remaining in the same state.

The value of the "odds" can hence be used to tell the effect of the various categories of the explaining variables. A low odds value indicate the probability that the response will worsen after taking methotrexate is small for the corresponding combination of the factors. On the other hand, a high odds value means that the liver pathology has a great chance of worsening for the categories of the factors it belong to. An odds value of one means that the liver pathology has an equal chance of worsening or remaining the same as the initial state before taking methotrexate.

In such a way, the effect of methotrexate on liver pathology can be seen. Schedules, duration and cumulative dose of methotrexate which are found to be harmful to the liver can be avoided in future administration of the drug.

IV. RESULT (PRESENTATION & DISCUSSION)

The data listed in Table 1 were analysed as described in Section III (Method) for each of the six liver pathology responses. Factors which were associated with the worsening of the liver abnormalities were chosen in a stepwise manner.

For the 102 patients under study, 74 had one post-MTX liver biopsy, 18 had two, 8 had three and 2 had four post-MTX biopsies. Table 15 corresponds to the cross-tabulation of the scores of the histological findings of Fatty Metamorphosis in the first post-MTX biopsy versus the pre-MTX finding; the second post-MTX scores versus the first; the third versus the second and the fourth post-MTX scores versus the third one. Table 16 is the cross-tabulation of the scores of the final post-MTX biopsy (the first one for 74, the second one for 18, the third one for 8 and the fourth one for 2) versus the pre-MTX findings.

Similar sequential tabulations for the responses Nuclear Variability are given in Tables 17 & 18; for Periportal Inflammation in Tables 19 & 20; for Necrosis in Tables 21 & 22; for Fibrosis in Tables 23 & 24; and for Cirrhosis in Tables 25 & 26. However just the last biopsy result was used in the analysis for the 28 patients who had

more than one post-MTX liver biopsy.

The step-up process of selecting the explaining variables is described below. The following notations are used :

1	denotes	AGE
2		SEX
3		ALCOHOL
4		FACTORS
5		DURATION
6		SCHEDULE
7		DOSE

The above are the independent variables.

8	FATTY METAMORPHOSIS
9	NUCLEAR VARIABILITY
10	PERIPORTAL INFLAMMATION
11	NECROSIS
12	FIBROSIS
13	CIRRHOSIS

These six are the dependent variables.

A. Fatty Metamorphosis (8)

For each of the seven 2-way tables formed with the independent variables, the independent models (X/8), X = 1, ... , 7 were fitted. In these and all subsequent models fitted, the value ½ was added to every cell count, as recommended by Goodman [13]. The hypothesis tested under these models are :

H_0 : X8 = 0 X = 1, ... , 7

Factor	Model	D.F.	L-R X^2	X^2 over D.F. (= $\frac{\text{L-R } X^2}{\text{D.F.}}$)	Probability
Age	1/8	2	1.73	0.865	0.4217
Sex	2/8	1	0.34	0.340	0.5609
Alcohol	3/8	3	2.47	0.823	0.4804
Factors	4/8	1	0.44	0.440	0.5069
Duration	5/8	2	4.88	2.440	0.0873
Schedule	6/8	3	0.80	0.267	0.8498
Dose	7/8	2	13.10	6.550	0.0014

The model [7/8] had the largest X^2 over d.f. value (6.55) and the smallest probability (0.0014), so factors 7 Cumulative Dose is least independent of fatty metamorphosis. Therefore 7 is chosen to be the first explaining variable for 8.

For the six 3-way tables, the models fitted were [Y7/78] Y = 1, ... , 6. The corresponding hypothesis tested were

H_1 : Y8 = Y78 = 0

Factors: Dose +	Model	D.F.	L-R X^2	X^2 over D.F.	Probability
Age	17/78	6	6.26	1.0433	0.3945
Sex	27/78	3	0.67	0.2333	0.8802
Alcohol	37/78	9	6.36	0.7067	0.7032
Factors	47/78	3	1.49	0.4967	0.6847
Duration	57/78	6	4.96	0.8267	0.5491
Schedule	67/78	6	2.60	0.4333	0.8575

Since none of the models fitted are significant at 5% level, and the partial and marginal associations of the interactions (Y8) are all insignificant, so the null hypothesis H_1 is accepted in each case. In other words, Cumulative Dose of Methotrexate is the only factor which affects the worsening of fatty metamorphosis. Cumulative Duration of methotrexate therapy (5) which has the second smallest probability in the first iteration (Prob [5/8] = 0.0873) is not significant after 7 had been included as a factor.

$$X^2\ (58 = 0 | 578 = 0) = X^2[57/78] - X^2[57/58/78]$$
$$= 4.96 - 4.63$$
$$= 0.33$$

D.F. = 2, Probability = 0.8489

The test statistic is not significant and so accept (58) = 0 given that (578) = 0. However the interaction between 5 and 7 is highly significant.

$$X^2(57 = 0 \mid 578 = 0) = X^2[58/78] - X^2[57/58/78]$$
$$= 46.32 - 4.63 = 41.69$$

D.F. = 4, Probability < 0.0001

The model $H^8 = [57/78]$ is chosen to be an adequate model to explain the change in fatty metamorphosis rather than collapsing over 5 since 5 and 7 are highly correlated. Other independent variables are found to have no significant interaction with 7 judging from their respective marginal and partial association with 7.

$$\log m_{578(ijk)} = u + u_{5(i)} + u_{7(j)} + u_{8(k)} + u_{57(ij)} + u_{78(jk)}$$
$$i = j = 1, 2, 3; \; k = 1, 2$$

The log-odds (logits) of the estimated incidence of worsening of fatty metamorphosis is defined as

$$\text{logit } M^{8}_{57(ij)} = \log \frac{m_{578(ij2)}}{m_{578(ij1)}}$$
$$= (u_{8(2)} - u_{8(1)}) + (u_{78(j2)} - u_{78(j1)})$$
$$= 2(u_{8(2)} + u_{78(j2)})$$
$$= w_8 + w_{7(j)}$$
$$i = j = 1, 2, 3$$

$$= \{-0.294\} + \begin{matrix}\{-0.930\}\\ \{-0.262\}\\ \{\ 1.192\}\end{matrix}$$

(8) (7) for i=1,2,3

The logits are a linear function of the effect due to cumulative dose (7), but the same for all levels of cumulative duration(5). The observed and expected cell frequencies under H^8 are given below :

Cumulative Duration(5)	Cumulative Dose (7)	Fatty Metamorphosis	
		No change	Worsen
< 1 year	< 500 mg	15 (14.955*)	4 (4.045)
	500 - 1500 mg	4 (3.949)	2 (2.051)
	> 1500 mg	1 (0.079)	0 (0.921)
1 - 2 years	< 500 mg	7 (7.227)	2 (1.773)
	500 - 1500 mg	21 (22.381)	14(12.619)
	> 1500 mg	1 (0.079)	0 (0.921)
> 2 years	< 500 mg	2 (1.818)	0 (0.182)
	500 - 1500 mg	11 (9.669)	4 (5.331)
	> 1500 mg	2 (3.842)	12(10.158)

( ) - Expexted value

* The expected cell counts presented in this table was obtained by subtracting ½ from the fitted values given by the model, since ½ had been added to every cell in the analysis and estimation process.

The logits and odds that fatty metamorphosis would worsen are :

		Cumulative Duration			Cumulative Duration			For all Durations
		< 1 yr	1-2 yrs	> 2 yrs	<1 yr	1-2 yrs	> 2 yrs	
Cumulative Dose	< 500 mg	-1.22	-1.223	-1.223	0.294†	0.294†	0.294†	0.250*
	500-1500mg	-.556	-0.556	-0.556	0.573	0.573	0.573	0.556
	>1500 mg	0.897	0.897	0.898	2.454	2.454	2.454	3.000
		Logits			Odds			

† Calculated from fitted values given by the model
* Calculated from expected cell counts (= fitted value - ½)

B. Nuclear Variability (9)

The results of fitting the independent models (X / 9), X = 1, ... , 7 to the 2-way tables are

Factor	Model	D.F.	L-R X^2	X^2 over D.F.	Probability
Age	1/9	2	3.39	1.695	0.1834
Sex	2/9	1	6.93	6.93	0.0085
Alcohol	3/9	3	7.64	2.547	0.0541
Factor	4/9	1	2.99	2.990	0.0839

Duration	5/9	2	3.73	1.865	0.1552
Schedule	6/9	3	3.39	1.130	0.3356
Dose	7/9	2	1.50	0.750	0.4715

Factor 2 has the largest chi-square over d.f. value (6.93) and the smallest probability (0.0085) for the independent model tested, so sex (2) is the first explaining variable for 9. Models [Y2/29], Y = 1, 3, 4, 5, 6, 7 are fitted to the six 3-way tables.

Factors : Sex +	Model	D.F.	L-R X^2	X^2 over D.F.	Probability
Age	12/29	4	4.47	1.1175	0.3464
Alcohol	23/29	6	7.16	1.1933	0.3060
Factors	24/29	2	10.24	5.1200	0.0060
Duration	25/29	4	4.33	1.0825	0.3629
Schedule	26/29	4	6.30	1.5750	0.1775
Dose	27/29	4	7.99	1.9975	0.0918

Since the model [24/29] has a large modified chi-square statistic (5.120) and a highly significant probability value, so the hypothesis tested under this model :

H_1 : 49 = 249 = 0

is rejected at 5% level.

When the model [24/29/49] was fitted, the L-R X^2 = 3.86 with 1 degree of freedom and tail probability = 0.0494, hence the hypothesis

H_1^* : 249 = 0

is again rejected at 5% level. Therefore there is a second order interaction among sex, other factors and change in nuclear variability. Presence of Factors (4) is selected as the second explaining variable.

