Published in final edited form as:

Contemp Clin Trials. 2009 January ; 30(1): 2–7. doi:10.1016/j.cct.2008.10.001.

Modeling using baseline characteristics in a small multicenter clinical trial for Barrett's esophagus

Albert O. Shar^{a,*}, Marie A. Gaudard^b, Elisabeth I. Heath^c, Arlene A. Forastiere^d, Vincent W. Yang^e, and Stephen J. Sontag^f

^aThe Robert Wood Johnson Foundation, Princeton, NJ, USA

^bNorth Haven Group, Inverness, FL, USA

^cBarbara Ann Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI, USA

^dSidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, USA

^eEmory University School of Medicine, Atlanta GA, USA

^fHines VA Hospital, Hines IL, USA

Abstract

Objective—Utilizing data obtained during a multicenter investigation, this paper illustrates how the use of covariates and careful modeling techniques can be useful in assessing whether a negative outcome from a small multicenter clinical trial could be due to imbalance in baseline characteristics. The Chemoprevention for Barrett's Esophagus Trial (CBET) was a phase IIb, multicenter, randomized, placebo-controlled trial of celecoxib in patients with Barrett's esophagus. The primary outcomes for the original study were the proportion of biopsy samples exhibiting dysplasia in the celecoxib and placebo groups. The secondary and tertiary outcomes included histologic change and measurements of biologically relevant markers, including COX-1 and -2 mRNA, prostanoic levels, and methylation of tumor suppressor genes p16, APC, and E-cadherin. The original study reported no significant differences in primary, secondary or tertiary outcomes. In this paper, we focus on the results of one of the secondary measures, quantitative endoscopy (QE).

Design—The study utilizes data from 56 patients in the CBET for whom baseline (BL) QE and one-year follow-up QE (F04) studies were performed. Of these, 29 were treated with celecoxib (200 mg twice daily for a minimum of 48 weeks) and 27 received the placebo. These patients are segmented as to the presence or absence of circumferential, tongues or islands of Barrett's.

Measurements—The response of interest is total affected area at one year (Total F04); affected area at baseline (Total BL) is used as a covariate.

Results—Controlling for complexity and clinic, there is a significant treatment effect. In addition, there is significant evidence that the area of Barrett's involvement decreased for patients in the treatment group.

© 2008 Elsevier Inc. All rights reserved.

*Corresponding author. The Robert Wood Johnson Foundation, Route 1 & College Road East, PO Box 2316, Princeton, NJ 08543-2316, United States. Tel.: +1 609 627 5873; fax: +1 609 419 8333. ashar@rwjf.org.

Appendix A. Supplementary data Supplementary data associated with this article can be found, in the online version, at doi: 10.1016/j.cct.2008.10.001.

Conclusions—That there was a decrease for the celecoxib over the placebo group adds to the body of evidence that relates COX-2 specific inhibitors and cancer incidence.

Keywords

Barrett's esophagus; Covariate modeling; Chemoprevention; Quantitative endoscopy

1. Background

Esophageal adenocarcinoma is a disease rapidly rising in incidence and Barrett's esophagus is a premalignant condition in which normal squamous epithelium of the esophagus is replaced by specialized columnar mucosa. It occurs as a result of chronic gastroesophageal reflux and is associated with an increased risk of developing esophageal adenocarcinoma [1–5]. The five-year survival rate achieved from surgical resection of esophageal cancer is approximately 24%.

The Chemoprevention for Barrett's Esophagus Trial (CBET) was a phase IIb, multicenter, randomized, placebo-controlled trial of celecoxib in patients with Barrett's esophagus and low or high grade dysplasia. Descriptions of the trial and its outcomes have been documented [6,7]. In particular, the original analysis showed no significance in primary, secondary or tertiary outcomes. However, while the sample size (100 enrolled in the study) may have been powerful enough for the primary outcomes, there were only 58 patients for whom quantitative endoscopy (QE) area measurements were usable.

The purpose of this paper is to explore the effect of celecoxib on QE area in greater detail. Because of the smaller sample size available for QE area analysis, a test for a treatment effect that pools assignable sources of variation into the error term is not very powerful. In this paper, we use covariates in our analyses in an attempt to account for extraneous sources of variation. Our analysis illustrates how the use of covariates can help indicate whether a negative outcome from a small multicenter clinical trial may be due to imbalance in baseline characteristics.

