A Randomized, Dose-Ranging Study of D-Penicillamine Treatment in Scleroderma

by Yu Xia
June 12, 2019

Abstract

In conformity with the clinical trial protocol, the main objective of this paper is to evaluate the equivalence of drug efficacy in high and low dose of D-Penicillamine as treatment on Scleroderma patients. In primary analysis, a wide range of statistical analyses are conducted, based on the original health assessment questionnaire score (HAQ) which is continuous and the discretized version of HAQ, from two perspectives. For continuous HAQ, the main statistical method we use is GEE model on longitudinal data and subsequent Wald test. For discretized HAQ, Pearson's chi-squared test and Fisher's exact test are used to evaluate the equivalence between high and low dose. Furthermore, logistic regression and ordinal logistic regression are implemented to investigate the relationship between HAQ and some covariates of medical information. Besides, a secondary analysis is conducted for ancillary purposes.

Conclusions stay consistent throughout the primary analysis whatever the statistical method we use, leading to no difference of efficacy between high and low dose of D-Penicillamine. Some demographics and medical measurements are deemed related to the number of falls reported by the patient in the prior 6 months he/she had at baseline.

The paper is concluded by some remarks in discussion section.

Key words: D-Penicillamine, Scleroderma, clinical trial, longitudinal data, GEE

1 Introduction

Background of Scleroderma and Penicillamine

In this research, we focus on Scleroderma and some treatments of this disease. Scleroderma is a group of autoimmune diseases that may result in changes to the skin, blood vessels, muscles, and internal organs. The disease can be either localized to the skin or involve other organs in addition to the skin. Symptoms may include areas of thickened skin, stiffness, feeling tired, and poor blood flow to the fingers or toes with cold exposure.

Now, we can consider two different concentration of D-Penicillamine – high dose and low dose. Penicillamine is a medication primarily used for the treatment of Wilson's disease. It is also used for people with kidney stones who have scleroderma, high urine cystine levels, rheumatoid arthritis, and various heavy metal poisonings. It is taken by mouth.

Study Design

In this two-arm clinical trials study, the main goal is to judge whether high dose of D-Penicillamine improves the health assessment questionnaire (HAQ) scores more significantly than a low dose of D-Penicillamine after two-year treatment.

All patients were randomly assigned into either the case group (group = 1) or the control group (group = 0). The number of group 1 and group 2 are same for the equivalence-study requirement. Patients in the case group were given a high dose of D-Penicillamine for Scleroderma disease treatment while patients in the control group were given a low dose of D-Penicillamine. In two-year treatment, the researchers ask patients to come back 6-months a time to record their blood pressure, serum creatine phosphokinase, cardiomegaly, cardiothoracic ratio and other features for the test.

Main objective

In this two-arm clinical trials study, the main goal is to judge whether high dose of D-Penicillamine improves the health assessment questionnaire (HAQ) scores more than low dose of D-Penicillamine after two-year treatment.

Study Population

141 patients enrolled in this clinical-trial study in total. In 2-year this study, patients were asked to come back 6-months a time after the first baseline visit; therefore, for the duration of the study, each patient need to have 5 records for this study.

Overall Statistical Analysis

In this analysis, we will consider two independent sample t-test, longitudinal analysis, GEE models and Pearson's chi-squared test.

For the model construction, we will use logistic regression and ordinal logistic regression. Also, MLR, Poisson Regression, Negative Binomial Regression, Zero-Inflated Poisson Regression and Zero-Inflated Negative Binomial Regression will be considered.

Exclusion Criteria

However, subjects having any of the following criteria, either at baseline or any follow-up visits, will not be included in the study:

1) Records which do not have HAQ data cannot be used. Because in this case we need to focus on HAQ, missing data of HAQ are not in our consideration.

2) Patient records with a missing value in visit-date cannot be used because if we do not have the date of record, we cannot find the change of HAQ from their baseline.

After these exclusion criteria, we finally keep 134 subjects and 495 observations in the data set. In below analysis, this paper only focuses on 526 observations.

2 Descriptive Analysis

Using Exclusion Criteria, the dataset we use has 495 observations involving 134 subjects (Scleroderma patients). (described in **Table 1**).

Table 1

Variable	Description	Format
patid	patient medical record number	Character
visitdat	date of the visit	Datetime
sex	1 = female, 0 = male, 22 = unknown	Categorical
race	0 = White, 1 = Black or African American, 2 = Asian, 3 = Native Hawaiian or Other Pacific Islander, 4 = Other	Categorical
age	age in years	Float
weight	weight (lb.)	Float
group	1 = high dose, 0 = low dose	Categorical
haq	total health assessment questionnaire score	Float
sbp	systolic blood pressure	Float
cpk	serum creatine phosphokinase	Float
cardrate	cardiothoracic ratio	Float
cardmega	1 = cardiomegaly, 0 = no cardiomegaly	Categorical
nooffalls	number of falls 6 months prior to baseline	Float

2.1.1 Sample Characteristics

The sample characteristics of patients are summarized into **Table 2**. Also, sample characteristics based on dose of D- Penicillamine and completion of study are summarized into **Table 3**.