Four-way tables are formed by 2, 4, 9 and one of the five remaining factors. Since (249) has been found to be significant, so models [24Z/249] , Z = 1, 3, 5, 6, 7 are fitted. Hypothesis tested under these models are :

H_2 : Z9 = 2Z9 = 4Z9 = 24Z9 = 0 Z = 1,3,5,6,7

Factors : Sex + Factors +	Model	D.F.	L-R X^2	X^2 over D.F.	Probability
Age	124/249	8	3.80	0.4750	0.8750
Alcohol	234/249	12	12.20	1.0167	0.4298
Duration	245/249	8	5.91	0.7388	0.6573

Schedule	246/249	8	6.86	0.8575	0.5514
Dose	247/249	8	6.60	0.8250	0.5799

All the models fitted are not significant at 10% level, and so all five null hypothesis H_2 are accepted. However from the partial association and marginal association option of the computer output, it was noticed that for factor 7 (Dose), the marginal association of the interaction (279) is significant.

$$X^2(279 = 0) = X^2[27/29/79] - X^2[279]$$

$$= 25.11 - 19.96 = 5.15$$

D.F. = 2, Probability = 0.0763

The partial association of this interaction is

$$X^2(279 = 0 \mid 2479 = 0) = X^2[247/249/479] - X^2[247/249/279/479]$$

$$= 2.98 - 0.02 = 2.96$$

D.F. = 2, Probability = 0.2272

The above test for marginal association shows that (279) is significant in the 3-way table formed by collapsing over variable 4. It is equivalent to the test for partial association in the model [27/29] considered on page 23. However, the test for partial association of (279) in the model [247/249] shows that, with variable 4 included in the model, (279) is no longer significant. In the following, the fit of the model H^9 = [247/249/279] is shown first, followed by the fit of the model H^9 = [247/249].

$$\log m_{2479(ijkl)} = u + u_{2(i)} + u_{4(j)} + u_{7(k)} + u_{9(l)} + u_{24(ij)} + u_{27(ik)} + u_{29(il)} + u_{47(jk)} + u_{49(jl)} + u_{79(kl)} + u_{247(ijk)} + u_{249(ijl)} + u_{279(ikl)}$$

The logits of the expected frequencies are :

$$\text{logit } M^{9}_{247(ijk)} = \log \frac{m_{2479(ijk2)}}{m_{2479(ijk1)}}$$

$$= u_{9(2)} - u_{9(1)} + u_{29(i2)} - u_{29(i1)} + u_{49(j2)} - u_{49(j1)} + u_{79(k2)} - u_{79(k1)} + u_{249(ij2)} - u_{249(ij1)} + u_{279(ik2)} - u_{279(ik1)}$$

$$= w_9 + w_{2(i)} + w_{4(j)} + w_{7(k)} + w_{24(ij)} + w_{27(ik)}$$

$$= \{-0.556\} + \begin{matrix}\{-0.896\}\\ \{0.896\}\end{matrix} + \begin{matrix}\{0.386\}\\ \{-0.384\}\\ \{-0.022\}\end{matrix} + \begin{matrix}\{0.56\}\\ \{-0.56\}\end{matrix} +$$

$$\uparrow \qquad \uparrow \qquad \uparrow \qquad \uparrow$$

$$(9) \qquad (2) \qquad (7) \qquad (4)$$

$$\begin{matrix} \{\ 0.438\} & \{-0.438\} \\ \{-0.438\} & \{\ 0.438\} \end{matrix} \quad + \quad \begin{matrix} \{-0.558\} & \{\ 0.434\} & \{\ 0.124\} \\ \{\ 0.558\} & \{-0.434\} & \{-0.124\} \end{matrix}$$

$$(2) \times (4) \qquad\qquad (2) \times (7)$$

The observed and expected cell frequencies are :

Dose (7)	Factors (4)	Sex (2)	Nuclear Variability: No Change	Nuclear Variability: Worsen
Less than 500 mg	No	Male	8 (7.973)	4 (4.026)
		Female	1 (1.070)	6 (5.928)
	Yes	Male	6 (6.027)	0 (-0.027)
		Female	1 (0.930)	4 (4.072)
500 - 1500 mg	No	Male	13 (13.178)	9 (8.8220
		Female	5 (5.324)	4 (3.679)
	Yes	Male	7 (6.822)	0 (0.178)
		Female	12 (11.676)	6 (6.321)
More than 1500 mg	No	Male	2 (1.849)	1 (1.151)
		Female	2 (1.606)	2 (2.394)
	Yes	Male	3 (3.151)	0 (-0.151)
		Female	2 (2.394)	3 (2.606)

The log-odds and the odds that the nuclear variability would worsen are :

Dose	Factors	Sex		Sex	
		Male	Female	Male	Female
Less than 500 mg	No	-0.626	1.406	0.5347	4.0796
	Yes	-2.6225	1.1623	0.0726	3.1972
500 - 1500 mg	No	-0.384	-0.3319	0.6811	0.7175
	Yes	-2.3795	-0.5795	0.0926	0.5602
More than 1500 mg	No	-0.3526	0.3178	0.7029	1.3742
	Yes	-2.3477	0.0707	0.0956	1.0733
		Logits		Odds	

On the other hand, if we used the model $H^9 = [247/249]$, which also fits the data satisfactorily, we obtain the following:-

$$\log m_{2479(ijkl} = u + u_{2(i)} + u_{4(j)} + u_{(7(k)} + u_{9(l)} + u_{24(ij)} + u_{27(ik)} + u_{29(il)} + u_{47(jk)} + u_{49(jl)} + u_{247(ijk)} + u_{249(ijl)}$$

The logits of the expected frequencies are:

$$\text{logit } M^{9}_{247(ijk)} = \log \frac{m_{2479(ijk2)}}{m_{2479(ijk1)}}$$

$$= u_{9(2)} - u_{9(1)} \quad u_{29(i2)} -$$

$$u_{29(i1)} \quad u_{49(j2)} - u_{49(j1)}$$

$$u_{249(ij2)} - u_{249(ij1)}$$

$$= w_{9} \quad w_{2(i)} \quad w_{4(j)} \quad w_{24(ij)}$$

$$\{-0.784\} \quad \{0.620\} \quad \{0.380\} \quad \{-0.380\}$$

$$= \{-0.674\} + \quad + \quad + \quad$$

$$\{0.784\} \quad \{-0.620\} \quad \{-0.380\} \quad \{0.380\}$$

$$\uparrow \quad \uparrow \quad \uparrow \quad \uparrow$$

$$(9) \quad (2) \quad (4) \quad (2) \times (4)$$

The observed and expected cell frequencies are:

Dose (7)	Factors (4)	Sex (2)	Nuclear Variability (9)	
			No Change	Worsen
Less than 500 mg	No	Male	8 (7.46)	4 (4.54)
		Female	1 (2.80)	6 (4.20)
	Yes	Male	6 (5.95)	0 (0.05)
		Female	1 (2.69)	4 (2.31)
500-1500 mg	No	Male	13 (13.59)	9 (8.41)
		Female	5 (3.63)	4 (5.37)
	Yes	Male	7 (6.87)	0 (0.13)
		Female	12 (9.61)	6 (8.39)
More than 1500 mg	No	Male	2 (1.95)	1 (1.05)
		Female	2 (1.57)	2 (2.43)
	Yes	Male	3 (3.18)	0 (-0.18)
		Female	2 (2.69)	3 (2.31)

The log-odds and the odds that the nuclear variability would worsen are as follows, for all three cumulative dose levels:-

Factors	Sex		Sex	
	Male	Female	Males	Females
No	-0.458	0.350	0.633	1.419
Yes	-2.458	-0.130	0.086	0.879
	Logits		Odds	

C. Periportal Inflammation (10)

Models fitted to the 2-way tables for testing independence are (X/10), X = 1, ... , 7

Factor	Model	D.F.	L-R X^2	X^2 over D.F.	Probability
Age	1/10	2	2.86	1.430	0.2387
Sex	2/10	1	1.41	1.410	0.2354

Alcohol	3/10	3	0.21	0.070	0.9755
Factors	4/10	1	3.88	3.880	0.0489
Duration	5/10	2	10.89	5.445	0.0043
Schedule	6/10	3	3.94	1.313	0.2682
Dose	7/10	2	4.42	2.215	0.1094

The model [5/10] had the largest chi-square over d. f. value and smallest probability. So Cumulative Duration of methotrexate therapy (5) is selected to be the first explaining variable for periportal inflammation.

Factors : Duration +	Model	D.F.	L-R X^2	X^2 over D.F	Probability
Age	15/5,10†	6	6.43	1.0717	0.3768
Sex	25/5,10	3	5.57	1.8567	0.1347
Alcohol	35/5,10	9	9.20	1.0222	0.4193
Factors	45/5,10	3	5,73	1.9100	0.1254
Schedule	56/5,10	6	3.80	0.6333	0.7060
Dose	57/5,10	6	1.73	0.2883	0.9429

The model [45/5,10] is significant at 15% level. To test if the second-order interaction is zero, the model [45/4,10/5,10] is fitted. The L-R chi-square statistic is 3.63 with 2 degress of freedom and the tail probability is 0.1627. Hence accept

$$H_1^* : (45,10) = 0$$

The partial association and marginal association of (4,10) are both slightly significant and so factor 4 (Factors) is chosen to be the second explaining variable for peri-portal inflammation.