1.1. Modeling considerations

In many clinical trial situations, there are auxiliary variables that may not be considered in analyzing the data. Sometimes, a treatment effect is so large that it is easily seen to be significant despite differences in these auxiliary measurements. However, there are situations where the variability that is conferred on the outcome measure by the auxiliary variables obscures what may be a real treatment effect. In such a case, it makes sense to account for the variation due to the auxiliary measures prior to testing for a treatment. This is the goal of analysis of covariance [8,9].

It is sometimes the case that the end measurement is dependent upon baseline characteristics in subtle ways. For example, in the case of a multicenter trial, there may be differences in outcomes that are attributable to the patient populations serviced by the specific clinics, or to other clinic-specific practices. In the case of a complex disease, there may be differences in outcomes that are attributable to a classification of the disease type or severity. Controlling for such auxiliary variables, as well as potential interaction effects, in modeling study outcomes can be critical, especially when the sample size is small.

By accounting for variation due to these auxiliary factors, a test for treatment effects can be much more sensitive than when that test is performed against the totality of background variation. In layman's terms, we are seeking a signal amid noise. If we can legitimately account for some of the noise, then the signal becomes easier to detect.

The statistical software utilized in our analysis is JMP® 7.0.2. The p -value used to define statistical significance is 0.05.

1.2. Quantitative endoscopy–CBET study

In the CBET, the surface area of Barrett's esophagus was measured with quantitative endoscopy utilizing an enhancement of the computer image analysis system [10–13]. The system (US Patent #7,011,625) transforms photographs of Barrett's esophagus into two-dimensional maps, and the surface area of Barrett's esophagus is calculated from the reconstructed images. The key steps in capturing the images are relatively simple and are described in [14]. The images in Fig. 1 illustrate the QE process.

Digitized images taken during the upper endoscopy procedure in selected clinics were sent to one expert study investigator for evaluation. The investigator was masked to the patient's treatment assignment. Standard procedure guidelines for imaging were used by all study gastroenterologists. The design called for quantitative endoscopy to be performed at baseline, at 6 months, at one year, and then yearly for up to 3 years. However, baseline QE was performed for 81 of the original study patients, and of these, only 58 had QE recorded at the one year follow-up interval.

The CBET measurements of total surface area affected by Barrett's esophagus obtained using quantitative endoscopy at baseline and at the one year follow-up (48 weeks of treatment) are used in the current study. In the original analysis, 58 patients were determined to have had QE measurements both at baseline and one year later. Based on a Wilcoxon rank sum test, the original analysis [6] showed no statistically significant differences by treatment assignment in the median change of total surface area affected by Barrett's esophagus as measured by quantitative endoscopy after one year (48 weeks of treatment) ($p=0.12$). Also, based on a Wilcoxon rank sum test, no statistically significant differences by grade of dysplasia were detected (low grade $p=0.17$, high grade $p=0.34$).

1.3. Analysis background

The current paper provides an alternative analysis of this data. Differences due to clinics are taken into account. Degree of dysplasia is replaced by a characterization of disease complexity, as defined below.

Although the original analysis included 58 patients with QE measurements at baseline and one year, only 56 patients are included in the current analysis, since validation of the raw data indicated that two of the 58 patient records used in the original analysis were in error. The CBET involved six different clinics; Table 1 gives the breakdown by clinic for our 56 patients. Note that 52 of the patients were treated in four clinics: CPMC (Columbia Presbyterian Medical Center), HVA (Hines VA Medical Center), JHH (Johns Hopkins Hospital) and UCLA (UCLA Medical Center). Only three patients were treated at UHC (University Health Center of Cleveland) and one at MAYO (Minnesota).

In [14], it is shown that there was a statistically significant difference in the patient complexity mix at the four primary CBET clinics (CPMC, HVA, JHH, and UCLA).

The Barrett's esophagus lesion is not a simple geometric shape. A description of possible shapes can be found in [15]. As defined there, these shapes can be classified as some combination of circumferential, tongues, and islands. "The classic long-segment Barrett's esophagus has a pink salmon-colored columnar mucosa... The proximal margin of the Barrett's esophagus may be horizontal (circumferential), or have irregular, tongue-shaped projections (tongues) [there may also be] pale islands or zones of regenerated or residual squamous epithelium (islands)."