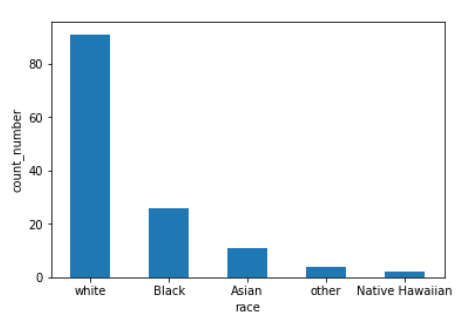
Table 2. Characteristics of the patient

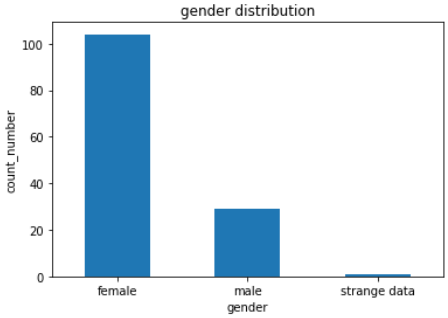
	age	weight	sbp	cpk	haq	nooffalls
count	134	134	134	133	134	124
mean	45.04	155.36	121.72	71.8769913	1.04	0.55
std	13.33	37.20	17.57	71.5166626	0.69	1.03
min	19.79	85.75	90.66	9.25	0	0
25%	35.05	133.18	110	33	0.37	0
50%	44.25	148.67	118.80	47	1.01	0
75%	52.06	168.42	133.04	80	1.57	1
max	101	370.20	179	495.66	2.68	5

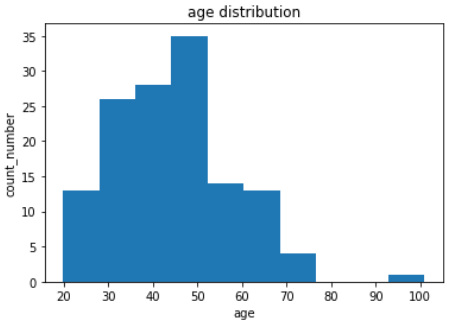
Table 3. Characteristics of the patient

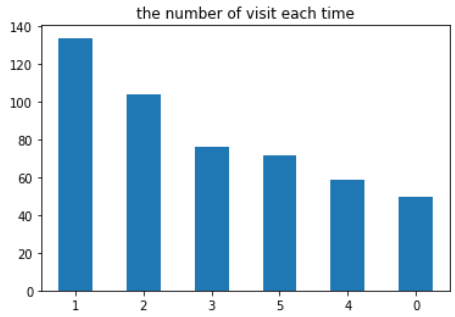
Characteristic	Dose of D-Penicillamine		Completion of Study	
	High-Dose Group (n = 64)	Low-Dose Group (n = 70)	Completers Group (n = 67)	Non-completers Group (n = 67)

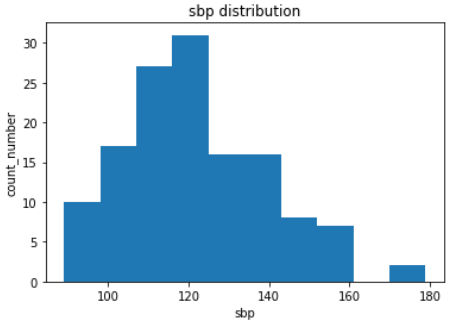
sex				
male	18 (28.12)	11 (15.71)	16 (22.73)	13 (20.59)
female	46 (71.88)	58 (82.86)	50 (75.76)	54 (79.41)
unknown	0 (0)	1 (1.43)	1 (1.52)	0 (0)
race				
White	45 (70.31)	46 (65.71)	44 (65.15)	47 (70.59)
Black	12 (18.75)	14 (20)	14 (21.21)	12 (17.65)
Asian	5 (7.81)	6 (8.57)	5 (7.58)	6 (8.82)
Pacific	1 (1.56)	1 (1.43)	1 (1.52)	1 (1.47)
Other	1 (1.56)	3 (4.29)	3 (4.55)	1 (1.47)
age	42.53	45.99	45.2	44.84
weight	157.98	151.66	157.08	153.54
haq	1.05	1.03	0.89	1.20
sbp	121.83	118.67	119.29	123.84
cpk	76.83	86.24	68.58	75.16
cardrate	0.46	0.47	0.45	0.48
cardmega				
no cardiomegaly	55 (85.94)	61 (87.14)	59 (89.39)	57 (83.82)
cardiomegaly	9 (14.06)	8 (11.43)	7 (10.61)	10 (14.71)
nooffalls	0.17 ± 0.39	0.38 ± 0.77	0.71	0.25 ± 0.52