Factors : Duration + Factors +	Model	D.F.	L-R X^2	X^2 over D.F.	Probability
Age	5,10†/4,10†/145	14	10.69	0.7636	0.7099
Sex	5,10/4,10/245	8	8.63	1.0788	0.3743
Alcohol	5,10/4,10/345	20	17.49	0.8745	0.6209
Schedule	5,10/4,10/456	14	14.59	1.0421	0.4066
Dose	5,10/4,10/457	14	10.70	0.7643	0.7092

Since none of these models are significant, the corresponding null hypothesis that

$$H_2 : (Z,10) = (Z4,10) = (Z5,10) = (45,10) = (45Z,10) = 0$$

are accepted. Therefore an adequate model for periportal inflammation is $H^{10} = [45/4,10/5,10]$. In other words, other factors (4) and cumulative duration of methotrexate therapy (5) have first-order interactions with periportal inflammation respectively.

$$\log m_{45,10(ijk)} = u + u_{4(i)} + u_{5(j)} + u_{10(k)} + u_{45(ij)} + u_{4,10(ik)} + u_{5,10(jk)}$$

$$i = 1,2;\ j = 1,2,3;\ k = 1,2$$

The logits are $\text{Logit } M^{10}_{45(ij)} = \log \dfrac{m_{45,10(ij2)}}{m_{45,10(ij1)}}$

$$= w_{10} + w_{4(i)} + w_{5(j)}$$

$$= \{-0.868\} + \begin{matrix}\{-0.38\}\\ \{\ 0.38\}\end{matrix} + \begin{matrix}\{-0.360\}\\ \{-0.558\}\\ \{\ 0.919\}\end{matrix}$$

(10) (4) (5)

The expected and observed cell frequencies are :

Other Factors (4)	Cumulative Duration (5)	Periportal Inflammation	
		No Change	Worsen
No	< 1 year	15 (14.498)	2 (2.501)
	1 - 2 years	24 (22.702)	2 (3.307)
	> 2 years	7 (8.800)	8 (6.192)
Yes	< 1 year	6 (6.504)	3 (2.497)
	1 - 2 years	13 (14.297)	6 (4.693)
	> 2 years	8 (6.199)	8 (9.810)

The logits and the odds are :

		Factors		Factors	
		No	Yes	No	Yes
Cumulative Duration	< 1 year	-1.6090	-0.8489	0.2001	0.4279
	1-2 years	-1.8074	-1.0471	0.1641	0.3509
	> 2 years	-0.3291	0.4312	0.7196	1.5390
		Logits		Odds	

† Note : Comma has to be inserted between 5 and 10 (i.e., 5,10) and between 4 and 10 (i.e., 4,10) to differentiate between the two pairs of variables for the interactions between 5 and 10, 4 and 10 respectively. The same notation is employed hereafter for interactions involving variables 10, 11, 12 and 13.

D. Necrosis (11)

Independent models (X/11), X = 1, ... , 7 are fitted to the 2-way tables

Factor	Model	D.F.	L-R X^2	X^2 over D.F.	Probability
Age	1/11	2	0.84	0.420	0.6580
Sex	2/11	1	0.54	0.540	0.4605
Alcohol	3/11	3	10.19	3.3967	0.0170
Factors	4/11	1	3.71	3.710	0.0539
Duration	5/11	2	4.69	2.345	0.0960
Schedule	6/11	3	4.71	1.5700	0.1944
Dose	7/11	2	2.35	1.1750	0.3084

Factors 3 and 4 have close chi-square over d.f. statistics and probabilities. Factor 3 is chosen to be the first explaining variable because of its small probability indicates that the interaction between alcohol intake and necrosis is significant at 5% level (probability = 0.017)

Factors : Alcohol +	Model	D.F.	L-R X^2	X^2 over D.F.	Probability
Age	13/3,11	8	3.82	0.4775	0.8726
Sex	23/3,11	4	2.68	0.6700	0.6124
Factors	34/3,11	4	3.26	0.8150	0.5149
Duration	35/3,11	8	6.18	0.7725	0.6274
Schedule	36/3,11	8	8.18	1.0225	0.4173
Dose	37/3,11	8	3.25	0.4063	0.9174

All the models fitted appear to be adequate. However, factor 4 is checked since it is significantly associated with necrosis in the first iteration though the model [34/3,11] has a small L-R chi-square statistic.

$$X^2[34/3,11/4,11] = 0.35$$

D.F. = 3, Probability = 0.9501

Hence accept the hypothesis that there is no second order interaction between 3, 4 and 11, i.e.

$$H_1^* : (34,11) = 0$$

But it was found that the partial association of (4,11) is fairly significant.

$$X^2(4,11 = 0 \mid 34,11 = 0) = X^2[34/3,11] - X^2[34/3,11/4,11]$$

$$= 3.26 - 0.35 = 2.91$$

D.F. = 1, probability = 0.0880

The marginal association is

$$X^2(4,11 = 0) = X^2[4,11] - X^2[4/11]$$

$$= 48.05 - 44.38 = 3.67$$

D.F. = 1, probability = 0.0553

Both are significant at 10% level, and so accept the hypothesis that $(4,11 \neq 0 \mid 34,11 = 0)$. So the second explaining variable for necrosis is the presence of other factors (4).

Factors : Alcohol + Factors +	Model	D.F.	L-R X^2	X^2 over D.F.	Prob
Age	3,11/4,11/134	19	6.33	0.3332	0.9970
Sex	3,11/4,11/234	11	5.57	0.5064	0.9002
Duration	3,11/4,11/345	19	8.15	0.4289	0.9851
Schedule	3,11/4,11/346	19	9.47	0.4984	0.9641
Dose	3,11/4,11/347	19	4.70	0.2474	0.9996

The above models are all adequate and so no third

factor is chosen. Change in necrosis is thus associated with alcoholic intake (3) and with the presence of other factors (4). An adequate model is $H^{11} = [34/3,11/4,11]$.

$$\log m_{34,11(ijk)} = u + u_{3(i)} + u_{4(j)} + u_{11(k)} + u_{34(ij)} + u_{3,11(ik)} + u_{4,11(jk)}$$

$$i = 1,2,3,4;\ j = 1,2;\ k = 1,2$$

$$\text{logit } M^{11}_{34(ij)} = \log \frac{m_{34,11(ij2)}}{m_{34,11(ij1)}} = w_{11} + w_{3(i)} + w_{4(j)}$$

$$= \{-1.474\} + \begin{Bmatrix} -0.710 \\ 0.816 \\ -0.776 \\ 0.670 \end{Bmatrix} + \begin{Bmatrix} -0.408 \\ 0.408 \end{Bmatrix}$$

(11) (3) (4)

The expected and observed cell frequencies are :

Other Factors	Alcoholic Intake	Necrosis	
		No Change	Worsen
No	Non-drinker	14 (14.386)	1 (0.615)
	1-3 drinks/week	20 (19.579)	6 (6.419)
	1-3 drinks/day	8 (7.911)	0 (0.089)
	> 3 drinks/day	4 (4.124)	1 (0.877)
Yes	Non-drinker	11 (10.614)	1 (1.385)
	1-3 drinks/week	12 (12.421)	10 (9.581)
	1-3 drinks/day	2 (2.089)	0 (-0.089)
	> 3 drinks/day	5 (4.876)	3 (3.124)

The logits and the odds are :

		Other Factors		Other Factors	
		No	Yes	No	Yes
Alcoholic Intake	Non-drinker	-2.5916	-1.7743	0.0749	0.1696
	1-3 drinks/week	-1.0654	-0.2482	0.3446	0.7802
	1-3 drinks/day	-2.6589	-1.8408	0.0700	0.1587
	> 3 drinks/day	-1.2114	-0.3944	0.2978	0.6741
		Logits		Odds	

E. Fibrosis (12)

The independent models (X/12), X = 1, ... , 7 are again fitted to the seven two-way tables.

Factors	Model	D.F.	L-R X^2	X^2 over D.F.	Probability
Age	1/12	2	2.98	1.490	0.2250
Sex	2/12	1	2.23	1.115	0.1355
Alcohol	3/12	3	2.77	0.923	0.4287
Factors	4/12	1	1.40	1.400	0.2360
Duration	5/12	2	9.51	4.755	0.0086
Schedule	6/12	3	7.25	2.417	0.0643
Dose	7/12	2	8.42	4.210	0.0148

Factor 5 and Fibrosis are least independent of each other. So cumulative duration of methotrexate therapy (5) is selected to be the first factor to explain the change of fibrosis in the psoriatic patients.

Factors: Duration +	Model	D.F.	L-R X^2	X^2 over D.F.	Prob
Age	15/5,12	6	5.36	0.8933	0.4980
Sex	25/5,12	3	3.46	1.1533	0.3257
Alcohol	35/5,12	9	5.18	0.5756	0.8187
Factors	45/5,12	3	0.45	0.1500	0.9309
Schedule	56/5,12	6	5.36	0.8933	0.5002
Dose	57/5,12	6	2.73	0.4550	0.8421

All the models are not significant and so the null hypothesis that

H_1 : Y,12 = 5Y,12 = 0 Y = 1, 2, 3, 4, 6, 7

are accepted. However 7 and 12 are significantly dependent of each other in the first iteration (Prob [7/12] = 0.0148 X^2 over d.f.= 4.21), so the 3-way table with variables 5, 7 and 12 are inspected.

