# DATA ANALYSIS FOR A CLINICAL TRIAL

# DATA ANALYSIS FOR A CLINICAL TRIAL OF THE MANAGEMENT OF URINARY TRACT INFECTIONS IN RESIDENTIAL LONG-TERM CARE FACILITIES

By

XIWU LIU

A Project

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2004) McMaster University

(Statistics) Hamilton, Ontario

TITLE: Data Analysis for a Clinical Trial of the Management of Urinary Tract Infections in Residential Long-term Care Facilities

AUTHOR: Xiwu Liu

SUPERVISOR: Professor S. D. Walter

NUMBER OF PAGES: vi, 50

# Abstract

The main object of the research is to analyze the effect of the clinical intervention algorithms proposed for reducing antibiotic use for older adults in long-term care facilities (LTCFs) by managing urinary tract infections (UTIs). 20 paired nursing homes were enrolled in the 12-month study. Within each pair, one was randomized to use of the intervention algorithms and the other to use of regular management. Cluster-level paired t-tests (unweighted and weighted) and regression analyses (unweighted and weighted) were used in the analysis of the data. Paired t-tests show that the algorithms did not significantly reduce the antibiotic use, the number of urine cultures or the antibiotic use for urinary infections in most months. However, they did reduce the proportion of antibiotic use for urinary infections significantly in most months. Regression analysis indicates that the difference between the control group and intervention group has no significant increasing or decreasing trend with time (month). And the algorithms significantly reduced the antibiotic use for urinary infections, number of cultures and the proportions through the 12-month study. The analyses reached a similar conclusion using nonparametric methods and weighted analysis.

# Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Stephen Walter for proving me with the opportunity to work in this research topic, for his expert guidance and support towards the completion of this project.

I would also like to thank to Dr. Mark Loeb and Stephanie Smith for providing the data and explaining the related background. Thanks to Carmine Sidhu for checking grammar errors.

Special thanks to Prof. Peter MacDonald and Prof. Charlie Goldsmith for reading the report and giving their advice and support.

Thanks are extended to all of my instructors for their teaching and guidance. Thanks to my family and friends for their constant support.

# Table of Contents

# Chapter 1

# Introduction

Antibiotics are frequently used for older adults who reside in long-term care facilities (LTCFs). Prospective studies (Nicolle et al., 1983, 1987) showed that antibiotic use for asymptomatic bacteriuria has no benefit and in fact is harmful. Loeb et al. (2002) proposed diagnostic and therapeutic algorithms to reduce the use of antibiotics in residents of LTCFs. The algorithms describe signs and symptoms for which it would be appropriate to send urine cultures or to prescribe antibiotics.

Twenty pairs of Ontario residential LTCFs (nursing homes) were selected for the clinical trail. The nursing homes were matched on number of beds and the proportion of residents having urinary catheters. Within each pair, one was randomized to use intervention algorithms and the other to use regular management. The study was conducted over a 12-month period.

In this study, the nursing home will serve as the cluster of analysis. I will analyze the differences between the control group and intervention group in rates of over-

all antibiotic use (including antibiotic courses per 1000 resident-days and antibiotic defined daily dosage per 1000 resident-days), in rates of antibiotic use for urinary infections (including antibiotic courses for urinary infection per 1000 resident-days and antibiotic defined daily dosage for urinary infection per 1000 resident-days) and in rates of urine cultures (the number of urine cultures per 1000 resident-days). The differences between the two groups in the proportions of antibiotic use for urinary infections (including the antibiotic courses and defined daily dosage) will also be considered.

Paired t-tests and Wilcoxon signed rank tests will be used to analyze the differences for all of the variables above in each research month. Due to the variation of the sample size in each cluster, the methods of weighting paired t-tests (after transforming the data if needed) will be conducted to improve the power of the unweighted paired t-tests. Linear regression and weighted linear regression analysis will be performed to describing the trends of the differences in these variables with the time (month). The analyses will be conducted mainly using S-Plus software in Windows system.

The report is arranged as follows: Chapter 2 demonstrates the research background including significance of the research, algorithms and data description. Chapter 3 introduces the objectives of the analysis and main statistical methods. Chapters 4 and 5 give the processes of the data analysis and corresponding conclusions.

# Chapter 2

# Research Background

In this chapter, I will introduce the research background related to the project. The detailed description of the research significant, intervention algorithms and study population can be seen in Loeb et al. (2002).

## 2.1 Significance

Antibiotics are frequently used for older adults who reside in long-term care facilities (LTCFs). Several risk factors are associated with inappropriate antibiotic use, including developing multi-drug antibiotic resistance and drug-related adverse effects (Nicolle et al., 1996; Strausbaugh et al., 1996; Loeb et al., 2001). Increasing the use of antibiotics also significantly increases costs.

Urinary tract infections are the most common indications for prescribing antibiotics for residents in LTCFs. 30% to 56% of all prescriptions for antibiotics in nursing

homes are prescribed for urinary tract infections (Warren et al., 1991; Zimmer et al., 1986). However, the diagnosis for UTIs is difficult since there are no clear symptoms in this population. Asymptomatic bacteriuria, the presence of bacteria in the urine in the absence of urinary symptoms, occurs in up to 50% of older institutionalized women and 35% of institutionalized older men (Abrutyn et al., 1991; Nicolle et al., 1983). Previous evidence (Nicolle et al., 1987; Ouslander et al., 1995) has shown that antibiotic treatment for asymptomatic bacteriuria has no benefit and in fact is harmful. However, the preliminary study indicated that about one third of all prescriptions for urinary indications are for asymptomatic bacteriuria. Therefore, reducing inappropriate antibiotic use for urinary infections is important for optimizing antibiotic use in LTCFs.

## 2.2 The Intervention Algorithm

A diagnostic and a treatment algorithm (Figure 2.1 and Figure 2.2) for optimizing the antibiotic use in LTCFs was proposed by Loeb et al. (2002).