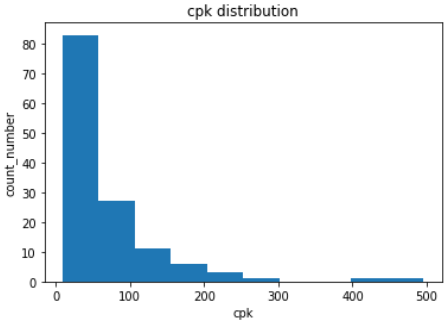












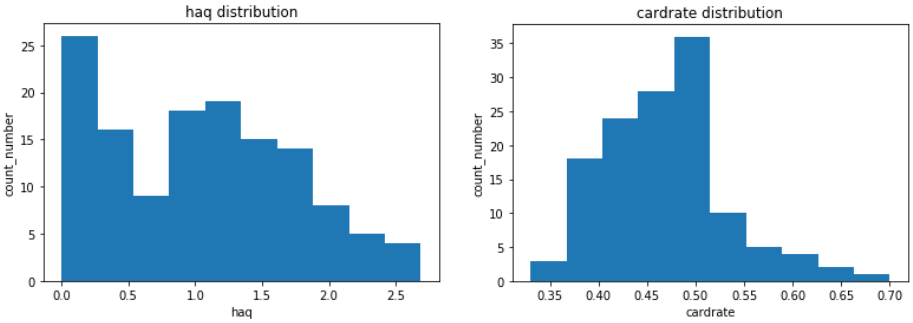
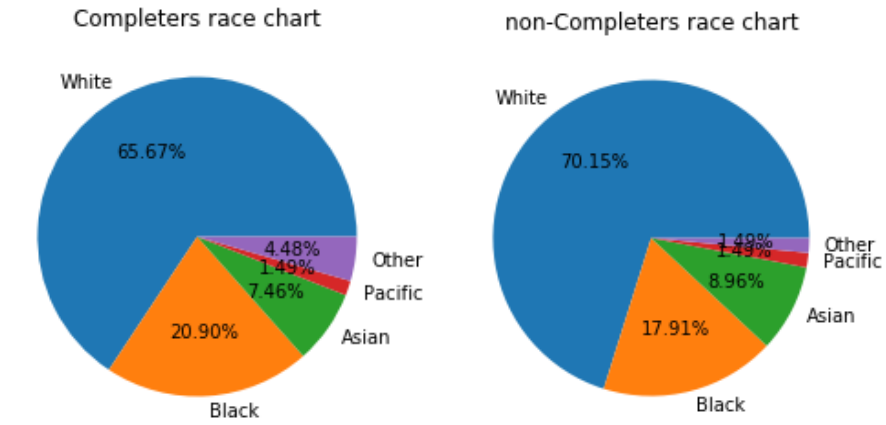
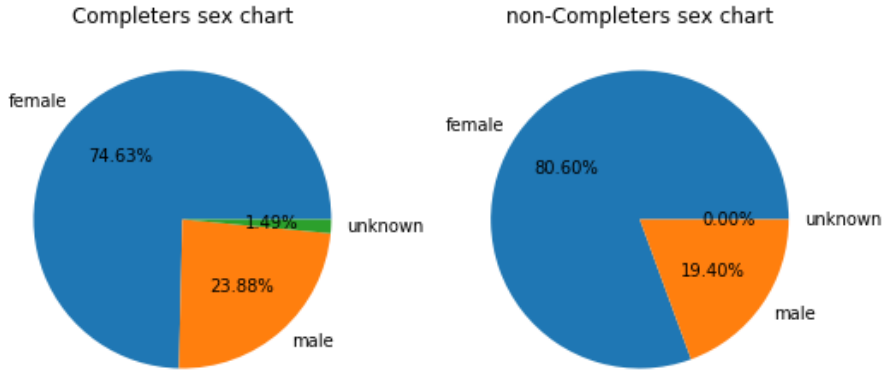