Based on this classification of forms, in [14] the authors introduced the concept of Barrett's esophagus **complexity** using the following scheme. Note that the term "complexity", in this context, is not intended to imply a ranking, but simply a characterization.

C– if the Barrett's was classified as circumferential

CTI– if there was circumferential plus at least one tongue and/or island

TI– if there was no circumferential and at least one tongue and/or island.

Our interest is to model Total F04 in order to see if there was a treatment effect and, if so, to increase our understanding of what might be contributing to such an effect. A model for Total F04 should take into account the initial baseline QE measurement, Total BL. Also, given our findings in [14], clinic should be included as a covariate. Complexity class measures another baseline characteristic that may impact Total F04. For example, it is unknown why some patients have tongues while others do not or why some clinics have more patients with circumferential disease than do others.

In addition to individual effects of clinic and complexity on Total F04, one might consider an interaction effect between these two factors. Clinics differ based on investigators, practitioners and locations within the country. There are different patient demographics (racial and gender differences, co-morbid conditions), the participating clinics included private, public, and VA hospitals; there were differences in equipment (endoscopes, video recording devices, etc.); and differences in the assessment of Barrett's esophagus by operators. Given these potential sources of differential effects, it makes sense to explore the possibility of an interaction between clinic and complexity.

A total of 81 patients enrolled in the CBET study had QE measurements recorded at baseline, and 56 of these also had QE measurements recorded at one year. The idea of imputation to the larger population of 81 patients was considered, but, since our interest is in modeling a time-based trend, we considered it prudent to proceed with a complete case analysis. A two-sided unequal variances t-test and a Wilcoxon rank sum test comparing Total BL for the 25 patients with only baseline QE measurements with Total BL for the 56 patients with both baseline and one-year QE measurements was not significant ($p=0.3788$ and $p=0.5362$, respectively), indicating that baseline QE measurements for the 56 patients we will study are not inconsistent with measurements for those who did not continue in the study.

1.4. Statistical analysis

The results in this section are based on the 56 patients with QE measurements at baseline and at one year. Of these patients, 52 were treated at four major clinics. Our plan for the statistical analysis consists of fitting several models, all of which suggest that, when appropriate covariates are accounted for, there is a significant treatment effect. We will:

1. fit a fixed effects model, considering the four major clinics in the study as fixed. This model assumes that these four clinics are of specific interest.
2. fit a mixed model, where the six clinics involved are considered to be a random sample from a population of clinics. This allows us to generalize to a larger population of clinics.
3. fit a logistic model to an indicator variable called Direction, which takes on the values "Increase" for an increase in QE measurement between baseline and one year and "Decrease" for a decrease.

For both [1] and [2], we will fit a model to the Total F04 QE measurements, using Total BL as a covariate, and using as explanatory variables: Treatment, Clinic, Complexity (complexity at baseline), and the Clinic by Complexity interaction. As mentioned above, this last term is included to account for differential effects on F04 related to complexity classes across clinics. For the logistic model, the nominal response, Direction, includes the effects of Total BL. The logistic model contains only Treatment and Clinic as factors.

1.4.1. Fixed effects model—The fixed effects model is based on the 52 patients enrolled in the four major clinics. The reason for this restriction is that inclusion of the four patients from UHC and Mayo caused technical issues (too few degrees of freedom and aliasing of effects).

We began by fitting a model containing the factor of interest, Treatment, and the covariates Total BL, Clinic, and Complexity, and all two-way interactions of these four factors. We note that, in a typical covariance analysis, one would assume that there is no interaction between the covariates and the factor of interest. Nonetheless, there is no harm in exploring such interactions. We then reduced this model using backwards elimination with a cut-off p -value of 0.10. The resulting model contained the terms Treatment, Total BL, and the Clinic*Complexity interaction. Using the convention that a model should include the lower order terms involved in a significant interaction, we retained the main effects Clinic and Complexity.

We checked the residual plot for deviations from the regression assumptions and found the plot to be acceptable. We tested the Studentized residuals for normality using the Shapiro-Wilk goodness of fit test, and obtained a nonsignificant p -value (0.0607). Based on this evidence, and the fact that ANOVA is somewhat robust to deviations from normality, we concluded that the model provided an adequate fit to the data.