## 2.3 Introduction of the Data

Twenty pairs of Ontario residential LTCFs (nursing homes) were chosen for this study. Only free standing, community-based residential LTCFs were eligible. The facilities were matched on two variables associated with the use of antibiotics for

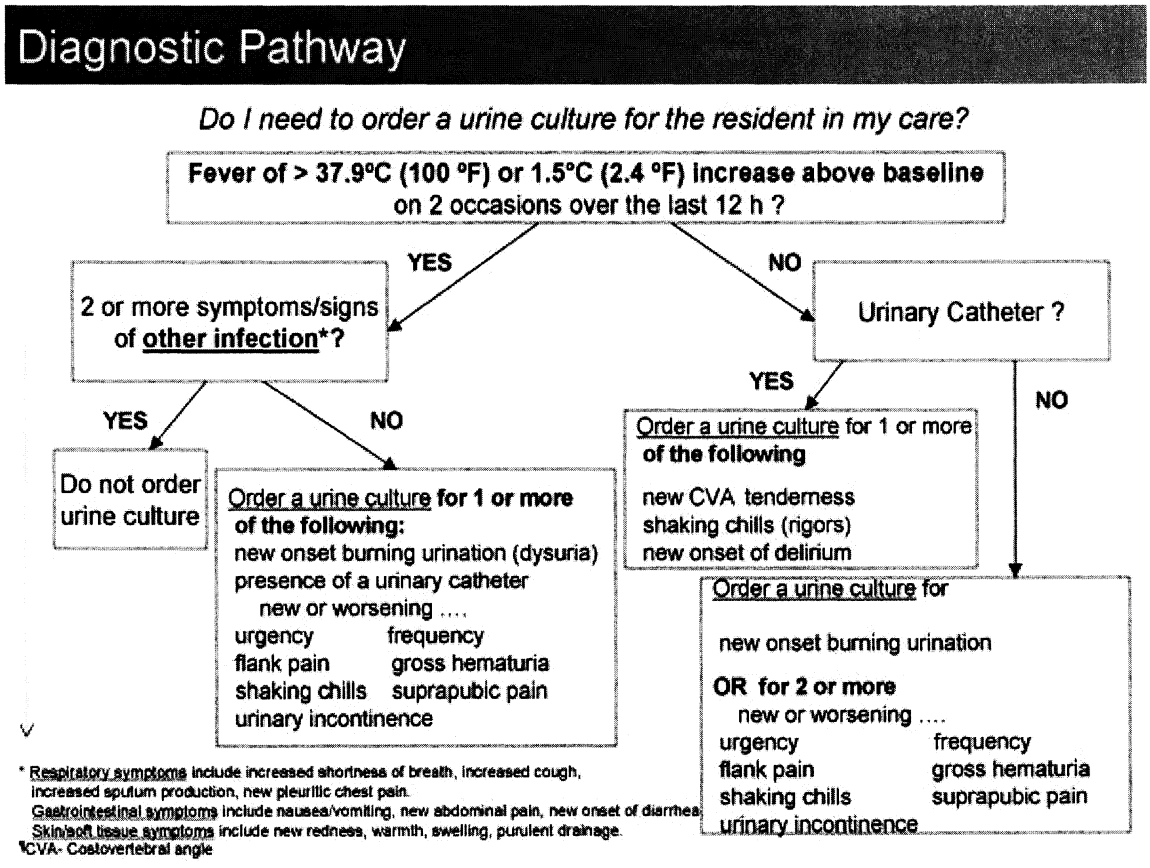


Figure 2.1: Diagnostic algorithm

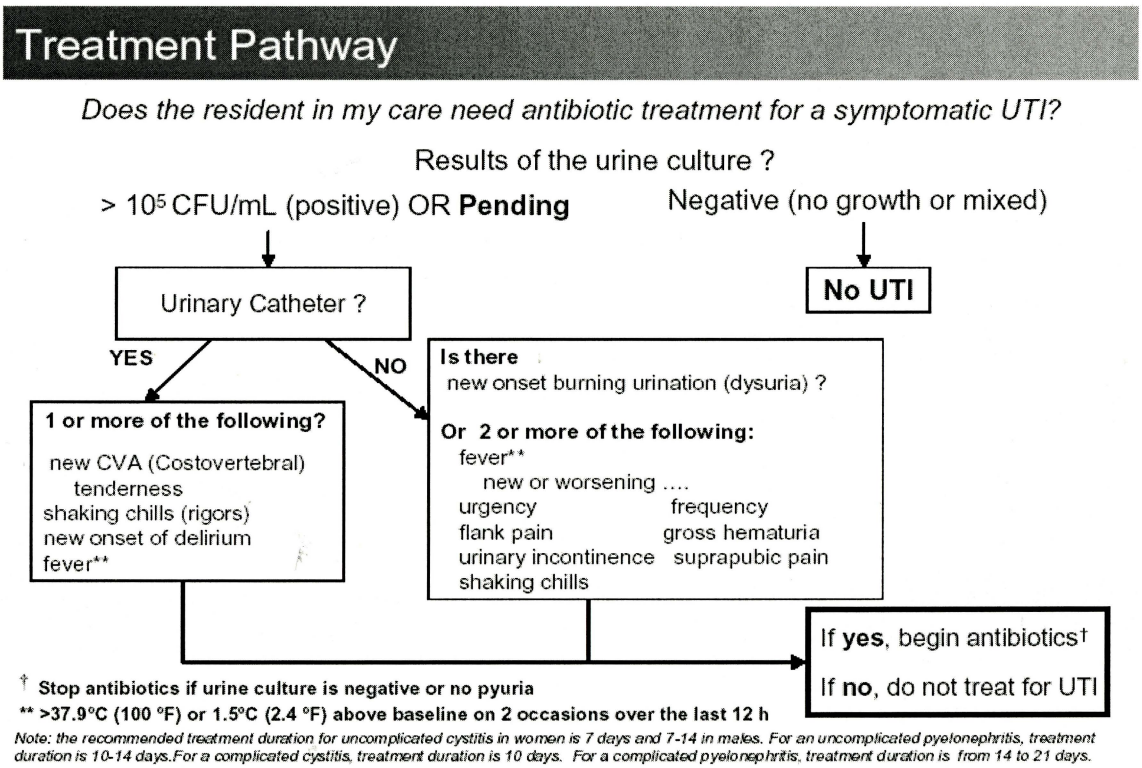


**Figure 2**
Treatment algorithm. This algorithm allows physicians and nurses to optimize antibiotic use in residents with suspected infections.

Figure 2.2: Treatment algorithm

presumptive urinary tract infections: size (number of beds) and the proportion of residents who had indwelling urinary catheters. Within each of the 10 pairs of LTCFs, one was randomized to the intervention (clinical algorithm), the other half to "usual" management.

The study lasted for 12 months. The data collected for each facility included the bed sizes, number of catheterized individuals, resident-days (person-days), the name of the antibiotic, start and stop date, reason for the prescription, as well as urinary symptoms leading to the prescription, whether a urine culture was ordered and the result of the urine culture.