X^2[57/5,12/7,12] = 0.54

D.F. = 4, Probability = 0.9692

Therefore the interaction (57,12) equals to zero. The marginal association of (7,12) is

$$X^2(7,12 = 0) = X^2[7/12] - X^2[7,12]$$

$$= 64.75 - 56.96 = 7.79$$

D.F. = 2 Probability = 0.0399

The partial association is

$$X^2(7,12 = 0 \mid 57,12 = 0) = X^2[57/5,12] - X^2[57/5,12/7,12]$$

$$= 2.73 - 0.54 = 2.19$$

D.F. = 2 Probability = 0.3353

It can be seen that the interaction (7,12) is slightly significant. It has been noted that 5 and 7 are highly correlated. Hence both cumulative duration (5) and cumulative dose (7) are factors affecting the change in fibrosis, and an adequate model for this set of data is $H^{12} = [57/5,12/7,12]$.

$$\log m_{57,12(ijk)} = u + u_{5(i)} + u_{7(j)} + u_{12(k)} + u_{57(ij)} + u_{5,12(ik)} + u_{7,12(jk)}$$

$$i = j = 1, 2, 3; \quad k = 1,2$$

The logit is

$$\text{logit } M^{12}_{57(ij)} = \log \frac{m_{57,12(ij2)}}{m_{57,12(ij1)}}$$

$$= w_{12} + w_{5(i)} + w_{(j)}$$

$$= \begin{matrix} \{-1.3040\} \\ \uparrow \\ (12) \end{matrix} + \begin{matrix} \{-0.350\} \\ \{-0.296\} \\ \{0.648\} \\ \uparrow \\ (5) \end{matrix} + \begin{matrix} \{-0.620\} \\ \{-0.014\} \\ \{0.634\} \\ \uparrow \\ (7) \end{matrix}$$

The observed and expected cell frequencies for the three-way table are :

Cumulative Dose	Cumulative Duration	Fibrosis	
		No Change	Worsen
< 500 mg	< 1 year	18 (17.634)	1 (1.366)
	1 - 2 years	8 (8.521)	1 (0.479)
	> 2 years	2 (1.845)	0 (0.155)
500-1500 mg	< 1 year	5 (5.389)	1 (0.611)
	1 - 2 years	30 (29.526)	5 (5.474)
	> 2 years	10 (10.084)	5 (4.916)
> 1500 mg	< 1 year	1 (0.970)	0 (0.030)
	1 - 2 years	1 (0.949)	0 (0.051)
	> 2 years	7 (7,082)	7 (6.918)

The log-odds and the odds that fibrosis would worsen is :

		Cumulative Dose			Cumulative Dose		
		<500mg	500-1500mg	>1500mg	<500mg	500-1500mg	>1500mg
Duration	<1 yr	-2.2740	-1.6678	-1.0201	0.1029	0.1887	0.3605
	1-2 yrs	-2.2208	-1.6146	-0.9669	0.1085	0.1990	0.3803
	>2 yrs	-1.2754	-0.6700	-0.0219	0.2793	0.5117	0.9784
		Logits			Odds		

F. Cirrhosis (13)

Since there is only one patient who had cirrhosis before the methotrexate therapy and cirrhosis was coded only as absent or present, so just the 101 patients who did not have cirrhosis initailly are analysed. Independence between factors and response is again postulated.

H_0 : $X,13 = 0$ $\quad X = 1, \ldots, 7$

Factor	Model	D.F.	L-R X^2	X^2 over D.F.	Probability
Age	1/13	2	1.19	0.5950	0.5508
Sex	2/13	1	0.01	0.0100	0.9142
Alcohol	3/13	3	2.68	0.8933	0.4439
Factors	4/13	1	0.00	0.0000	0.9928
Duration	5/13	2	4.80	2.4000	0.0906
Schedule	6/13	2	0.558	0.2790	0.7607
Dose	7/13	2	5.00	2.5000	0.0821

There are only 7 patients who acquired cirrhosis after the methotrexate therapy out of the 101 under study. None of the above models is significant at 5% level. However factor 7 and 13 have a X^2over d.f. of 2.50 and a probability 0.0821, so the hypothesis

$$H_0 : \gamma(5,13) = 0$$

is rejected at 10% level. Thus 7 (Dose) is the first explaining variable for cirrhosis.

Factors : Dose +	Model	D.F.	L-R X^2	X^2 over D.F.	Prob
Age	17/7,13	6	1.95	0.3250	0.9244
Sex	27/7,13	3	1.78	0.5933	0.6202
Alcohol	37/7,13	9	5.11	0.5678	0.8249
Factors	47/7,13	3	2.13	0.7100	0.5465
Duration	57/7,13	6	2.98	0.4967	0.8110
Schedule	67/7,13	6	5.48	0.9133	0.5144

The models are all adequately fitted and so the corresponding null hypothesis that

H_1 : (Y,13) = (Y7,13) = 0 Y = 1, ... , 6

are accepted. Since 7 and 5 are correlated, the 3-way table with factors 5, 7 and 13 is studied.

X^2[57/5,13/7,13] = 2.11

D.F. = 4 Probability = 0.7159

So there is no second order interaction. Both the partial and marginal association of the interaction between 5 and 13 are small, so this effect is not significant and is thus hypothesized to be zero. Hence there is only one explain-

ing variable for acquiring cirrhosis, namely the cumulative dose of methotrexate taken. The model that is fitted is [57/7,13].

$$\log m_{57,13(ijk)} = u + u_{5(i)} + u_{7(j)} + u_{13(k)} + u_{57(ij)} + u_{7,13(jk)}$$

$$i = j = 1, 2, 3; \quad k = 1, 2$$

The logit that a patient would have cirrhosis after taking methotrexate is

$$\text{logit } M^{13}_{57(ij)} = \log \frac{m_{57,13(ij2)}}{m_{57,13(ij1)}}$$

$$= w_{13} + w_{7(j)} \quad \text{for all } i$$

$$= \{-2.152\} + \begin{matrix}\{-0.122\} \\ \{\ 0.982\}\end{matrix}$$

$$\uparrow \qquad \uparrow$$

$$(13) \qquad (7)$$

The observed and expected cell frequencies for the data are :

Cumulative Duration	Cumulative Dose	Cirrhosis	
		No Change	Worsen
< 1 year	< 500 mg	18 (17.609)	0 (0.391)
	500 - 1500 mg	5 (5.847)	1 (0.153)
	> 1500 mg	1 (1.026)	0 (-0.026)
1 - 2 years	< 500 mg	9 (9.031)	0 (-0.031)
	500 - 1500 mg	34 (32.144)	1 (2.856)
	> 1500 mg	1 (1.026)	0 (-0.026)
> 2 years	< 500 mg	2 (2.359)	0 (-0.359)
	500 - 1500 mg	13 (14.008)	2 (0.992)
	> 1500 mg	11 (10.947)	3 (3.053)

The logits and the odds that a patient who did not have cirrhosis initially would acquire it after taking methotrexate are :

		Logits			Odds		
		Cumulative Duration					
		<1 yr	1-2 yrs	>2 yrs	<1 yr	1-2 yr	>2 yr
Dose	<500mg	-3.0118	-3.0118	-3.0109	0.0492	0.0492	0.0493
	500-1500mg	-2.2742	-2.2749	-2.2749	0.1029	0.1028	0.1028
	>1500mg	-1.1692	-1.1692	-1.1699	0.3106	0.3106	0.3105

V. CONCLUSION

From the results, it can be seen that the six liver pathology are affected by different factors. Out of the seven independent variables under study, both age at which methotrexate therapy began and schedule of administering the drug have no significant effect on any response. Therefore the age of the psoriatic patients would have no association with worsening or acquiring liver disease after taking methotrexate.

Besides the schedule of taking the drug orally every day which has been found to be hepatotoxic, the other three schedules of administration are not significantly different from one another and have not been found to affect the change in liver pathologies.

Fatty metamorphosis is dependent on the cumulative dose of methotrexate taken. It has a high chance of worsening after taking over 1500 mg of the drug (odds = 2.4542). The relative odds that the fatty metamorphosis would worsen for the cumulative dose over 1500 mg against that under 1500 mg (i.e. those <500 mg. and those with 500-1500 mg) is 2.454/0.4603 = 5.332. Therefore the odds that it would go worse increases rather markedly with increasing dose.

Nuclear variability is affected jointly by the sex

of the patient and the presence of other factors, and by sex and cumulative dose jointly as well. It is noticed from the odds of nuclear variability would worsen that females have much greater chance of getting worse. The relative odds that females would worsen against male is 1.0769/0.4048 = 2.6604. Those who have previous liver disease, diabetes obesity et. al. are expected to have a higher chance of worsening. However for nuclear variability those who do not have such factors are more subjectable to worsen. The relative odds of no other factors against some is 0.8529/0.4705 = 1.8128, which is a rather unusual result. The second order interaction of sex and dose with nuclear variability (279) is only marginally significant (Prob = 0.0763), so the inclusion of this effect is rather debatable. However, since cumulative dose (7) is a subject of interest in this investigation, so (279) is included in the model H^9 to show any particular effect that methotrexate might have on nuclear variability. It is noticed from the odds value that the probability of worsening in nuclear variability increases with increasing dose of methotrexate for male patients. Nevertheless, females who have only small cumulative dose (under 500 mg) and have no previous liver disease, nor diabetes, et. al. have very high probability of getting worse in nuclear variability (odds = 4.079). The odds is not as high for females with greater cumulative dose (odds = 1.3742). This may be due to the

fact that there are only 9 female patients who have taken more than 1500 mg of methotrexate and so the effect cannot be seen as distinctly. Further investigation with larger sample size is recommended to establish a more conclusive relationship amongst nuclear variability, sex, cumulative dose of methotrexate and presence of other factors.

Periportal inflammation is associated with the cumulative duration of taking methotrexate and with the presence of other factors. The chance that those who have previous liver disease or diabetes or obesity, et. al. would get worse in periportal inflammation doubles the chance of those who do not. The relative odds of presence against absence of other factors is 0.6497/0.2842 = 2.2847. Moreover the probability increases with the duration as well. Patients who have the therapy for less than two years have odds of worsening way below one. The relative odds of those who have taken methotrexate for over two years against those for under two years is 1.0625/0.25 = 4.25, which is quite significant.

The worsening of necrosis is not affected by methotrexate in any way. Patients with presence of other factors such as previous liver disease, diabetes, obesity et. al. have twice the chance of those with none in getting worse in necrosis, and the probability increases with the consumption of alcohol. However none of the odds exceeds one, so necrosis have fairly high chance of undergoing no change after the administration of methotrexate to psoriatic patients.

Fibrosis is associated with the cumulative duration and with cumulative dose of methotrexate. This agrees with findings of previous studies. Of the 20 patients in the "Worsen" category, 19 of them acquired fibrosis after taking methotrexate and 1 of them who had it initially got worse. The odds of acquiring fibrosis increases with increasing duration and dose of the drug. The relative odds of those taking the drug over 2 years against those under 2 years is 0.6585/0.1667 = 3.95. The probability increases progressively with cumulative dose. All the odds are less than 1, meaning that the probability of acquiring fibrosis is less than that of remaining in the same state. However the odds for patients who have methotrexate for more than two years and over 1500 mg is approximately 1, and so it becomes equiprobable. Therefore methotrexate would be hepatotoxic if large dose has been taken for an extended long period.