Figure 1. Distributions of different variable





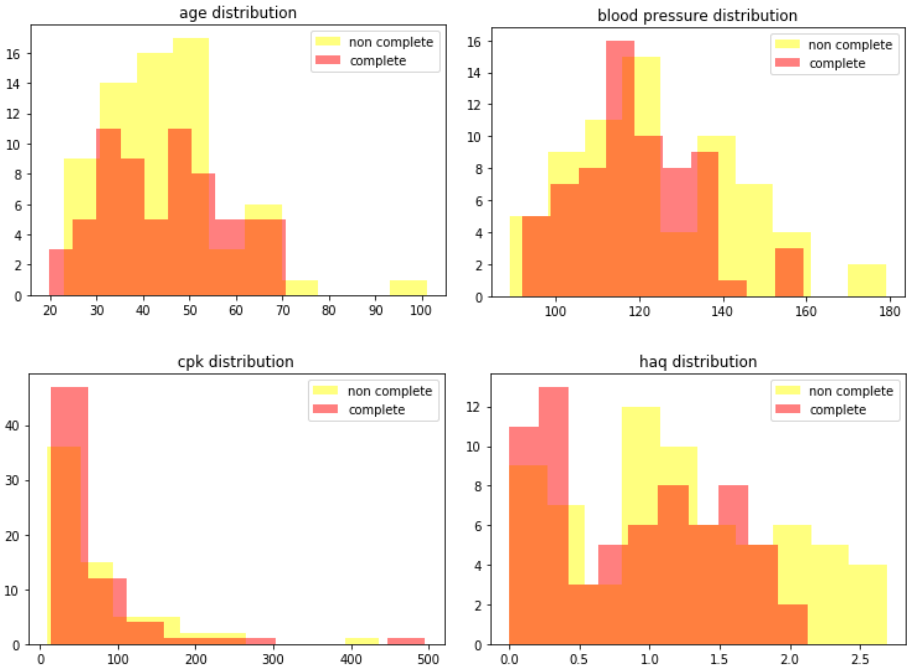


Figure 2. Distribution of completers and non-completers

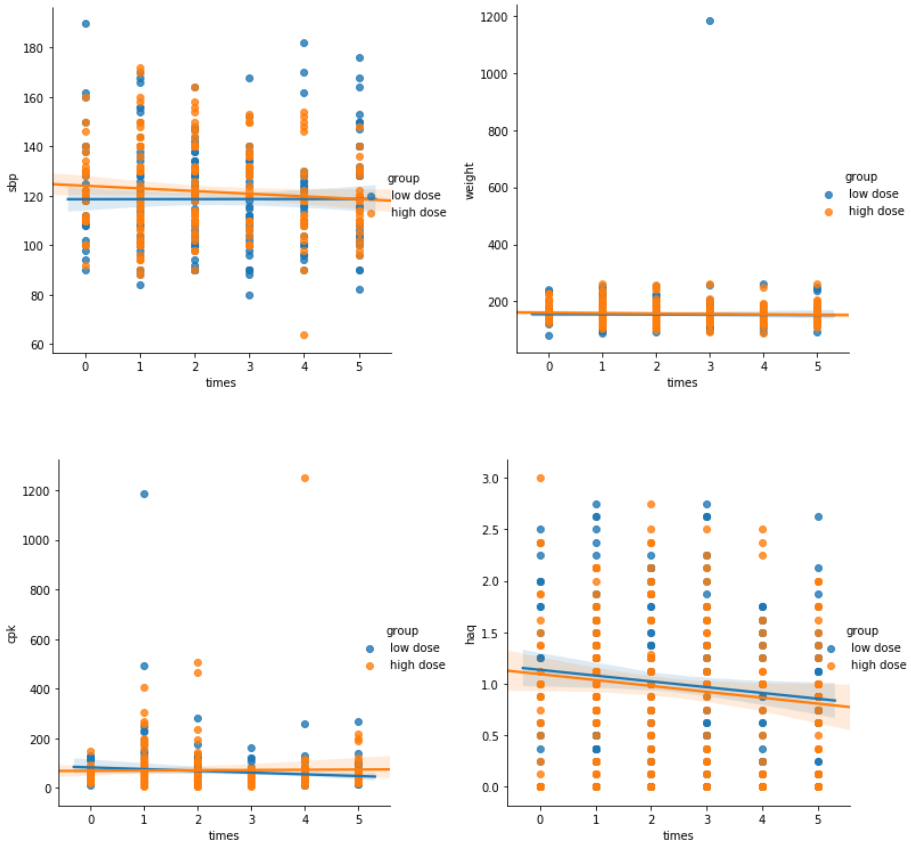


Figure 3. differences in patients treated in the two groups (high-dose, low-dose)

Findings:

1) In the table 3, we notice that HAQ of non-completers is higher than that of completers; meanwhile, SBP and CPK shows the similar trend.

2) In this dataset, the number of females is much more than male.

3) The HAQ score in the completer group is slightly different from that of the non-completer group.

4) The percentage of completers and non-completers in White is higher than other

5) Both groups of patients' HAQ scores decrease over time overall, and patients treated with high-dose of D-Penicillamine showed a larger magnitude of decrease.

6) For patients who were treated with high-dose of D-Penicillamine, their SBP decreases faster than that of patients who are in low-dose group.