For each facility and each month through the one-year study, the summary data include: 1) the bed sizes and resident-days, 2) the antibiotic courses prescribed for urinary indications, 3) the total number of courses of antibiotics used, 4) urine cultures ordered, 5) the dose of antibiotics used for urinary indications, and 6) the total dose of antibiotics used. For the quantitative component of this study, the nursing home will serve as the cluster of analysis.

# Chapter 3

# Objectives of the Project and Main Statistical Methods

## 3.1 The Objectives of the Project

The primary aim of the original study (Loeb et al., 2002) was to determine if the clinical algorithms for managing urinary tract infections (UTIs) in older adults in residential long-term care facilities (LTCFs) can reduce the overall use of antibiotics in LTCFs.

The related study questions include the following:

- Does the use of the treatment algorithm reduce the number of urine cultures ordered for residents in LTCFs without urinary symptoms?
- Does the use of the treatment algorithm reduce the number of antibiotic courses

and the dose of the antibiotics prescribed for UTIs?

- Does the time (research month) affect the above results through the 12 months?

In this project, I will analyze the differences between the control group and intervention group. The main analysis results will include:

1. Is there a difference between the two groups in rates of overall antibiotic courses used in each month (antibiotic courses (AB) per 1000 resident-days and defined daily dosage (DDD) per 1000 resident-days)?

2. Is there a difference between the two groups in rates of antibiotic use for urinary indications in each month (antibiotic courses per 1000 resident-days for urinary indication (AB-UTI) and defined daily dosage per 1000 resident-days for urinary indication (DDD-UTI))?

3. Is there a difference between the two groups in rates of the number of cultures in each month (the number of cultures per 1000 resident-days)?

4. Is there a difference between the two groups in the proportion of antibiotic dosage for urinary indications in overall antibiotic dosage in each month (DDD-UTI/DDD)?

5. Is there a difference between the two groups in the proportions of antibiotic courses for urinary indications in each month (AB-UTI/AB)?

6. Are there trends in the differences of antibiotic use (including AB, AB-UTI, DDD, DDD-UTI, cultures, AB-UTI/AB and DDD-UTI/DDD) with time?

Paired t-test and Wilcoxon signed rank test will be used to analyze questions 1-5. In addition to these two methods, standard Mantel-Haenszel chi-square test and Liang's test can also be used in question 5. Regression analyses will be conducted to answer question 6. Weighted analyses will be performed to improve the precision of the present analyses. In the next sections, I will introduce these main statistical methods.

## 3.2 Paired t-test for Cluster Randomization Trials

Suppose that $y_{ij}$, where $i = 1, 2$ denotes the control group and intervention group respectively, and $j = 1, 2, \cdots, n$ denotes the cluster, is the rate or proportion of the clusters. They are independent for every $j$, and $y_{1j}$ and $y_{2j}$ are pairwise related for every $j$.

Under the assumption that the paired differences between the two groups $d_j = y_{1j} - y_{2j}$ are normally distributed with common variances, the paired t-test (Donner and Klar, 2002; Montgomery, 1997) can be used to discuss the difference of the two groups.

Let $\mu_1$ and $\mu_2$ be the means of the matched groups $y_{1j}$ and $y_{2j}$ respectively, and $\mu_d = \mu_1 - \mu_2$.

The hypothesis is

$$H_0 : \ \mu_1 = \mu_2 \ \text{vs.} \ H_1 : \ \mu_1 \neq \mu_2,$$

that is equivalent to

$$H_0 : \mu_d = 0 \text{ vs. } H_1 : \mu_d \neq 0.$$

Under the hypothesis $H_0$, the test statistic is

$$t_0 = \frac{\sqrt{n}\bar{d}}{S_d} \sim t_{n-1},$$

where $\bar{d} = \frac{1}{n}\sum_{j=1}^{n} d_j$ (the sample mean of the differences), $S_d = \sqrt{\frac{\sum_{j=1}^{n}(d_j - \bar{d})^2}{n-1}}$ (the sample standard deviation of the differences).

The $(1-\alpha)100\%$ confidence interval for the mean $\mu_d$ is given by

$$\bar{d} \pm t_{1-\alpha/2,n-1} S_d.$$

The main difficulty with this approach is that the underlying assumption, which $d_j$'s are normally distributed with common variances, is not satisfied if there is substantial variation in the cluster size. However, empirical research (Korn, 1984; Donner and Donald, 1987; Gail et al., 1996) suggests that the statistics $t_0$ is fairly robust to departures from the assumption of normal distribution with common variances.

## 3.3 Wilcoxon Signed Rank Test

The paired t-test assumes a normal distribution for the data. However, the assumption is usually not satisfied in practice. In this section, I will introduce a distribution-free test called Wilcoxon signed rank test (Armitage et al., 2002; Donner and Klar, 2002).

The process is :

1. compute $|d_j| = |y_{1j} - y_{2j}|$.

2. order $|d_j|$s from smallest to largest (zero values are ignored) and assign a rank for every one.

3. compute the exact statistic $T_0$ or large sample statistic $T^*$. $T_0$ is denoted by the sum of the ranks assigned to the positive values. $T^*$ is given by

$$T^* = \frac{|T_0 - \frac{1}{4}n(n+1)|}{\sqrt{n(n+1)(2n+1)/24}}.$$

If there are ties among the $|d_j|$s, assign each of the observations in a tied group the average of the integer ranks that are associated with the tied group. After that, compute the $T_0$ and

$$T^* = \frac{|T_0 - \frac{1}{4}n(n+1)|}{\sqrt{Var_0(T_0)}},$$

where $Var_0(T_0) = (24)^{-1}[n(n+1)(2n+1) - \frac{1}{2}\sum_{j=1}^{g} t_j(t_j-1)(t_j+1)]$, where $g$ denotes the number of tied groups of nonzero $|d_j|$s and $t_j$ is the size of tied group $j$.

The $(1-\alpha)100\%$ confidence interval is determined by

$$(W^{(c_\alpha)}, W^{(M+1-c_\alpha)}),$$

where $M = n(n+1)/2$, $W^{(1)} \leq \cdots \leq W^{(M)}$ are the ordered values of the $(d_i + d_j)/2$, $1 \leq i \leq j \leq n$, and

$$c_\alpha = \frac{n(n+1)}{2} + 1 - t_{\alpha/2} \quad \text{for exact test}$$

$$c_\alpha = \frac{n(n+1)}{4} - z_{\alpha/2}\sqrt{\frac{n(n+1)(2n+1)}{24}} \quad \text{for large sample test}$$

where $t_{\alpha/2}$ and $z_{\alpha/2}$ are the upper $\alpha/2$th percentile points of the null distribution of $T_0$ and standard normal distribution respectively.