Though cumulative dose is a factor affecting cirrhosis, only 7 patients acquired it out of the 101 patients studied. The chance of getting cirrhosis for patients who had over 1500 mg of methotrexate is about three times those who have under 1500 mg. The relative odds of cumulative dose over 1500 mg against dose under 1500 mg is 0.3103/0.0833 = 3.72. Yet the odds is still small for high cumulative dose (odds = 0.3103). Therefore

the occurrence of cirrhosis in psoriatic patients cannot be attributed to methotrexate directly even though large cumulative dose is found to be harmful to the liver.

In a general sense, methotrexate is not harmful to the liver if the amount taken is not large and for a short period. However it has certain adverse effects hepatically if the cumulative dose becomes large and/or the cumulative duration is too long. Patients who have other factors such as previous liver disease, diabetes, obesity et. al. have higher chance of getting worse in liver pathology. Females have a remarkably high chance of worsening in necrosis. Alcohol intake is found to be associated with liver disease, which has been pointed out before in other retrospective studies. Thus methotrexate can be employed for psoriasis so long as the drug is not taken daily orally, is used for less than two years and the total dose is not too large. Otherwise, methotrexate tends to have adverse effect on liver pathology.

VI. APPENDIX I : THEORY (THE LOG-LINEAR MODEL)

The present set of data under study consists of 13 variables from each of 102 individuals are discrete multivariates. The units of the sample are classified according to the categories they belong to. Linear model approach like multiple regression, multivariate analysis of variance cannot be used owing to the fact that most of the variables are not quantitative.

Different authors have proposed different methods for multivariate analysis of m qualitative (m = 2, 3,..) variables forming a m-way contingency table. Birch [3], Goodman [13, 14, 15], Bishop [4, 5], Fienberg [11], Ku & Kullback [19], Bhapkar & Koch[2], et. al. have investigated extensively on the various aspects of fitting "Log- linear model" to multidimensional contingency tables. Literature on model fitting, hypothesis testing, expected frequencies and effect estimation, interactions among multiple classifications and partitioning are well written but too numerous to be discussed here. Only the details of fitting a log-linear model to a four way contingency table is presented.

Consider individuals of a sample of size N classified according to the variables 1, 2, 3 and 4 with

levels 1_i, 2_j, 3_k and 4_l respectively, where i = 1, .. , r; j = 1, .. , s; k = 1, .. , t; l = 1, .. , u. The total number of entries in the table is N, and the number of counts for the elementary cell (i, j, k, l) corresponding to the ith level of 1, the jth level of 2, the kth level of 3 and the lth level of 4 is denoted by n_{ijkl}.

When the cell frequencies are summed over a particular variable, then the subscript for the variable is replaced by a '+'. For example,

$$n_{ijk+} = \sum_{l=1}^{u} n_{ijkl}$$

$$n_{ij++} = \sum_{k=1}^{t} \sum_{l=1}^{u} n_{ijkl}$$

and $$n_{++++} = \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \sum_{l=1}^{u} n_{ijkl} = N$$

The sums of elementary cell counts can be arranged in tables of non-elementary cells. These tables have fewer dimensions than the array of elementary cells, and are known as configurations, denoted by the letter C. For example, the three dimensional array (n_{ijk+}) obtained by summing over the fourth variable is the configuration C_{123}; C_{24} consists of cell counts (n_{+j+l}).

Let $m_{ijkl} = E (n_{ijkl})$. m_{ijkl} is the theoretical value postulated by a particular model and $\hat{m}_{ijkl}$ is the

maximum likelihood estimate of this theoretical value.

The mathematical model known as the "Log-linear Model" for expected cell frequencies which is described below was introduced by Darroch (1962), further amplified by Birch [3], Bishop [4], and Bishop & Mosteller [5].

Let $e_{ijkl} = \log (m_{ijkl})$, where log refers to the natural logarithm throughout (for simplicity, assume $m_{ijkl} > 0$). Analogous to the analysis of variance, using the notation of Birch [3], decompose e_{ijkl} as follows :

$$
\begin{aligned}
e_{ijkl} &= \log (m_{ijkl}) \\
&= u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + \\
&\quad u_{12(ij)} + u_{13(ik)} + u_{14(il)} + u_{23(jk)} + \\
&\quad u_{24(jl)} + u_{34(kl)} + u_{123(ijk)} + \\
&\quad u_{124(ijl)} + u_{134(ikl)} + u_{234(jkl)} + \\
&\quad u_{1234(ijkl)}
\end{aligned} \qquad (1)
$$

where $\sum_{i=1}^{r} u_{1(i)} = 0, \ldots.$

$$\sum_{i=1}^{r} u_{12(ij)} = \sum_{j=1}^{s} u_{12(ij)} = 0 , \ldots.$$

$$\sum_{i=1}^{r} u_{123(ijk)} = \sum_{j=1}^{s} u_{123(ijk)} = \sum_{k=1}^{t} u_{123(ijk)} = 0 , \ldots.$$

$$\sum_{i=1}^{r} u_{1234(ijkl)} = \sum_{j=1}^{s} u_{1234(ijkl)} =$$

$$\sum_{k=1}^{t} u_{1234(ijkl)} = \sum_{l=1}^{u} u_{1234(ijkl)} = 0$$

The u's represent the possible "effects" of the four variable on e_{ijkl}. The main effects are $u_{1(i)}$, $u_{2(j)}$, $u_{3(k)}$ and $u_{4(l)}$; the interaction effects are $u_{12(ij)}$, ... , $u_{123(ijk)}$, ... , $u_{1234(ijkl)}$. The parameter u satisfies the condition that

$$\{\exp u\}\{\sum_{i,j,k,l} \exp (u_{1(i)} + ... + u_{1234(ijkl)})\}$$

$$= 0$$

Using this notation, Birch has shown that models corresponding to different hypothesis are defined by omitting one or more terms from (1). For instance, postulating the hypothesis that there is no four-factor effect, then $u_{1234(ijkl)} = 0$ for all i, j, k, l, or more briefly, $u_{1234} = 0$.

The principle of hierarchy states that if any u-term is set equal to zero, all its higher relatives must also be set equal to zero. Conversely, if any u-term is non-zero, its lower order relatives must be present in the log-linear model. Thus if $u_{12} = 0$, then $u_{123} = u_{124} =$

$u_{1234} = 0$. On the other hand, if u_{123} is present in the model, u_{12}, u_{13}, u_{23}, u_1, u_2, u_3 must also be present. In this context, only hierarchical models are considered in subsequent analysis.

For hierarchical models, the minimal sufficient statistics always consist of a set of non-redundant configurations. The non-elementary cell frequencies of these configuration are unbiased estimates of their expected values under the particular model postulated. Estimates of every elementary cell of the whole contingency table can be obtained from them. For example, for the model where $u_{1234} = u_{234} = 0$, the minimal sufficient configurations are C_{123}, C_{124} and C_{134}. The model is represented by [123/124/134]. The maximum likelihood estimates of m_{ijk+}, m_{ij+l}, m_{i+kl} are

$$\hat{m}_{ijk+} = n_{ijk+}, \quad \hat{m}_{ij+l} = n_{ij+l}, \quad \hat{m}_{i+kl} = n_{i+kl}$$

This result, shown by Birch, is true not only for the multinomial obtained when N is fixed, but also true for product multinomial obtained when one or more configurations are fixed, provided that the u-terms corresponding to the fixed configurations are included in the model.

Bhapkar [1], then later Bhapkar & Koch [2], and

Bishop [4] have emphasized the difference between "Factor" variables that classify the unit of observation according to " a description of the subpopulation of units to which it belongs ", and "Response" variables that classify according to " a description of what happens to it during and/or after the experiment ". If variables 1, 2 and 3 are factors and variable 4 is the response, the first three variables determine the r x s x t sampling strata. In other words, C_{123} with entries (n_{ijk+}) is fixed by the sampling plan, and so only models retaining u_{123} are permissible. If the model with $u_{123} = u_{1234} = 0$ is fitted, then

$$\hat{m}_{ijk+} \neq n_{ijk+}$$

which contradicts the design of the experiment.

Unique set of maximum likelihood estimates for every elementary cell can be derived from the above mentioned sufficient statistics alone. For some models, cell estimates may be written directly as functions of the sufficient statistics. When direct estimates do not exist, iterative procedures must be used. Bishop, Fienberg & Holland [6] gave detailed accounts of distinguishing these two classes of models. Only a few general rules are given here.

Direct estimates exist if the set of minimal

sufficient statistics for a model consists of only two configurations, at least one two-factor effect must be absent. The general form of direct estimates is such that the numerator has entries from each sufficient configuration; the denominator, entries from redundant configurations caused by overlapping; terms in powers of N appear either in the numerator or in the denominator to ensure the right order of magnitude. Consider the model with sufficient configurations C_{123}, C_{124}. The direct cell estimate is given by

$$\hat{m}_{ijkl} = \frac{n_{ijk+}\, n_{ij+l}}{n_{ij++}}$$

Two overlapping configurations are said to be linked to each other. When each pair of configurations are linked, a closed loop is formed, for example, C_{12}, C_{13} and C_{23}. In any dimension, iterative methods are necessary when such closed loop exist, whether all or only some of the minimal set of sufficient configurations are involved. The method of iterative proportional fitting is such that it always converges to the required unique set of maximum likelihood estimates, and to any desired degree in the elementary cell estimates. Any set of starting values may be used that conforms to the model being fitted. When direct estimates exist, the procedure yields the exact

estimates in one cycle. Consider the model with $u_{123} = u_{124} = u_{1234} = o$. The sufficient configurations are C_{134}, C_{234} and C_{12}, and the model is represented by [134/234/12]. The three configurations are linked to each other and a closed loop is formed. To fit the elementary cells iteratively,

$$\hat{m}^{0}_{ijkl} = 1$$

is commonly adopted. The preliminary estimates are adjusted to fit C_{134}, C_{234} and C_{12} successively.