The ANOVA table for this model is given in Table 2. The effect tests are given in Table 3.

The overall model is highly significant. The effect tests assess whether a factor or interaction explains significant variation, given that all other effects are included in the model. As expected, Total BL is highly significant. Interestingly, but not surprisingly given our earlier discussion, the Clinic*Complexity interaction is also significant. Treatment is significant with a p -value of 0.0262, indicating that, given the other effects in the model, Treatment explains additional variability.

To gain some insight into the underlying structure of this model, consider Fig. 2. Fig. 2 plots the least squares means (predicted values) for Total F04, for each combination of Complexity level and Clinic. Note that the effect of complexity on the least squares mean for Total F04 is different for CPMC, than, say, for HVA. To some extent, CPMC and JHH are similar in their effect on Total F04 across Complexity levels. But HVA and UCLA differ from CPMC and JHH in their effect, especially for the CTI category. The least squares means are shown in Table 4.

For each clinic, contrasts were used to test complexity differences. These F-tests only show significance at the .05 level for CPMC ($p=0.0061$). Similarly, for each complexity class, contrasts were used to test clinic differences. The only significant F-test was for CTI ($p=0.0382$). These results suggest that the low predicted value of Total F04 for CPMC in the CTI complexity class is driving this interaction.

Finally, the least squares means table (Table 5) shows that the predicted Total F04 value for the celecoxib group is 26.28, while, for the placebo group, it is 31.99. Based on our model, we expect a reduction in total QE area of 5.71 units for those patients taking celecoxib rather

than the placebo. A 95% level confidence interval for this mean difference (placebo–celecoxib) is 0.71 to 10.71, with a standard error of 2.47.

We note, at this point, that for both the placebo and celecoxib groups, the mean change in QE area, Total F04–Total BL, is negative, implying a reduction in total area over the year. The mean differences for all 56 patients are given in Table 6. A two-sided t-test and a Wilcoxon rank sum test of the null hypothesis that the mean BL and F04 measurements are equal were conducted for each treatment group. The results, shown in Table 6, indicate that this difference is significantly non-zero for the celecoxib group, but not for the placebo group.

1.4.2. Mixed model—Here we proceed under the assumption that the actual clinics utilized in the study are not of intrinsic interest, but rather they are of interest as representatives of the class of all potential clinics where such a study might be conducted. This means that Clinic and the Clinic*Complexity interaction should be modeled as random effects [8,9,16]. In addition, treating these as random effects allows us to include the four patients from UHC and Mayo in the analysis. Also, since this model estimates random components for Clinic and the Clinic*Complexity interaction, aliasing and degree of freedom issues do not surface as they would have in a fixed effects model. The fitting methodology is Restricted Maximum Likelihood (REML).

The model we fit to Total F04 contains, as predictors, Total BL, Clinic, Complexity, and Clinic*Complexity. Table 7 presents the fixed effect tests. (Following the principle of effect hierarchy [8], we retain Complexity in the model because its interaction with Clinic is a random effect in the model.) Note that Treatment is significant at the .05 level ($p=0.0421$). So, we see that the mixed model lends support to the results of the fixed model analysis.

1.4.3. Logistic model—Yet another way to analyze this data is in terms of a nominal response. We define a variable called *Direction* as follows: Direction takes on the value “Increase” if a patient’s QE area measurement increases between the baseline and the one-year follow-up, and “Decrease” if that value decreases. This removes issues of non-normality from the modeling process. Using a logistic model, we model the probability of a decrease between baseline and one year. Our interest is in whether or not there is a treatment effect.

Because of the nature of the logistic model, we will treat Clinic as a fixed effect and, to be able to estimate the required parameters, we will only look at the four main clinics. The numbers of patients with increases and decreases in QE area are given in Table 8, broken down by Clinic and Treatment.

Neither a likelihood ratio nor a Pearson chi-square test of Direction by Treatment is significant ($p=0.1638$ and $p=0.1652$, respectively). We explored a logistic model for Direction containing the predictors Treatment, Clinic, and Complexity. (Because of the structure and sparsity of the data, it was not feasible to include the Clinic by Complexity interaction.) This model was highly significant ($p=0.0081$). The effect tests indicated significance for Treatment ($p=0.0370$) and Clinic ($p=0.0035$), while Complexity was not significant ($p=0.1971$).