## 3.4 Standard Mantel-Haenszel Chi-square test and Liang's test

Suppose that the data from a paired cluster randomization trial can be summarized in a series of $n$ $2 \times 2$ contingency tables ($j = 1, 2, \cdots, n$):

| | Control | Intervention | Total |
|---|---|---|---|
| Event | $a_{1j}$ | $a_{2j}$ | $r_{1j}$ |
| No event | $m_{1j} - a_{1j}$ | $m_{2j} - a_{2j}$ | $r_{2j}$ |
| Total | $m_{1j}$ | $m_{2j}$ | $m_j$ |

In addition to using paired t-test and Wilcoxon signed rank test, the difference of the proportions between the two groups can also be tested by Standard Mantel-Haenszel chi-square test (Donner and Klar, 2002; Mantel and Haenszel, 1959) and Liang's test (Donner and Klar, 2002; Liang, 1985; Liang et al., 1986).

Denote the ratios $R_j = p_{1j}(1 - p_{2j})/p_{2j}(1 - p_{1j})$, where $p_{ij} = a_{ij}/m_{ij}$ $(i = 1, 2)$.

The hypothesis is

$$H_0 : \ R = 1 \quad \text{vs} \quad H_1 : \ R \neq 1.$$

Then, Mantel-Haenszel one degree of freedom chi-square statistic is given by

$$\chi^2_{\text{MH}} = \frac{\left( \sum_{j=1}^{n} (a_{1j} - m_{1j} r_{1j}/m_j) \right)^2}{\sum_{j=1}^{n} m_{1j} m_{2j} r_{1j} (m_j - r_{1j}) / \left( m_j^2 (m_j - 1) \right)};$$

Liang's one degree of freedom chi-square statistic is given by

$$\chi^2_{\mathrm{L}} = \frac{\left(\sum_{j=1}^{n}(a_{1j} - m_{1j}r_{1j}/m_j)\right)^2}{\sum_{j=1}^{n}(a_{1j} - m_{1j}r_{1j}/m_j)^2}.$$

The problem of the Mantel-Haenszel chi-square test is that p-value is biased downward in the presence of clustering (Donner and Klar, 2002). Liang's test needs at least 25 pairs to guarantee the precision of the test (Donner and Klar, 2002).

## 3.5 Regression Model

Suppose that a continuous response variable $y_i$ $(i = 1, 2, \cdots, n)$ has normal distribution with homogeneous variance $\sigma^2$, and $x_i$ is the dependent variable. The linear model is given by

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

where the $\varepsilon_i$ is independently and identically distributed $N(0, \sigma^2)$ random variable (Montgomery, 1997; Woolson and Clarke, 2002; Draper and Smith, 1981).

So, the least squares or maximum likelihood estimators of $\beta_0$ and $\beta_1$ are given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

The variances of the estimators $\beta_0$ and $\beta_1$ are given by

$$Var(\hat{\beta}_0) = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right),$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

The covariance of the estimators $\beta_0$ and $\beta_1$ are given by

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\sigma^2 \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

When using the $t$ statistics for testing coefficients as well as other forms of inference on subsets (e.g. confidence interval on mean response, prediction interval, etc.), the parameter $\sigma^2$ can be estimated by the mean squared error

$$S^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{n-2}.$$

## 3.6 Weighted Analysis

In practice, the number of subjects per cluster will vary, either by design or by subject attribute. Therefore, cluster-level analyses which give equal weight to all clusters may be imprecise. Weighted cluster-level analyses should be conducted to account for such variability (Bland and Kerry, 1997; Marubini et al., 1988).

### 3.6.1 Choice of Weights

**Theorem:** Let $t_i$ $(i = 1, 2, \cdots, n)$ be a random variable with the variance $\sigma_i^2$, $\sum_i w_i t_i$ $(w_i > 0)$ be a linear combination of $t_i$s, where $\sum_i w_i = 1$. If $t_i$s are independent or uncorrelated, then the variance of the linear combination $\sum_i w_i t_i$ gets to the

minimum when the weights are proportional to the reciprocal of the variances, that is $w_i/w_j = \sigma_j^2/\sigma_i^2$.

**Proof.**

$$Var(\sum_i w_i t_i) = \sum_i w_i^2 Var(t_i) = \sum_i w_i^2 \sigma_i^2.$$

Setting the partial derivatives with respect to $w_i$s equal to zero, we get

$$w_i \sigma_i^2 = w_n \sigma_n^2,$$

which is equivalent to the quoted condition.

The theory shows that the precision of these analyses is maximized when the cluster-level summary score is weighted inversely proportional to the reciprocal of its variance or estimated variance.

### 3.6.2 Weighted Paired t-test

Define $y_{ij}$ $(i = 1, 2,\ j = 1, 2, \cdots, n)$ and $d_j = y_{1j} - y_{2j}$ to be same as those in section 3.2. $d_j$ may have different variances because the cluster sizes may differ.

Suppose that $d_j$ is $N(\mu_j, \sigma_j^2)$ ($\mu_j$ and $\sigma_j$ are unknown) and $\mu$ is the common value of $\mu_j$.

The weighted sample mean is given by

$$\bar{d} = \frac{\sum_{j=1}^{n} w_j d_j}{\sum_{j=1}^{n} w_j},$$

where $w_j = 1/\sigma_j^2$ chosen by the theorem in the last section.

The weighted sample variance is given by

$$G = \sum_{j=1}^{n} w_j (d_j - \bar{d})^2 / \sum_{j=1}^{n} w_j.$$

The hypothesis is

$$H_0 : \ \mu = 0 \ \text{vs.} \ H_1 : \ \mu \neq 0.$$

So, under the hypothesis $H_0$, the test statistic

$$t_0 = \sqrt{n-1}\bar{d}/\sqrt{G} \sim t_{n-1}.$$

The best estimate of $\mu$ is the weighted mean $\bar{d}$. The $(1-\alpha)100\%$ confidence interval for $\mu$ is given by

$$\bar{d} \pm t_{1-\alpha/2,n-1}\sqrt{G}/\sqrt{n-1}.$$

# Chapter 4

# Analysis of the Data

## 4.1 Comparison of the Means of the Intervention Group and Control Group

Let $y_{ij}$ be the antibiotic courses, urine cultures or antibiotic dosage in cluster $j$ of the group $i$, and $m_{ij}$ be the number of resident-days in cluster $j$ of the group $i$, where $i = 1$ represents the control group and $i = 2$ represents the intervention group and $j = 1, 2, \cdots, n$. The cluster rate is $r_{ij} = 1000 y_{ij}/m_{ij}$, i.e. the rate per 1000 resident-days. Let $t_k$ be the $k$th study month, $k = 1, 2, \cdots, 12$.