Fitting to C_{134}

$$\hat{m}^{1}_{ijkl} = \hat{m}^{0}_{ijkl} \frac{n_{i+kl}}{\hat{m}^{0}_{i+kl}}$$

Fitting to C_{234}

$$\hat{m}^{2}_{ijkl} = \hat{m}^{1}_{ijkl} \frac{n_{+jkl}}{\hat{m}^{1}_{+jkl}}$$

Fitting to C_{12}

$$\hat{m}^{3}_{ijkl} = \hat{m}^{2}_{ijkl} \frac{n_{ij++}}{\hat{m}^{2}_{ij++}}$$

A quantity δ is predeterminded such that the three-step cycle is repeated until a complete cycle does not cause any cell to change by more than this amount, i.e.

$$\left| \hat{m}_{ijkl}^{3\alpha} - \hat{m}_{ijkl}^{3\alpha-3} \right| < \delta \qquad \text{for all } i, j, k, l$$

After estimating the elementary cells, either directly or by iterative method, the parameters of the model can by found.

$$u = \frac{1}{rstu} \sum_{i,j,k,l} e_{ijkl}$$

$$u_{1(i)} = \frac{1}{stu} \sum_{j,k,l} e_{ijkl} - u$$

$$u_{12(ij)} = \frac{1}{tu} \sum_{k,l} e_{ijkl} - u - u_{1(i)} - u_{2(j)}$$

et. al.

Goodman [13] recommended replacing n_{ijkl} by $n_{ijkl} + \frac{1}{2}$ for both analysis and estimation. This adjustment of the n's reduces both the asymptotic bias and mean square error of the u's.

Models may be fitted to help understanding complex data in the table, or it may be used to test certain hypothesis. More accurate estimates of "expected" frequencies are obtained from models than the original data by themselves. So it is not the optimal model that is of interest, but one which fits the data adequately and provides more stable cell estimates. It is still acceptable if the number of cells in the configuration to be fitted is kept as small

as possible. The goodness of fit for the hierarchical model H with estimates m_{ijkl} can be calculated in two ways. The classical one is the Pearson chi-square test statistic,

$$\chi^2(H) = \sum_{i,j,k,l} \frac{(n_{ijkl} - m_{ijkl})^2}{m_{ijkl}}$$

The second goodness-of-fit test based on likelihood ratio citerion, is the likelihood ratio chi-square test statistic

$$\chi^2(H) = 2 \sum_{i,j,k,l} n_{ijkl} \log \left(\frac{n_{ijkl}}{\hat{m}_{ijkl}} \right)$$

Both of these two statistic, under the null hypothesis H, are distributed asymptotically to a chi-square distribution with degrees of freedom equal to the number of cells minus the number of parameters fitted in the model.

A virtue of the likelihood ratio chi-square statistic is that it can be "partitioned". Let H and H' be two hierarchical models with cell estimates $\hat{m}_{ijkl}$ and $\hat{m}'_{ijkl}$. The set of parameters assumed to be zero under H', ζ', is a subset of those under H, ζ. ζ^* is the set of parameters that are in ζ but not in ζ', and H* is the hypothesis that the u's in ζ^* are zero. The hypothesis that H* is true assuming H' is true (H*|H') can be tested by the statistic

$$\chi^2(H^*|H') = \chi^2(H) - \chi^2(H')$$

$$= 2 \sum_{i,j,k,l} n_{ijkl} \log \left(\frac{\hat{m}_{ijkl}}{\hat{m}_{ijkl}} \right)$$

$$= 2 \sum_{i,j,k,l} m_{ijkl} \log \left(\frac{\hat{m}_{ijkl}}{\hat{m}_{ijkl}} \right)$$

This statistics is also asymptotically distributed as a chi-square with degrees of freedom equal to the number of u's in ζ^* if H* is true. By such a method the significance of a particular effect u_θ can be tested, i.e.

$$H^* : u_\theta = 0$$

by letting ζ^* consist of a single parameter u_θ only.

The test that the marginal k-factor (k = 2, 3, 4) interaction is zero is defined as a test that the k-factor interaction is zero in the k-dimensional marginal subtable indexed by those k factors. For instance, for the 2-factor interaction u_{23}, obtain the marginal subtable C_{23} by collapsing over variables 1 and 4. Test either the fit of the model [2/3] applied to this collapsed table, or the difference in the tests of fit of the two models [23] and [2/3] applied to the original table,

The test that the partial k-factor effect is zero is the difference between tests-of-fit of two hierarchical models which differ only by that k-factor effect, and such that all factors not among those k-factors are treated

symmetrically. A reasonable choice is the full k-order model and the k-order model excluding the k-factor effect to be tested. Thus for u_{23}, the test for partial association is the difference between the tests-of-fit for the two models [12/13/14/23/24/34] and [12/13/14/24/34].

Under the null hypothesis that u_θ equal to zero, both the partial association and marginal association test statistics are asymptotically distributed as chi-square with degrees of freedom equal to those associated with the parameter. These two tests indicate clearly the relative magnitude of the difference which is likely to be found when the parameter u_θ is added or deleted from the model. Hence if both statistics are large, the k-factor interaction is significant and is required in the model. If both are small, the interaction is not needed. The rest are uncertain and need to be tested. Thus effects can be screened according to their importance, which makes the task of selecting effects to be put into models much simpler.

VII. APPENDIX II : TABLES

Age (1)	Sex (2)	Alcohol (3)	Factors (4)	Duration (5)	Schedule (6)	Dose (7)	FM (8)	NV (9)	PI (10)	Necrosis (11)	Fibrosis (12)	Cirrhosis (13)
3	2	1	2	2	4	2	2	2	2	2	1	1
3	2	2	2	2	4	2	2	2	1	2	1	1
2	2	1	2	1	2	1	2	2	2	1	1	1
2	2	1	2	3	4	3	2	2	2	1	2	1
3	1	2	1	1	4	1	1	1	1	2	1	1
2	1	2	1	2	4	2	1	1	1	1	1	1
3	1	2	2	3	2	2	1	1	1	1	1	1
3	1	2	1	2	4	2	1	1	1	1	1	1
3	1	2	1	3	4	3	1	1	1	1	1	1
3	1	4	2	2	4	2	2	1	2	1	1	1
3	2	2	2	3	4	1	1	1	1	1	1	1
3	2	2	2	3	4	2	1	1	2	1	1	1
1	0	0	1	3	4	3	2	2	2	2	2	2
2	2	2	2	1	4	2	1	1	1	1	1	1
3	1	2	2	3	4	2	1	1	2	2	2	1
3	1	2	1	2	4	2	2	1	1	1	2	2
3	1	2	1	2	4	2	1	1	1	1	1	1
2	2	2	2	1	4	2	1	1	1	1	2	1
3	1	3	1	2	4	2	2	2	1	1	1	1
1	1	1	1	3	4	2	1	2	1	1	2	1
2	2	2	1	2	3	3	1	2	1	2	1	1
2	1	0	1	3	3	3	2	2	2	1	1	1

(Cont'd)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
2	1	4	2	3	2	1	1	1	1	1	1	1
2	1	4	1	1	3	1	1	1	1	1	1	1
3	2	2	2	2	3	2	1	1	1	1	1	1
2	1	3	1	1	3	1	1	1	1	1	1	1
1	1	3	2	1	3	1	1	1	2	1	1	1
3	2	1	1	3	4	3	2	1	1	1	1	1
3	2	2	1	1	3	1	1	2	1	1	1	1
2	1	3	1	3	2	2	1	1	1	1	1	1
2	2	2	1	1	3	3	1	1	1	1	1	1
2	1	2	1	1	2	2	1	1	1	1	1	2
2	1	2	2	2	4	2	1	1	1	2	2	1
3	2	2	1	1	2	1	2	2	1	1	1	1
3	2	2	1	2	2	1	1	2	1	1	1	1
3	1	2	1	3	2	2	2	2	2	2	2	1
2	2	2	2	3	2	2	2	2	2	1	1	1
1	2	2	2	3	2	2	1	1	1	2	1	1
2	1	4	1	3	2	2	1	2	1	2	1	1
3	2	4	2	3	1	3	2	1	2	1	2	2
3	2	2	1	3	2	2	2	2	2	1	2	1
3	1	3	2	3	2	3	1	1	1	1	1	1
1	1	1	1	1	4	2	2	2	2	2	1	1
1	1	1	1	1	4	1	1	2	2	1	1	1
3	2	2	1	3	4	2	2	2	2	2	1	1
1	1	1	1	2	2	2	1	1	1	1	1	1
1	1	0	1	2	2	2	1	1	1	1	1	1
1	1	4	1	1	3	1	2	2	1	1	2	1
3	1	4	2	3	4	3	2	1	1	2	1	1
1	1	3	1	2	4	1	1	1	1	1	1	1