2.2 Outlier Detection

Table 4: different outliers

Variable	Outlier	Available Obs.
sex	22	494
weight	2	493
sbp	24	488
cpk	10	485
age	1	494

In sex column, one observation is 22, which has no meaning; therefore, we need to delete that record; In the weight column, 2 observations are larger than 1000, which has no meaning; therefore, we need to delete that record;

In cpk column, 10 observations are larger than 400, in the histogram above, we know that these are outliers, however, we do not know whether we can delete these observations or not. Therefore, we still keep these data for other tests.

In age column, 1 observation are larger than 90, in the histogram above, we know that these are outliers, however, we do not know whether we can delete these observations or not. Therefore, we still keep these data for other tests.

2. 3 Summary of Multiple Visits

Table 5: The number of patients who showed up at each visit

Visit	Count
1	134
2	104
3	76
4	67
5	72

As the data shown above, we focus on whether high dose of D-pen improves the HAQ scores more significantly than a low dose of D- Penicillamine at the end of the study in many different ways. Firstly, we wish to determine which data in these two groups we want to consider in our analysis.

3 Primary Analysis

The main research question in the study is whether high dose of D-pen improves the health assessment questionnaire (HAQ) scores more significantly than a low dose of D- Penicillamine at the end of the study.

The first way is to consider the HAQ variable as continuous variable: in this way, we use five different data analysis methods to build model and determine whether patients in the two groups experienced a significant change in their HAQ scores at the end of the study.

The second way is to consider the HAQ variable as categorical variable and determine drug efficacy is to work with a discretized version of HAQ: in this way, we use two ways to determine the HAQ level and test whether the drug equally affects the completers in the two treatment groups at the end of 24 months.

3.1 Efficacy Evaluation Using Original HAQ (continuous)

3.1.1 Statement of Methods

This way is to consider the HAQ variable as continuous variable and determine drug efficacy is to work with it. We first calculate the difference of HAQ score, then we use these five methods for available data:

$$diff\_haq = Haq(final\ visit) - Haq(baseline)$$

Five methods:

(1) Using only available data (i.e. patients who had both baseline and 2-year data)

(2) Using data from completers only (i.e. patients who showed up for all scheduled visits)

(3) If patients have missing data at 2 years after baseline, impute them by carrying forward the last available data to 2 years

(4) Working with data only collected at visit dates that are within 10 days of the scheduled visit date; otherwise they are deem outside the visit window and so unacceptable

(5) Using ALL available HAQ data from each patient.

Method 1: Using only available data (i.e. patients who had both baseline and 2-year data):

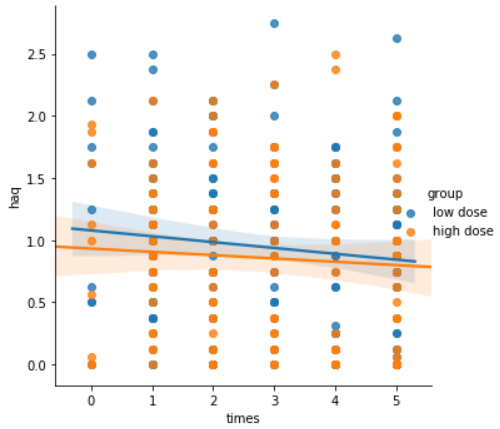


Figure 4. HAQ score regressed on visit in two study groups under method1 specific data

Method 2: Using data from completers only (i.e. patients who showed up for all scheduled visits)

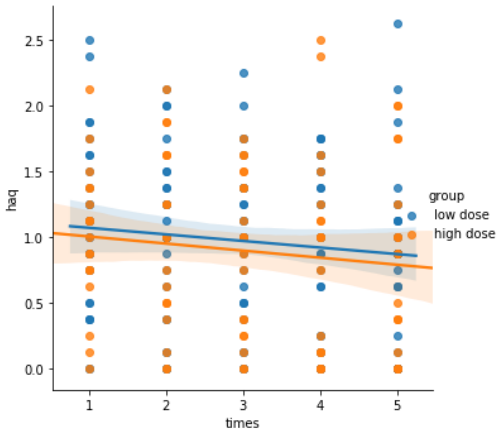


Figure 5. HAQ score regressed on visit in two study groups under method2 specific data

GEE Regression Results

Dep. Variable:	haq	No. Observations:	322
Model:	GEE	No. clusters:	67
Method:	Generalized Estimating Equations	Min. cluster size:	4
		Max. cluster size:	5
Family:	Gaussian	Mean cluster size:	4.8
Dependence structure:	Exchangeable	Num. iterations:	5
Date:	Tue, 11 Jun 2019	Scale:	0.450
Covariance type:	robust	Time:	21:46:52

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	1.0896	0.123	8.826	0.000	0.848	1.332
group	-0.1948	0.161	-1.206	0.228	-0.511	0.122
times	-0.0492	0.022	-2.261	0.024	-0.092	-0.007
group:times	0.0153	0.029	0.525	0.600	-0.042	0.072