### 4.1.1 Plot the data

Firstly, I give a draft picture by plotting the means of the rates $r_{ij}$ of antibiotic courses, urine cultures, antibiotic dosage, antibiotic courses and antibiotic dosage

used for urinary indications respectively vs. the month $t_k$.

The plots (Figure 4.1) show that there are no patterns between the months and the antibiotic use (including AB, AB-UTI, DDD, DDD-UTI and culture). The algorithms did not appear to reduce the overall antibiotic use (AB and DDD). However, the algorithms did appear to reduce the antibiotic use for urinary infection (AB-UTI and DDD-UTI) and the number of urine cultures in most months.

In the next section, I will use the numerical results to check these findings.

### 4.1.2 Numerical Analysis

According to the theory in the last chapter, paired t-test and Wilcoxon signed rank test can be used to analyze the difference of the rates between the control group and intervention group for each month.

The two side paired t-tests for the rates of overall antibiotic courses and dosage indicate that there were no significant differences in statistics between the two groups through the 12-month experiment. Although the plots demonstrate that the rates of antibiotic dosage and courses used for urinary infections and the rate of the number of urine cultures are reduced in most months by using intervention algorithms, the differences between the two groups in these months are still not statistically significant at the level $\alpha = 0.05$.

If $d_j$ does not have a normal distribution, then non-parameter analysis (Wilcoxon signed rank test) can be used to avoid the distribution assumption. Using Wilcoxon

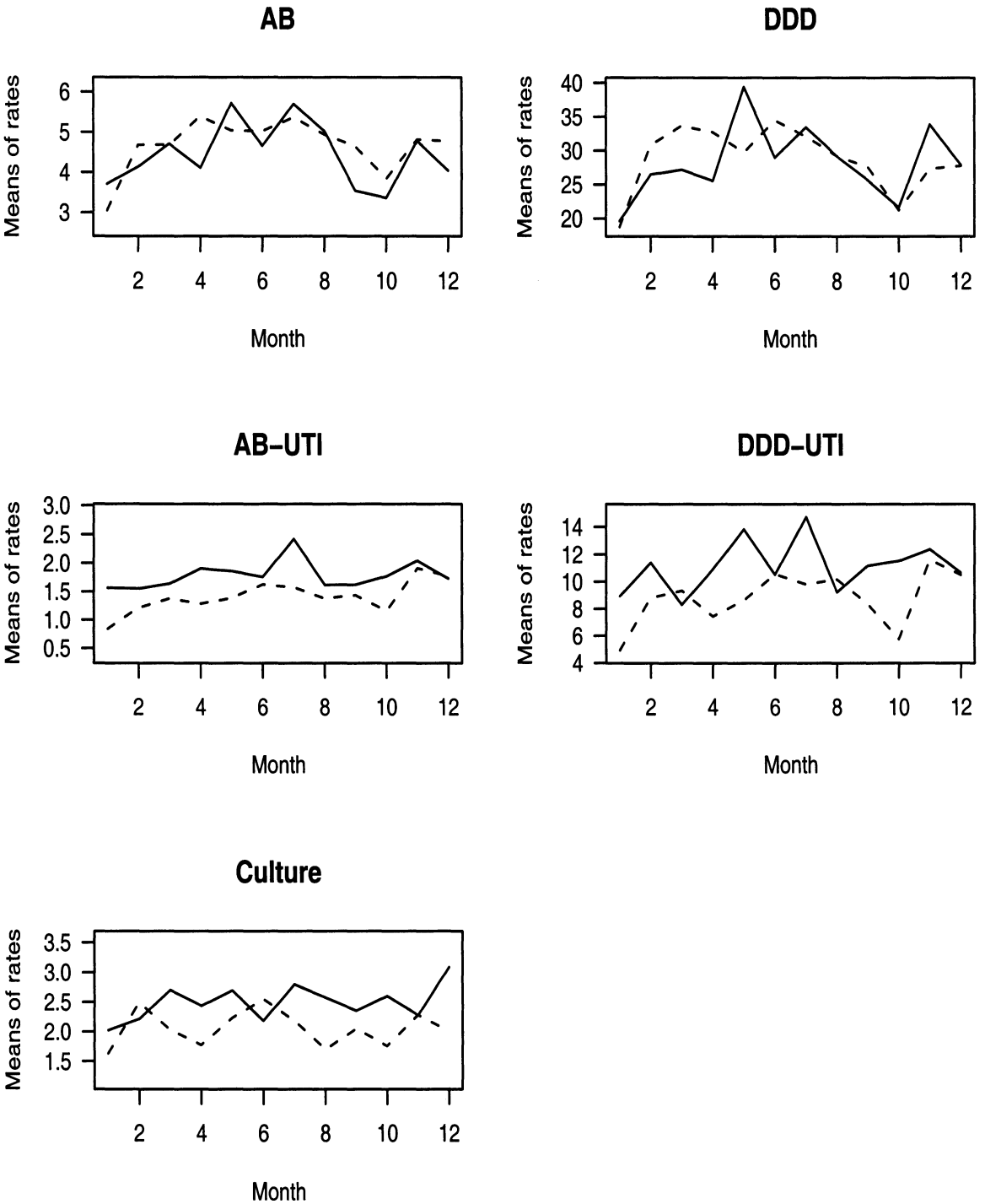


Figure 4.1: Data Plot, dashed line: intervention, solid line: control

signed rank test, I get a similar conclusion.

**Table 4.1. Paired t-test for AB Courses**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.66 | -0.55 | 0.07 | -1.18 | 0.71 | -0.37 | 0.34 | 0.09 | -1.10 | -0.48 | -0.31 | -0.57 |
| Conf.l | -1.12 | -4.81 | -1.70 | -3.26 | -3.47 | -2.47 | -2.17 | -2.63 | -3.06 | -2.03 | -2.88 | -2.97 |
| Conf.h | 2.43 | 3.71 | 1.83 | 0.89 | 4.89 | 1.73 | 2.84 | 2.80 | 0.85 | 1.06 | 2.26 | 1.82 |
| pvalue | 0.42 | 0.78 | 0.93 | 0.22 | 0.71 | 0.70 | 0.77 | 0.94 | 0.23 | 0.50 | 0.79 | 0.60 |