(Cont'd)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1	1	1	1	2	4	2	1	1	1	1	1	1
2	2	4	2	3	3	3	2	1	1	1	2	2
1	2	1	2	3	2	3	2	2	2	1	2	1
2	1	4	2	2	2	1	1	1	1	2	1	1
1	2	0	1	1	2	2	1	2	1	1	1	1
2	1	1	1	3	2	2	1	1	1	1	1	1
2	1	2	1	1	2	1	1	2	1	1	1	1
2	1	4	2	3	4	3	2	1	1	2	1	1
1	1	1	1	2	4	2	2	2	1	1	2	1
3	1	3	1	2	4	1	1	1	1	1	1	1
3	1	3	1	2	3	2	1	1	2	1	1	1
3	2	2	1	2	2	2	2	1	1	1	1	1
2	1	3	1	2	4	1	2	1	1	1	1	1
2	2	2	1	3	2	3	2	2	2	1	2	1
2	2	2	1	2	2	2	1	1	2	2	1	1
2	1	4	1	2	4	1	1	2	1	1	1	1
2	1	2	1	2	2	2	2	2	1	1	1	1
3	2	1	1	2	4	2	1	1	1	1	1	1
2	2	2	2	2	4	2	1	1	1	2	1	1
3	1	2	1	1	4	1	1	1	1	1	1	0
3	2	2	1	2	0	2	2	1	1	1	1	1
1	1	1	2	2	4	2	2	1	1	1	1	1
2	1	2	2	2	4	2	1	1	1	1	1	1
3	1	2	1	2	4	2	2	1	1	1	1	1
3	2	2	2	3	4	3	2	2	2	2	2	1
2	2	2	1	2	4	1	1	1	1	2	2	1
2	2	2	2	2	4	2	1	1	1	1	1	1
2	2	2	2	2	4	2	1	1	2	2	2	1

(Cont'd)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
3	2	2	2	2	4	2	1	1	1	1	1	1
2	2	2	2	2	4	1	2	2	2	1	1	1
3	2	1	2	2	4	2	2	2	1	1	1	1
2	1	2	1	2	4	2	2	2	1	1	1	1
3	2	2	2	2	4	2	1	1	1	2	2	1
2	1	4	1	2	4	2	1	1	1	1	1	1
3	2	1	2	2	4	2	1	1	1	1	1	1
3	2	2	2	2	4	2	2	2	2	2	1	1
3	1	4	2	1	4	1	1	1	2	1	1	1
2	2	2	2	2	4	2	1	1	1	1	1	1
3	1	2	2	2	4	1	1	1	2	2	1	1
1	2	1	2	3	1	2	1	2	2	1	2	2
2	1	2	1	1	2	1	1	1	1	1	1	1
2	2	1	1	1	2	1	1	2	1	1	1	1
2	2	1	1	2	2	2	1	2	1	1	1	1
2	2	1	1	1	4	1	1	2	1	1	1	1
2	2	1	2	1	2	1	1	2	1	1	1	1
2	2	1	1	1	2	1	1	2	1	1	1	1
2	1	1	2	1	2	1	2	1	1	1	1	1
1	2	1	2	1	2	1	1	2	1	1	1	1
2	1	1	2	1	2	2	2	1	1	1	1	1
3	1	3	1	3	2	2	1	2	2	1	1	2
2	2	1	1	3	2	2	1	1	1	1	1	1
2	1	1	1	3	2	3	2	1	2	1	1	1

Table 1 : Raw Data of the 102 Patients in the Prospective Study

1	Less than 25 years old	17
2	25 to 50 years old	45
3	More than 50 years old	40
	Total	102

Table 2 : Marginal Totals of (1) - AGE
Methotrexate Therapy Started

1	Male	53
2	Female	48
	Total	101

Table 3 : Marginal Totals of (2) - SEX

1	Non-drinker	27
2	1 to 3 drinks per week	48
3	1 to 3 drinks per day	10
4	More than 3 drinks per day	13
	Total	98

Table 4 : Marginal Totals of (3) - ALCOHOL intake

1	No	58
2	Yes	44
	Total	102

Table 5 : Marginal Totals of (4) - Presence of Other FACTORS

1	Less than 1 year	26
2	1 to 2 years	45
3	More than 2 years	31
	Total	102

Table 6 : Marginal Totals of (5) - Cumulative DURATION of Methotrexate Therapy

1	Daily oral	2
2	Weekly oral single dose	35
3	Weekly intramuscular or intravenous single dose	11
4	Weekly oral of divided dose	53
	Total	101

Table 7 : Marginal Totals of (6) - Predominant Dose SCHEDULE

1	Less than 500 mg	30
2	500 to 1500 mg	56
3	More than 1500 mg	16
	Total	102

Table 8 : Marginal Totals of (7) - Cumulative DOSE of Methotrexate taken

1	No Change	64
2	Worsen	38
	Total	102

Table 9 : Marginal Totals of (8) - FATTY METAMORPHOSIS

1	No Change	62
2	Worsen	40
	Total	102

Table 10 : Marginal Totals of (9) - NUCLEAR VARIABILITY

1	No Change	73
2	Worsen	29
	Total	102

Table 11 : Marginal Totals of (10) - PERIPORTAL INFLAMMATION

1	No Change	79
2	Worsen	23
	Total	102

Table 12 : Marginal Totals of (11) - NECROSIS

1	No Change	82
2	Worsen	20
	Total	102

Table 13 : Marginal Totals of (12) - FIBROSIS

1	No Change	94
2	Worsen	7
	Total	101

Table 14 : Marginal Totals of (13) - CIRRHOSIS

In the following tables, 1 denotes None
2 denotes Slight
3 denotes Moderate
4 denotes Marked

1st Post-MTX

Pre-MTX \ 1st Post-MTX	1	2	3	4	
1	36	16	4	1	
2	2	21	12	0	
3	0	4	4	1	
4	0	0	1	0	
					102

2nd Post-MTX

1st Post MTX \ 2nd Post-MTX	1	2	3	4	
1	7	1	1	0	
2	1	9	3	0	
3	0	1	4	1	
4	0	0	0	0	
					28

3rd Post-MTX

2nd Post MTX \ 3rd Post-MTX	1	2	3	4	
1	1	2	0	0	
2	1	3	1	0	
3	0	0	2	0	
4	0	0	0	0	
					10

4th Post-MTX

3rd Post MTX \ 4th Post-MTX	1	2	3	4	
1	0	0	1	0	
2	0	0	0	0	
3	0	0	1	0	
4	0	0	0	0	
					2

Table 15 : Post-MTX Biopsies vs Pre-MTX Biopsy on Fatty Metamorphosis

Pre-MTX \ Post-MTX	1	2	3	4	
1	33	17	6	1	
2	3	19	12	1	
3	0	2	6	1	
4	0	0	1	0	
					102

Table 16 : Last Post-MTX Biopsy vs Pre-MTX Biopsy on Fatty Metamorphosis

1st Post-MTX

Pre-MTX \ 1st Post-MTX	1	2	3	4	
1	15	22	2	0	
2	5	38	9	1	
3	1	5	3	0	
4	0	1	0	0	
					102

2nd Post-MTX

1st Post-MTX \ 2nd Post-MTX	1	2	3	4	
1	4	1	1	0	
2	2	12	2	2	
3	0	1	3	0	
4	0	0	0	0	
					28

3rd Post-MTX

2nd Post-MTX \ 3rd Post-MTX	1	2	3	4	
1	1	1	0	0	
2	2	3	0	1	
3	0	1	1	0	
4	0	0	0	0	
					10

4th Post-MTX

3rd Post-MTX \ 4th Post-MTX	1	2	3	4	
1	0	0	0	0	
2	0	1	1	0	
3	0	0	0	0	
4	0	0	0	0	
					2

Table 17 ; Post-MTX Biopsies vs Pre-MTX Biopsy on Nuclear Variability

Pre-MTX \ Post-MTX	1	2	3	4
1	13	22	3	1
2	7	33	11	2
3	2	3	3	1
4	0	1	0	0

102

Table 18 : Last Post-MTX Biopsy vs Pre-MTX Biopsy on Nuclear Variability

1st Post-MTX

Pre-MTX \ 1st Post-MTX	1	2	3	4
1	60	16	5	0
2	6	9	4	0
3	0	1	1	0
4	0	0	0	0

102

2nd Post-MTX

1st Post-MTX \ 2nd Post-MTX	1	2	3	4
1	15	2	0	0
2	2	5	1	0
3	0	2	1	0
4	0	0	0	0

28

3rd Post-MTX

2nd Post-MTX \ 3rd Post-MTX	1	2	3	4
1	3	3	0	0
2	1	0	1	0
3	0	1	1	0
4	0	0	0	0

10

4th Post-MTX

3rd Post-MTX \ 4th Post-MTX	1	2	3	4
1	0	0	1	0
2	1	0	0	0
3	0	0	0	0
4	0	0	0	0

2

Table 19 : Post-MTX Biopsies vs Pre-MTX Biopsy on Periportal Inflammation

Pre-MTX \ Post-MTX	1	2	3	4
1	57	18	6	0
2	6	8	5	0
3	0	1	1	0
4	0	0	0	0

102

Table 20 : Last Post-MTX Biopsy vs Pre-MTX Biopsy on Periportal Inflammation

1st Post-MTX

Pre-MTX \ 1st Post-MTX	1	2	3	4
1	51	18	0	0
2	11	17	0	0
3	2	3	0	0
4	0	0	0	0

102

2nd Post-MTX

1st Post-MTX \ 2nd Post-MTX	1	2	3	4
1	10	5	0	0
2	4	7	1	0
3	0	1	0	0
4	0	0	0	0

28

3rd Post-MTX

2nd Post-MTX \ 3rd Post-MTX	1	2	3	4
1	3	4	0	0
2	0	3	0	0
3	0	0	0	0
4	0	0	0	0

10

4th Post-MTX

3rd Post-MTX \ 4th Post-MTX	1	2	3	4
1	0	0	0	0
2	1	0	1	0
3	0	0	0	0
4	0	0	0	0

2

Table 21 : Post-MTX Biopsies vs Pre-MTX Biopsy on Necrosis

Pre-MTX \ Post-MTX	1	2	3	4
1	46	21	2	0
2	10	18	0	0
3	2	3	0	0
4	0	0	0	0

102

Table 22 : Last Post-MTX Biopsy vs Pre-MTX Biopsy on Necrosis

1st Post-MTX

Pre-MTX \ 1st Post-MTX	1	2	3	4
1	77	14	2	0
2	2	3	2	0
3	0	0	2	0
4	0	0	0	0

102

2nd Post-MTX

1st Post-MTX \ 2nd Post-MTX	1	2	3	4
1	18	5	1	1
2	0	2	0	1
3	0	0	0	0
4	0	0	0	0

28

3rd Post-MTX

2nd Post-MTX \ 3rd Post-MTX	1	2	3	4
1	6	0	0	0
2	1	1	0	0
3	0	0	0	0
4	0	0	1	1

10

4th Post-MTX

3rd Post-MTX \ 4th Post-MTX	1	2	3	4
1	1	1	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