Skew:	0.3383	Kurtosis:	-0.7352
Centered skew:	0.2387	Centered kurtosis:	1.2282

Method 3: If patients have missing data at 2 years after baseline, impute them by carrying forward the last available data to 2 years

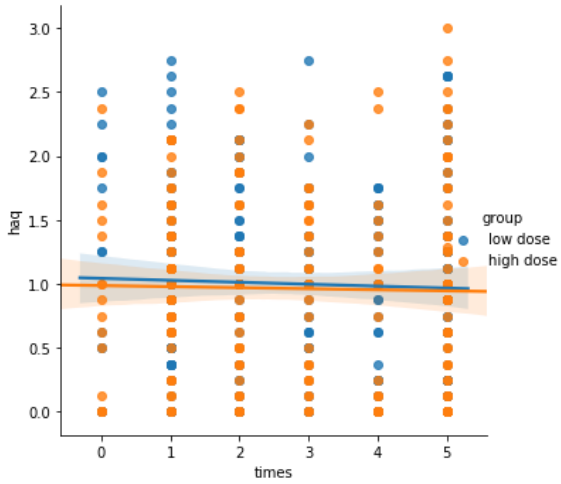


Figure 6. HAQ score regressed on visit in two study groups under method3 specific data

GEE Regression Results

Dep. Variable:	haq	No. Observations:	480
Model:	GEE	No. clusters:	134
Method:	Generalized Estimating Equations	Min. cluster size:	1
		Max. cluster size:	5
Family:	Gaussian	Mean cluster size:	3.6
Dependence structure:	Exchangeable	Num. iterations:	5
Date:	Tue, 11 Jun 2019	Scale:	0.522
Covariance type:	robust	Time:	21:55:11

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	1.1109	0.095	11.651	0.000	0.924	1.298
group	-0.0107	0.127	-0.084	0.933	-0.260	0.239
times	-0.0339	0.019	-1.740	0.082	-0.072	0.004
group:times	0.0019	0.027	0.071	0.943	-0.051	0.055

Skew:	0.3630	Kurtosis:	-0.7046
Centered skew:	0.0031	Centered kurtosis:	1.2729

Method 4: Working with data only collected at visit dates that are within 10 days of the scheduled visit date; otherwise they are deem outside the visit window and so unacceptable

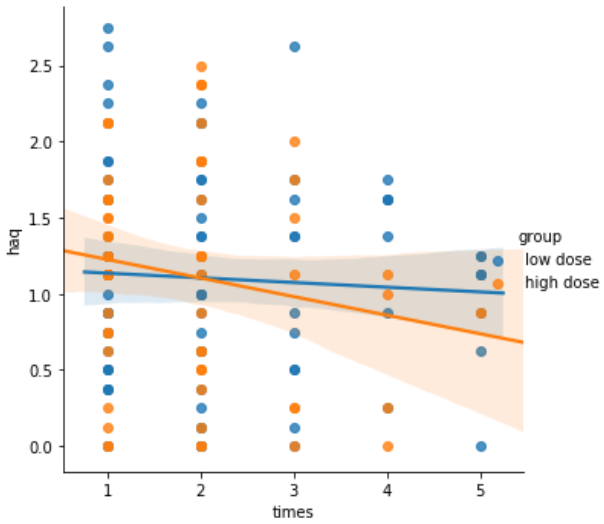


Figure 7. HAQ score regressed on visit in two study groups under method4 specific data

GEE Regression Results

Dep. Variable:	haq	No. Observations:	147
Model:	GEE	No. clusters:	59
Method:	Generalized Estimating Equations	Min. cluster size:	1
		Max. cluster size:	5
Family:	Gaussian	Mean cluster size:	2.5
Dependence structure:	Exchangeable	Num. iterations:	5
Date:	Tue, 11 Jun 2019	Scale:	0.491
Covariance type:	robust	Time:	21:57:39

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	1.1711	0.164	7.135	0.000	0.849	1.493
group	0.1231	0.212	0.580	0.562	-0.293	0.539
times	-0.0179	0.047	-0.381	0.703	-0.110	0.074
group:times	-0.0676	0.069	-0.981	0.327	-0.203	0.067

Skew:	0.1786	Kurtosis:	-0.7076
Centered skew:	-0.2922	Centered kurtosis:	1.4607

Method 5: Using ALL available HAQ data from each patient.