**Table 4.2. Paired t-test for DDD**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.88 | 4.90 | -6.51 | -6.39 | 10.19 | -5.50 | 1.38 | -0.03 | -1.92 | 0.39 | 5.19 | 1.81 |
| Conf.l | -6.48 | -7.60 | -28.08 | -17.42 | -17.99 | -35.42 | -17.42 | -18.77 | -17.01 | -8.19 | -16.02 | -9.63 |
| Conf.h | 8.24 | 17.40 | 15.06 | 4.64 | 38.36 | 24.42 | 20.17 | 18.72 | 13.16 | 8.97 | 26.41 | 13.24 |
| pvalue | 0.79 | 0.39 | 0.51 | 0.22 | 0.43 | 0.69 | 0.87 | 1.00 | 0.78 | 0.92 | 0.59 | 0.73 |

**Table 4.3. Paired t-test for Culture**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.41 | -0.33 | 0.68 | 0.71 | 0.41 | -0.36 | 0.61 | 0.88 | 0.30 | 0.84 | 0.09 | 0.98 |
| Conf.l | -0.70 | -2.42 | -0.25 | -0.18 | -1.47 | -2.08 | -1.08 | -0.23 | -0.80 | -0.15 | -0.57 | -0.85 |
| Conf.h | 1.51 | 1.75 | 1.60 | 1.61 | 2.30 | 1.35 | 2.30 | 1.98 | 1.41 | 1.84 | 0.75 | 2.82 |
| pvalue | 0.42 | 0.72 | 0.13 | 0.10 | 0.63 | 0.64 | 0.43 | 0.11 | 0.55 | 0.09 | 0.77 | 0.25 |

**Table 4.4. Paired t-test for AB-UTI**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.81 | 0.34 | 0.31 | 0.68 | 0.62 | 0.13 | 0.85 | 0.24 | 0.11 | 0.53 | 0.07 | -0.07 |
| Conf.l | -0.15 | -1.28 | -0.33 | -0.72 | -0.41 | -1.06 | -0.32 | -0.85 | -0.67 | -0.45 | -1.22 | -1.14 |
| Conf.h | 1.77 | 1.96 | 0.94 | 2.08 | 1.64 | 1.32 | 2.02 | 1.33 | 0.89 | 1.50 | 1.35 | 1.00 |
| pvalue | 0.09 | 0.65 | 0.30 | 0.29 | 0.20 | 0.81 | 0.14 | 0.63 | 0.75 | 0.25 | 0.90 | 0.88 |

**Table 4.5. Paired t-test for DDD-UTI**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 4.31 | 6.19 | -1.03 | 4.16 | 6.48 | -0.34 | 3.20 | -1.36 | 2.60 | 5.36 | 0.75 | 0.16 |
| Conf.l | -0.77 | -1.97 | -6.76 | -0.57 | -0.97 | -7.90 | -4.00 | -8.67 | -3.84 | -0.96 | -7.03 | -5.02 |
| Conf.h | 9.38 | 14.34 | 4.69 | 8.90 | 13.93 | 7.21 | 10.40 | 5.94 | 9.04 | 11.69 | 8.53 | 5.34 |
| pvalue | 0.09 | 0.12 | 0.69 | 0.08 | 0.08 | 0.92 | 0.34 | 0.68 | 0.38 | 0.09 | 0.83 | 0.94 |

where Est denotes the estimator of the mean of $d_j$s, Conf.l and Conf.h denote the lower bound and higher bound respectively of 95% confidence interval of the mean of $d_j$s.

### 4.1.3 Weighted Analysis

The above paired t-tests are imprecise because the bed sizes vary in the clusters from 101 to 378. In this section, I will use the weighted paired t-test instead of the above analysis to improve the power of the tests.

If the observations $y_{ij}$ have Poisson distributions with the parameter $\lambda_{ij}$ during the $t_{ij}$ resident-days of follow-up, then $E(y_{ij}) = Var(y_{ij}) = \lambda_{ij}t_{ij}$. The variance of the rate $r_{ij} = y_{ij}/t_{ij}$ is equal to $\lambda_{ij}/t_{ij}$. For the paired cluster analysis in this research, $y_{1j}$ and $y_{2j}$ are uncorrelated. Therefore, the $Var(d_j) = Var(r_{1j} - r_{2j}) = Var(r_{1j}) + Var(r_{2j}) = \lambda_{1j}/t_{1j} + \lambda_{2j}/t_{2j}$. I use the sample means of the rates to estimate the $\lambda_{ij}$s here since $E(r_{ij}) = E(y_{ij}/t_{ij}) = \lambda_{ij}$. So the weight can be given by the reciprocal of $r_{1j}/t_{1j} + r_{2j}/t_{2j}$.

If the observations $y_{ij}$ have continuous distributions, then the variances of $r_{ij}$ are

often proportional to the reciprocal of the cluster sizes $n_{ij}$. Larger cluster size will have smaller variance. Since $r_{1j}$ and $r_{2j}$ are uncorrelated in the paired cluster study and they are matched according to the cluster size, $Var(d_j) = Var(r_{1j}) + Var(r_{2j}) \propto \frac{1}{n_{1j}} + \frac{1}{n_{2j}}$. Therefore, the empirical weights can be given by $\frac{1}{\frac{1}{n_{1j}} + \frac{1}{n_{2j}}}$.

The number of urine cultures and antibiotic courses (AB and AB-UTI) have Poisson distributions. So, the weights are given by the reciprocal of $r_{1j}/t_{1j} + r_{2j}/t_{2j}$. The antibiotic dosage (DDD and DDD-UTI) have continuous outcome. So, the weights are given by $\frac{1}{\frac{1}{n_{1j}} + \frac{1}{n_{2j}}}$, where $n_{ij}$ is the sample size (the number of beds) of the cluster $ij$. Weighting the paired t-tests gets the following results.