Since Complexity was not significant, it was removed from the model. This reduced model is significant with $p=0.0070$, as shown in Table 9. The effect tests, displayed in Table 10, show that, with Clinic as a covariate, Treatment is highly significant.

The estimated probability of a decrease in QE area over the one year period for someone in the celecoxib group is 0.8428, while for someone in the placebo group, the estimated probability of a decrease is 0.5283. This leads to an odds ratio for a decrease in QE area over the one-year period of 4.79 for the celecoxib versus the placebo group.

2. Conclusion

The larger reduction in lesion area for the celecoxib group is consistent with the initial hypothesis of the CBET trial. For the continuous response F04, treatment group significance was realized when the model controlled for Total BL, Clinic, and the Clinic*Complexity class interaction effect. For the nominal change indicator, Direction, treatment group significance was observed once Clinic was accounted for.

There is a saying, attributed to the famous statistician George E. P. Box, asserting “All models are wrong, but some are useful.” This paper has attempted to provide useful statistical models for QE area measurements as they relate to treatment effects for QE area in the CBET study. These models use covariates to control for baseline variation that might otherwise obscure treatment effects. We have attempted to illustrate how, in the case of a small multicenter clinical trial, the use of covariates to explain imbalance in baseline characteristics can increase the potential to discern significant treatment effects.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We very much appreciate the efforts of the reviewers, whose careful critiques and valuable insights led to significant improvements in this paper. This work is that of the authors and does not imply any endorsement by or position of the Robert Wood Johnson Foundation.

Funding disclosure Funding for the CBET trial came from Pfizer and the NCI. There was no additional funding for this particular study. There are no financial conflicts of interest. However, we note that Albert Shar holds the patent to the methodology used to calculate QE area. This work is that of the authors and does not imply any endorsement by or position of the Robert Wood Johnson Foundation.

References

- [1]. Koppert L, Wijnhoven B, van Dekken H, Tilanus H, Dinjens W. The molecular biology of esophageal adenocarcinoma. *J Surg Oncol*. 2005; 92(3):169–90. [PubMed: 16299787]
- [2]. Devesa SS, Blot WJ, Fraumeni JF Jr. Changing patterns in the incidence of esophageal and gastric carcinoma in the United States. *Cancer*. 1998; 83:2049–53. [PubMed: 9827707]
- [3]. Brown LM, Silverman DT, Pottner LM, et al. Adenocarcinoma of the esophagus and esophagogastric junction in white men in the United States: alcohol, tobacco, and socioeconomic factors. *Cancer Causes Control*. 1994; 5:333–40. [PubMed: 8080945]
- [4]. Drewitz DJ, Sampliner RE, Garewal HS. The incidence of adenocarcinoma in Barrett’s esophagus: a prospective study of 170 patients followed 4.8 years. *Am J Gastroenterol*. 1997; 92:212–5. [PubMed: 9040193]
- [5]. Shaheen NJ, Crosby MA, Bozymski EM, Sandler RS. Is there publication bias in the reporting of cancer risk in Barrett’s esophagus? *Gastroenterology*. 2000; 11:333–8.
- [6]. Heath EI, Canto MI, Wu TT, et al. Chemoprevention for Barrett’s esophagus trial. Design and outcome measures. *Dis Esophagus*. 2003; 16(3):177–86. [PubMed: 14641306]
- [7]. Heath EI, Canto MI, Piantadosi S, et al. Secondary chemoprevention of Barrett’s Esophagus with celecoxib. *J Natl Cancer Inst*. 2007; 99(7):545–57. [PubMed: 17405999]
- [8]. Montgomery, DC. Design and analysis of experiments. Sixth Edition. Wiley; New York: 2005.