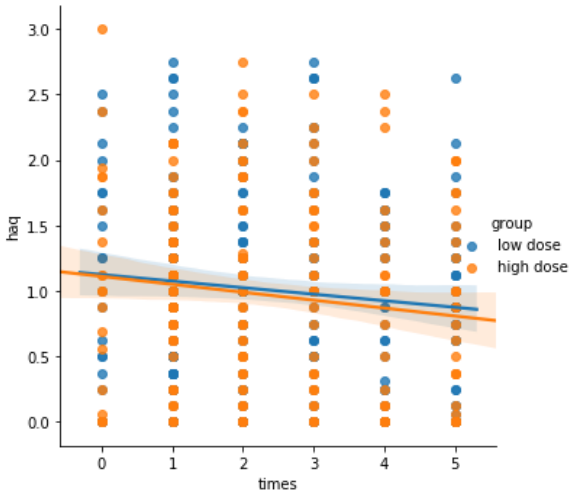


Figure 8. HAQ score regressed on visit in two study groups under method 5 specific data

GEE Regression Results

Dep. Variable:	haq	No. Observations:	480
Model:	GEE	No. clusters:	134
Method:	Generalized Estimating Equations	Min. cluster size:	1
		Max. cluster size:	5
Family:	Gaussian	Mean cluster size:	3.6
Dependence structure:	Exchangeable	Num. iterations:	5
Date:	Tue, 11 Jun 2019	Scale:	0.522
Covariance type:	robust	Time:	21:59:16

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	1.1109	0.095	11.651	0.000	0.924	1.298
group	-0.0107	0.127	-0.084	0.933	-0.260	0.239
times	-0.0339	0.019	-1.740	0.082	-0.072	0.004
group:times	0.0019	0.027	0.071	0.943	-0.051	0.055

Skew:	0.3630	Kurtosis:	-0.7046
Centered skew:	0.0031	Centered kurtosis:	1.2729

Model Assumptions

For all of the models, each record of HAQ score is independently and identically distributed; For all of the models, the data are continuous: Satisfied since the difference two HAQ scores is a continuous measurement; The data follow normal distributions

3.1.2 Results of five different methods

Table 6. Summary of univariate outliers

Method	p-value	Significance
Using only available data	0.418	Not significant
Completers only	0.228	Not significant
All with missing imputed	0.933	Not significant
Working with data only collected at visit dates	0.562	Not significant
All data	0.933	Not significant

We firstly use two-sample t test in the first method, p-value represents there is no statistical evidence to show that high dose group and low dose group have different impact on HAQ. In the method 2 to 5, we use Longitudinal GEE model, p-value (p-value> 0.05) represents there is no statistical evidence to show that high dose group and low dose group have different impact on HAQ. Because we need complete data to satisfy the GEE model assumptions, method 2 is the most recommended method among these. Method 4 also own complete data, but the number of data it uses is very small.

3.2 Efficacy Evaluation Based on Discretized HAQ

This way is to consider the HAQ variable as categorical variable and determine drug efficacy is to work with a discretized version of HAQ: in this way, we use two ways to determine the HAQ level and judge whether the drug equally affects the completers in the two treatment groups at the end of 24 months.

3.2.1 Binary HAQ

Definition: binary variable where patients are deemed to have improved if their HAQ scores had dropped by at least 30% at the end of the 24th month from their baseline.

In this design, we use two methods to judge the difference.

Method 1: Pearson's chi-squared test

Table 7. 2×2 contingency table

	Improved	Not Improved
High Dose	28	26
Low Dose	22	16

Table 8. Pearson's chi-squared test

Method	p-value	Significance
Pearson's chi-squared test	0.719	Not significant

Method 2: Logistic Regression

Table 9. Result of logistic regression

Covariate	Estimate	Std. Error	z-value	Pr(>|z|)
Intercept	-0.0741	0.385	-0.192	0.847
group	-0.2443	0.604	-0.405	0.686

Then we take *cardratio, sbp, cpk* into consideration and build the logistic regression.

Table 10. Result of logistic regression

Covariate	Estimate	Std. Error	z-value	Pr(>|z|)
Intercept	3.1164	3.394	0.918	0.359
group	-0.7404	0.717	-1.032	0.302
cardrate	-12.5409	7.348	-1.707	0.088
sbp	0.0221	0.024	0.933	0.351
cpk	0.0006	0.002	0.299	0.765
G-cardrate	-16.9507	12.3948	1.3676	0.28
G- sbp	0.0467	0.0432	-0.204	0.173
G- cpk	-0.0034	0.017	1.0799	0.84

Overall, there is no statistically significant difference of efficacy between high and low dose of D-Penicillamine in terms of p-value>0.05. Also, the p-value of each interaction term is not significant. Therefore, there is no difference of efficacy between high and low dose of D-Penicillamine depended on any covariates

3.2.2 Multi-class HAQ

Definition: break up the HAQ variable into 3 categories: low (value is smaller than 1), moderate (value is at least 1 and smaller than 1.5) and high (value is 1.5 or larger).

Use these two discretized versions of HAQ and ascertain whether the drug equally affects the completers in the two treatment groups at the end of 24 months.