**Table 4.6. Weighted Paired t-test for AB Courses**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.20 | -0.14 | -0.82 | -0.95 | -0.36 | 0.05 | 0.06 | 0.42 | -1.14 | -0.23 | -0.57 | -0.72 |
| Conf.l | -1.09 | -2.48 | -3.03 | -2.74 | -3.32 | -1.41 | -2.43 | -2.07 | -3.02 | -1.60 | -2.88 | -2.68 |
| Conf.h | 1.49 | 2.20 | 1.38 | 0.83 | 2.61 | 1.50 | 2.55 | 2.90 | 0.73 | 1.15 | 1.73 | 1.25 |
| pvalue | 0.73 | 0.90 | 0.41 | 0.25 | 0.79 | 0.95 | 0.96 | 0.71 | 0.20 | 0.72 | 0.58 | 0.42 |

**Table 4.7. Weighted Paired t-test for AB-UTI Courses**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.30 | 0.35 | -0.04 | 0.16 | 0.49 | 0.25 | 0.88 | 0.19 | 0.25 | 0.33 | 0.03 | 0.09 |
| Conf.l | -0.30 | -0.47 | -0.70 | -0.50 | -0.48 | -0.63 | -0.26 | -0.55 | -0.54 | -0.58 | -1.27 | -0.92 |
| Conf.h | 0.90 | 1.18 | 0.62 | 0.83 | 1.46 | 1.14 | 2.03 | 0.93 | 1.04 | 1.24 | 1.33 | 1.10 |
| pvalue | 0.28 | 0.35 | 0.88 | 0.57 | 0.27 | 0.53 | 0.11 | 0.58 | 0.48 | 0.43 | 0.96 | 0.84 |

**Table 4.8. Weighted Paired t-test for Urine Culture**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.35 | 0.26 | 0.79 | 0.71 | 0.40 | 0.02 | 0.68 | 0.94 | 0.08 | 0.58 | 0.16 | 0.79 |
| Conf.l | -0.76 | -1.32 | -0.02 | -0.22 | -1.27 | -1.40 | -0.77 | -0.17 | -0.86 | -0.32 | -0.44 | -0.86 |
| Conf.h | 1.46 | 1.84 | 1.60 | 1.64 | 2.06 | 1.44 | 2.13 | 2.05 | 1.02 | 1.47 | 0.75 | 2.43 |
| pvalue | 0.49 | 0.71 | 0.05 | 0.12 | 0.60 | 0.98 | 0.31 | 0.09 | 0.84 | 0.18 | 0.56 | 0.30 |

**Table 4.9. Weighted Paired t-test for DDD**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.09 | 3.14 | -5.71 | -4.60 | 9.35 | -0.87 | 4.83 | 4.71 | 0.49 | 2.27 | 6.65 | 2.58 |
| Conf.l | -7.08 | -7.70 | -25.91 | -15.26 | -17.56 | -29.15 | -15.58 | -13.56 | -15.36 | -6.00 | -12.34 | -8.67 |
| Conf.h | 7.27 | 13.97 | 14.49 | 6.06 | 36.25 | 27.41 | 25.24 | 22.97 | 16.34 | 10.54 | 25.63 | 13.84 |
| pvalue | 0.98 | 0.52 | 0.54 | 0.35 | 0.45 | 0.95 | 0.61 | 0.57 | 0.95 | 0.55 | 0.44 | 0.61 |

**Table 4.10. Weighted Paired t-test for DDD-UTI**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 3.45 | 5.03 | -0.82 | 4.19 | 6.68 | 0.10 | 3.83 | -0.15 | 4.11 | 6.00 | 1.76 | 1.80 |
| Conf.l | -0.63 | -2.25 | -6.09 | -0.40 | -1.10 | -6.48 | -3.52 | -7.30 | -3.04 | -0.32 | -6.57 | -3.85 |
| Conf.h | 7.52 | 12.31 | 4.45 | 8.78 | 14.45 | 6.67 | 11.17 | 7.00 | 11.26 | 12.31 | 10.08 | 7.45 |
| pvalue | 0.09 | 0.15 | 0.73 | 0.07 | 0.08 | 0.97 | 0.26 | 0.96 | 0.22 | 0.06 | 0.64 | 0.49 |

Compared with the unweighted case, the differences between the two groups are not significant at the level $\alpha = 0.05$. However, the confidence intervals of the means are much shorter than unweighted case. This indicates that the power of the tests has been improved.

The qqplots (see appendix) for the differences between the two groups after weighting indicate that most differences look normal, only a few of them depart from the

normal distribution moderately. By doing log transformation prior to taking differences, the qqplots do not have much improvement. Since the paired t-test is robust to departures from the normality assumption, the analysis results are valid even for some departure data.

## 4.2 Comparison of the Proportions of the Control Group and Intervention Group

The above analysis demonstrates that the algorithms did not reduce the antibiotic use (including AB, AB-UTI, DDD, DDD-UTI and culture) significantly in most months. Can the algorithms reduce the proportions of the antibiotic used for urinary tract infections (AB-UTI and DDD-UTI) in the overall antibiotic use (AB and DDD)? In this section, I will answer this question.

### 4.2.1 Plot the Proportions

The plot (Figure 4.2) of the mean of AB-UTI/AB shows that the algorithms reduce the proportions in all 12 months. The proportions in the control group fluctuate in the 12-month period, but the trend of the proportions in the intervention group seem to be increasing with the time. The plot (Figure 4.2) of the mean of DDD-UTI/DDD shows that the algorithm seem to reduce the proportions except for the 6th and 11th month. There is an increasing trend in the intervention group, however it is not

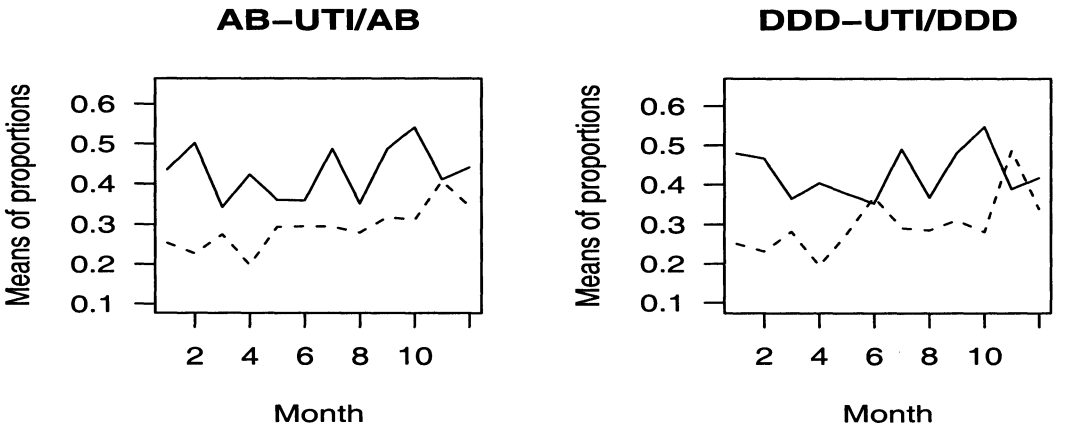


Figure 4.2: Data Plot, dashed line: intervention, solid line: control

statistically significant.

### 4.2.2 Numerical Analysis

Let $p_{ij}$ be the proportion (AB-UTI/AB and DDD-UTI/DDD), ($i = 1, 2$ denote the control and intervention group respectively, $j = 1, 2, \cdots, 10$ denote the clusters), $d_j = p_{1j} - p_{2j}$ be the paired difference.