2

Table 23 : Post-MTX Biopsies vs Pre-MTX Biopsy on Fibrosis

Pre-MTX \ Post-MTX	1	2	3	4
1	72	18	3	0
2	2	3	1	1
3	0	0	2	0
4	0	0	0	0

102

Table 24 : Last Post-MTX Biopsy vs Pre-MTX Biopsy on Fibrosis

In the following tables, 1 denotes Absence (No)

2 denotes Presence (Yes)

1st Post-MTX

Pre-MTX \ 1st Post-MTX	1	2
1	95	6
2	0	1

102

2nd Post-MTX

1st Post MTX \ 2nd Post-MTX	1	2
1	24	2
2	1	1

28

3rd Post-MTX

2nd Post MTX \ 3rd Post-MTX	1	2
1	9	0
2	0	1

10

4th Post-MTX

3rd Post MTX \ 4th Post-MTX	1	2
1	1	1
2	0	0

2

Table 25 ; Post-MTX Biopsies vs Pre-MTX Biopsy on Cirrhosis

		Post-MTX 1	Post-MTX 2
Pre-MTX	1	94	7
	2	0	1

102

Table 26 : Last Post-MTX Biopsy vs Pre-MTX Biopsy on Cirrhosis

APPENDIX III -: DATA COLLECTION FORM
Retrospective and Prospective Liver-MTX Study

Do not write in this space
1 [ ][ ] [ ][ ][ ]
HOSP. PATIENT

1. Patient code number: _____

2. Age MTX started: [ ][ ] col.7-8

col.9
3. Sex: male [ ] 1
female [ ] 2

col.10
4. Alcohol intake: non-drinker [ ] 1
1-3 drinks per week [ ] 2
1-3 drinks per day [ ] 3
4 or more drinks per day [ ] 4

5. Duration of psoriasis (years): [ ][ ] col.11-12

6. Extent of psoriasis (% of body involved when considered for MTX, rule of nines): [ ][ ][ ] col.13-15

	1 Yes	2 No	
7. History of previous liver disease: hepatitis	[ ]	[ ]	col.16
jaundice	[ ]	[ ]	col.17
gall bladder	[ ]	[ ]	col.18
other (specify _____)	[ ]	[ ]	col.19
8. History of other disease: diabetes	[ ]	[ ]	col.20
obesity	[ ]	[ ]	col.21

9. Possible contributing factors to liver disease:

blood transfusions	[ ]	[ ]	col.22
hepatotoxins (specify _____)	[ ]	[ ]	col.23
10. Systemic drugs for psoriasis: corticosteroids	[ ]	[ ]	col.24
antimetabolites (specify _____)	[ ]	[ ]	col.25

11. Date MTX started: month [ ][ ] col.26-27
year [ ][ ] col.28-29

Retrospective and Prospective Liver-MTX Study

Do not write in this space

2 [ ][ ] [ ][ ][ ] [ ][ ][ ]
HOSP. PATIENT SLIDE

1. Patient code number: _____

2. Date of examination:
   month [ ][ ] col.10-11
   year [ ][ ] col.12-13

3. BSP (abs. no.) [ ][ ] col.14-15

	1 Nor.	2 Abn.	Units	
4. SGOT	[ ]	[ ]	[ ][ ][ ].[ ]	col.16-20
5. SGPT	[ ]	[ ]	[ ][ ][ ].[ ]	col.21-25
6. Alk. phos	[ ]	[ ]	[ ][ ][ ].[ ]	col.26-30

7. Physical findings:

	1 Yes	2 No	
significant hepatomegaly	[ ]	[ ]	col.31
jaundice	[ ]	[ ]	col.32
ascites	[ ]	[ ]	col.33
spider angioma	[ ]	[ ]	col.34
esophageal varices	[ ]	[ ]	col.35

8. Cumulative duration of MTX (years) [ ][ ].[ ] col.36-38

IF NO MTX HAS BEEN GIVEN, SKIP TO QUESTION 13.

9. % of time treated [ ][ ][ ] col.39-41

10. Predominant dose schedule: col.42

daily oral	[ ] 1
weekly oral single dose	[ ] 2
weekly intramuscular or IV single dose	[ ] 3
weekly 3-4 doses /36 hours p.o.	[ ] 4

11. Total cumulative dose (mg.) [ ][ ][ ][ ] col.43-46

12. Time elapsed between last dose of MTX and liver function studies (days) [ ][ ][ ] col.47-49

13. IF NO BIOPSY WAS DONE ON THIS DATE, SKIP THE REST OF THE FORM

Identification on slide _____

14. Pathology report:

	1 None	2 Slight	3 Mod.	4 Marked	
fatty metamorphosis	[ ]	[ ]	[ ]	[ ]	col.50
nuclear variability	[ ]	[ ]	[ ]	[ ]	col.51
periportal inflammation	[ ]	[ ]	[ ]	[ ]	col.52
necrosis	[ ]	[ ]	[ ]	[ ]	col.53
fibrosis	[ ]	[ ]	[ ]	[ ]	col.54
_____	[ ]	[ ]	[ ]	[ ]	col.55
_____	[ ]	[ ]	[ ]	[ ]	col.56

	Yes	No	
cirrhosis	[ ]	[ ]	col.57

IX. REFERENCES

[1] Bhapkar, V.P. (1961). *Some Tests for Categorical Data*. Ann. Math. Statist. 32, 72 - 83.

[2] Bhapkar, V.P. and Koch, G.G. (1968). *Hypothesis of No Interaction in Multidimensional Contingency Tables*. Technometrics 10, 107 - 124.

[3] Birch, M.W. (1963). *Maximum Likelihood in Three-way Contingency Tables*. Journal of Royal Statistical Society B25, 220 - 233.

[4] Bishop, Y.M.M. (1969). *Full Contingency Tables, Logits and Split Contingency Tables*. Biometrics Vol. 25, No. 2, 383 - 399.

[5] Bishop, Y.M.M. and Mosteller, F. (1969). *Smoothed Contingency Table Analysis*. Chap. IV-3 of The National Halethane Study, U.S. Govt. Printing Office, Washington, D.C., 238 - 272.

[6] Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis : Theory and Practice*. The MIT Press, Cambridge, Mass.

[7] Brown, M.B. (1976). *Screening Effects in Multi-dimensional Contingency Tables*. Applied Statistics Vol. 25, 37 - 46.

[8] Dahl, M,G,C,, Gregory, M,M, and Schener, P,J, (1972) *Methotrexate Hepatotoxicity in Psoriasis - Comparison of Different Dose Regimens*. British Medical Journal 11, 654 - 646.

[9] Dixon, W.J. and Brown, M.B. (1977). *Biomedical Computer Programs P Series*. University of California Press, Berkeley, Los Angeles, London.

[10] Draper, N.R. and Smith, H. (1966). *Applied Regression Analysis*. John Wiley and Son, Inc.

[11] Fienberg, S.E. (1970). *The Analysis of Multidimensional Contingency Tables*. Ecology, Vol. 51, No. 2, 419 - 433.

[12] Fleiss, J.L. (1973). *Statistical Methods for Rates and Proportions*. John Wiley and Sons, Inc. Toronto.

[13] Goodman, L.A. (1970). *The Multivariate Analysis of Qualitative Data : Interactions Among Multiple Classifications*. Journal of American Statistical Association, 65, 226 - 256.

[14] Goodman, L.A. (1971). *The Analysis of Multidimensional Contingency Tables : Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications*. Technometrics, Vol 13, No. 1, 33 - 61.

[15] Goodman, L.A. (1971). *Partitioning of Chi-square, Analysis of Marginal Contingency Tables and Estimation of Expected Frequencies in Multidimensional Contingency Tables.* Journal of American Statistical Association, 66, 339 - 344.

[16] Haberman, S.J. (1974). *The Analysis of Frequency Data.* Chicago Press, Chicago.

[17] Ku, H.H., Varner, R. and Kullback, S. (1971). *On the Analysis of Multidimensional Contingency Tables.* Journal of American Statistical Association, 66, 55 - 64.

[18] Kullback, S. and Fisher, M. (1973). *Partitioning Second Order Interaction in Three-way Contingency Tables.* Applied Statistics, 22, 172 - 184.

[19] Ku, H.H. and Kullback, S. (1974). *Log-linear Models in Contingency Table Analysis.* American Statistician, Vol. 28, No. 4, 115 - 122.

[20] Maxwell, A.E. (1961). *Analysing Qualitative Data.* Methuen and Company, Ltd., London, England.

[21] Nie, H.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K. and Bent, D.H. (1975). *Statistical Package for the Social Sciences.* McGraw-Hill Book Company.

[22] Podurgiel, B.J., et. al. (1973). *Liver Injury Associated with Methotrexate Therapy for Psoriasis.* Mayo Clinic Proceedings, Vol. 48, 787 - 792.

[23] Reese, L., Grisham, J., Aach, R. and Eisen, A. (1974). *Effects of Methotrexate on the Liver in Psoriasis.* Journal of Investigative Dermatology, Vol. 62, No. 6, 597 - 602.

[24] Weinstein, G. et. al. (1973). *Psoriasis - Liver - Methotrexate Interactions.* Archives of Dermatology, Vol. 108, 36 - 42.

[25] Dixon, W.J., and Massey, F.J. *Introduction to Statistical Analysis.* Third Edition, 1969. Mcgraw Hill, New Yor.