Table 11. 2×3 contingency table based on multi-class HAQ definition

	Low	Moderate	High
High Dose	26	12	0
Low Dose	36	14	4

Table 12. Fisher test

Method	p-value	Significance
Pearson's chi-squared test	0. 314	Not significant

Table 13. Result of multiclass logistic regression

Covariate	Estimate	Std. Error	z-value	Pr(>|z|)
Intercept	-0.9445	0.445	-2.12	0.034
group	0.1713	0.665	0.258	0.797
Covariate	**Estimate**	**Std. Error**	**z-value**	**Pr(>|z|)**
Intercept	-2.1972	0.745	-2.948	0.003
group	-23.8883	1.28E+05	0	0.793

Table 14. Result of ordinal logistic regression

Covariate	Estimate	Std. Error	z-value	Pr(>|z|)
group	-0.085	0.5878	-0.1329	0.794
cardrate	7.273	7.252	1.2240	0.321
sbp	-0.005	0.277	-0.4902	0.626
cpk	-0.0013	0.0193	-0.4421	0.328
G-cardrate	-5.9507	10.3948	1.3676	0.18
G- sbp	0.0967	0.1432	-0.204	0.183
G- cpk	-0.0054	0.017	1.0799	0.84

When we divide HAQ into three different levels, there is no statistically significant difference of efficacy between high and low dose of D-Penicillamine. In the fisher exact test, we obtain p-value is 0.314, which means there is no statistically significant difference of efficacy between high and low dose of D-Penicillamine. And we use multiclass logistic regression to check the answer again. Also, the p-value of each interaction term is not significant. Therefore, there is no difference of efficacy between high and low dose of D-Penicillamine depended on any covariates.

4 Secondary Analysis

4.1 Summary of result

The result of fitting MLR, ZIP and ZINB regression is summarized into **Table 15 and Table 16**. After AIC criteria, the zero-inflated models show an improvement on regression model.

4.1.1 Multiple Linear Regression

Table 15. Result of MLR

Covariate	Estimate	Std. Error	z-value	Pr(>|z|)
Intercept	0.8242	1.304	0.632	0.532
group	-0.262	0.239	-1.095	0.282
age	-0.0142	0.009	-1.504	0.143
sex	0.0814	0.325	0.251	0.804
weight	0.0007	0.004	0.192	0.849
sbp	-0.0003	0.006	-0.053	0.958
cpk	-0.0006	0.002	-0.238	0.813
cardmega	0.389	0.562	0.692	0.494
cardrate	-0.5399	2.154	-0.251	0.804
haq	-0.0034	0.174	-0.02	0.984
race = 0	0.3373	0.71	0.475	0.638
race = 1	0.1635	0.79	0.207	0.837
race =2	0.0714	0.868	0.082	0.935

4.1.2 ZIP Regression

Table 16. Result of ZIP

Baseline Variables	Poison-Estimate	Pr(>|z|)	Logit- Estimate	Pr(>|z|)
(Intercept)	1.7374	0.109	-15.167	0.958
sex	1.2351	0.005	0.2814	0.942
age	-0.0611	0.323	0.095	0.524
race = 0	0.068	0.71	-	-
race = 1	-0.3983	0.79	-	-
race =2	-0.5642	0.868	-	-
haq	0.6163	0.008	1.492	0.032
sbp	-0.0391	0.000	-0.193	0.312
weight	0.0206	0.037	0.083	0.323
cardmega	0.2393	0.544	-16.793	0.966

4.1.3 ZINB Regression

Table 17. Result of ZINB

Baseline Variables	NB - Estimate	Pr(>|z|)	Logit- Estimate	Pr(>|z|)
(Intercept)	1.6853	0.165	-0.455	0.278
sex	1.1638	0.010	21.450	0.471

age	-0.0096	0.437	-	-
race	-0.3760	0.034	-1.949	0.205
haq	0.6139	0.024	4.625	0.095
sbp	-0.0290	0.000	-0.254	0.070
weight	0.0112	0.069	-	-
cardmega	0.1625	0.660	-	-

haq**:** for a one-unit change in the HAQ score, the difference in the log-scale of expected number of falls patients changes by 1.492, controlling for other varaibles.

sex**:** when a patient's gender is presumed to transfer from female to male, the difference in the log-scale of expected counts of the number of falls patients changes by 0.2814, controlling for other varaibles.

5 Conclusion

In this report, we first overview description of the data and defined different types of participants: completers, non-completers based on the completion of the clinical trail. Then based on different groups (completers, non-completers and high dose, low dose), we plot several graphs to find the difference.

Then we use two ways of modelling to find he relationship between the change of HAQ score and group: first is to consider HAQ as continuous variable, second is to consider HAQ as different levels. For first one, we try both t-test and GEE models; both of them suggest that there is no statistical significance. In second way, we try Logistic regression and overall, there is no statistically significant difference of efficacy between high and low dose of D-Penicillamine in terms of p-value>0.05.