- [9]. Lentner, M.; Bishop, T. Experimental design and analysis. Second Edition. Valley Book Co.; 1986.
- [10]. Kim R, Baggott BB, Rose S, et al. Quantitative endoscopy: precise computerized measurement of metaplastic epithelial surface area in Barrett's esophagus. *Gastroenterology*. 1995; 108(2):360–6. [PubMed: 7835577]
- [11]. Kim R, Rose S, Shar AO, Weiner M, Reynolds JC. Extent of Barrett's metaplasia: a prospective study of the serial change in area of Barrett's measured by quantitative endoscopic imaging. *Gastrointest Endosc*. 1997; 45(6):456–62. [PubMed: 9199900]
- [12]. Shar AO, Baggott BB, Reynolds JC. Computer enhanced endoscopic visualization. *Proc SCAMC*. 1990:544–6.
- [13]. Shar AO, Weiner M, Kim R, Reynolds JC. Automated reconstruction of endoscopic images of the esophagus. *Proc SCAMC*. 1995:547–9.
- [14]. Shar AO, Gaudard MA, Heath EI, et al. Quantitative endoscopy in the Chemoprevention of Barrett's Esophagus trial. *Dis Esophagus*. 2008; 21
- [15]. Meinhard, C.; Tytgat, GNJ.; Lightdale, CJ. *Gastroenterological endoscopy*. Thieme; 2002.
- [16]. Box, GEP.; Hunter, WG.; Hunter, JS. *Statistics for experimenters*. 2nd ed. Wiley; New York: 2005.

Fig. 1.
QE process flow-raw image, outlined image, transformed image.

Fig. 2.
Least squares means by complexity class and clinic.

Table 1

Number of patients with baseline and one-year QE measurements by clinic

Clinic	Number of patients
CPMC	13
HVA	14
JHH	15
UCLA	10
UHC	3
MAYO	1

Table 2

ANOVA Table for fixed effects model

Source	df	Sum of squares	Mean square	F ratio	Prob >F
Model	13	30171.35	2320.87	34.66	<0.0001
Error	38	2544.68	66.97		
C. total	51	32716.03			

Table 3

Effect tests for fixed effects model

Source	df	Sum of squares	F Ratio	Prob > F
Treatment	1	358.29	5.35	0.0262
Total BL	1	14402.20	215.07	<0.0001
Clinic	3	134.42	0.67	0.5763
Complexity	2	18.99	0.14	0.8682
Clinic*Complexity	6	1010.73	2.52	0.0378

Table 4

Least squares means and standard errors by complexity and clinic

Clinic	LS Mean C (SE, N)	LS Mean CTI (SE, N)	LS Mean TI (SE, N)
CPMC	39.18 (3.25, 7)	21.52 (4.11, 4)	26.10 (5.97, 2)
HVA	31.43 (4.11, 4)	38.36 (4.89, 3)	24.81 (3.26, 7)
JHH	26.69 (5.10, 3)	25.15 (2.59, 10)	27.90 (5.94, 2)
UCLA	23.71 (9.02, 1)	32.05 (4.21, 4)	32.74 (3.86, 5)

Table 5

Least squares means for F04

Level	N	Least sq mean	Std error	Sample mean
Cele	27	26.28	1.76	26.76
Plbo	25	31.99	1.94	32.07

Table 6

Mean decreases by treatment group

Treatment	N	Mean (Total F04–Total BL)	df	Standard error	T-test p-value	Rank sum test p-value
Celecoxib	29	–4.73	28	1.83	0.0150	0.0065
Placebo	27	–1.49	26	1.84	0.4272	0.5581

Table 7

Fixed effect tests for mixed model

Source	<i>N</i> parameters	<i>df</i>	<i>df</i> den	<i>F</i> ratio	Prob > <i>F</i>
Treatment	1	1	48.12	4.36	0.0421
Total BL	1	1	50.01	266.49	<0.0001
Complexity	2	2	8.70	0.61	0.5638

Table 8

Number of patients with increases or decreases in QE area between baseline and one-year follow-up

Clinic	Treatment	Decrease	Increase
CPMC	Cele	5	3
	Plbo	2	3
HVA	Cele	6	0
	Plbo	6	2
JHH	Cele	7	0
	Plbo	6	2
UCLA	Cele	3	3
	Plbo	1	3
		36	16

Table 9

Whole model test for reduced logistic model

Model	–Log likelihood	df	Chi square	Pro>Nchi sq
Difference	7.06	4	14.11	0.0070
Full	25.04			
Reduced	32.10			

Table 10

Effect tests for reduced logistic model

Source	N parameters	df	Likelihood ratio chi square	Pro>Nchi sq
Treatment	1	1	4.59	0.0322
Clinic	3	3	12.17	0.0068