Antibiotic defined daily dosage for urinary infections (DDD-UTI) has a continuous outcome. According to the theory in Chapter 3, paired t-test and Wilcoxon signed rank test can be used to compare the proportions of the two groups (DDD-UTI/DDD). The tests indicate that the two groups are significantly different in the 1st, 2nd and 4th month at the level of $\alpha = 0.05$, are also significant in the 7th and 10th month at the level of $\alpha = 0.1$.

Antibiotic courses for urinary infections (AB-UTI) have a binomial distribution.

According to the theory in Chapter 3, Mantel-Haenszel test, Liang's test, paired t-test and Wilcoxon signed rank test can be used to compare the proportion of the two groups. Comparing these methods, Brookmeyer and Chen (1998) indicated that the t-test (perhaps after a suitable log or other transformation) performs well in many situations with as low as five pairs. Considering missing data that lead to fewer pairs than the original 10 pairs, I will mainly use paired t-tests to analyze the differences of the two groups here for AB-UTI/AB. The paired t-tests for AB-UTI/AB get similar results as those of DDD-UTI/DDD.

**Table 4.11. Paired t-test for DDD-UTI/DDD**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.22 | 0.25 | 0.08 | 0.21 | 0.14 | -0.03 | 0.17 | 0.08 | 0.18 | 0.26 | -0.07 | 0.07 |
| Conf.l | 0.05 | 0.03 | -0.06 | 0.12 | -0.10 | -0.26 | -0.03 | -0.15 | -0.16 | -0.03 | -0.30 | -0.11 |
| Conf.h | 0.38 | 0.47 | 0.23 | 0.30 | 0.38 | 0.20 | 0.37 | 0.32 | 0.51 | 0.55 | 0.15 | 0.26 |
| pvalue | 0.02 | 0.03 | 0.22 | 0.00 | 0.23 | 0.78 | 0.08 | 0.44 | 0.26 | 0.07 | 0.48 | 0.39 |

**Table 4.12. Paired t-test for AB-UTI/AB**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.17 | 0.27 | 0.08 | 0.22 | 0.09 | 0.06 | 0.19 | 0.07 | 0.16 | 0.22 | 0.02 | 0.08 |
| Conf.l | 0.02 | 0.06 | -0.02 | 0.08 | -0.13 | -0.08 | -0.01 | -0.11 | -0.16 | -0.07 | -0.15 | -0.04 |
| Conf.h | 0.33 | 0.49 | 0.18 | 0.35 | 0.31 | 0.21 | 0.39 | 0.26 | 0.47 | 0.51 | 0.20 | 0.21 |
| pvalue | 0.03 | 0.02 | 0.12 | 0.01 | 0.37 | 0.35 | 0.06 | 0.41 | 0.29 | 0.12 | 0.76 | 0.17 |

### 4.2.3 Weighted Analysis

Since the sample sizes vary from 101 to 378 in clusters, weights should be given to improve the power of the tests.

Suppose that the antibiotic courses for urinary tract infection $y_{ij}$s have binomial distributions with the parameters $p_{ij}$ and $m_{ij}$ ($m_{ij}$ is the overall antibiotic courses of cluster $ij$). So, $Var(y_{ij}) = m_{ij}p_{ij}(1-p_{ij})$. The variances of the proportions $\hat{p}_{ij} = y_{ij}/m_{ij}$ are equal to $p_{ij}(1-p_{ij})/m_{ij}$. Since the $\hat{p}_{1j}$ and $\hat{p}_{2j}$ are uncorrelated, the variances of the differences $\hat{p}_{1j} - \hat{p}_{2j}$ are $p_{1j}(1-p_{1j})/m_{1j} + p_{2j}(1-p_{2j})/m_{2j}$. So, the empirical weights can be given by the reciprocal of $1/m_{1j} + 1/m_{2j}$. Since the overall antibiotic courses are proportional to the sample sizes $n_{ij}$ (number of beds) in each cluster and $y_{1j}$ and $y_{2j}$ are paired by the sample sizes, the overall antibiotic courses $m_{ij}$ can be replaced by the sample sizes $n_{ij}$ in the weights. Therefore, the weights can be given by the reciprocal of $1/n_{1j} + 1/n_{2j}$.

The proportions of antibiotic dosage have a continuous distribution in $[0, 1]$. According to the analysis in section 2.5, the empirical weights can be given by the reciprocal of $1/n_{1j} + 1/n_{2j}$.

The weighted paired t-tests for the proportions of the antibiotic courses indicate that there are significant differences between the two groups in the 2nd, 4th and 7th month at the level of $\alpha = 0.05$, and the differences are significant in the 1st and 12th month at the level of $\alpha = 0.1$.

The tests of the proportions of the antibiotic daily dosage show that there are significant differences between the two groups in the 1st, 2nd, 4th and 7th month at the level of $\alpha = 0.05$, and the differences are significant in the 10th month at the level of $\alpha = 0.1$.

The comparison of the unweighted paired t-tests with weighted paired t-test indi-

cates that the weighted tests get shorter confidence intervals of the means. Therefore, using weights in the paired t-tests can improve the precision of the tests.

The weighted qqplots (see appendix) of the differences between the two groups also show that the assumption of normal distribution is appropriate.

**Table 4.13. Weighted Paired t-test for AB-UTI/AB**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.13 | 0.22 | 0.08 | 0.19 | 0.13 | 0.10 | 0.23 | 0.08 | 0.16 | 0.20 | 0.05 | 0.11 |
| Conf.l | -0.02 | 0.00 | -0.03 | 0.05 | -0.07 | -0.05 | 0.05 | -0.10 | -0.12 | -0.09 | -0.12 | -0.02 |
| Conf.h | 0.29 | 0.45 | 0.20 | 0.33 | 0.32 | 0.25 | 0.41 | 0.25 | 0.44 | 0.50 | 0.22 | 0.25 |
| pvalue | 0.08 | 0.05 | 0.12 | 0.01 | 0.18 | 0.18 | 0.02 | 0.34 | 0.22 | 0.15 | 0.54 | 0.09 |

**table 4.14. Weighted Paired t-test for DDD-UTI/DDD**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.19 | 0.23 | 0.11 | 0.19 | 0.18 | -0.05 | 0.18 | 0.09 | 0.17 | 0.25 | -0.07 | 0.14 |
| Conf.l | 0.04 | 0.01 | -0.06 | 0.10 | -0.06 | -0.29 | 0.01 | -0.13 | -0.12 | -0.05 | -0.28 | -0.06 |
| Conf.h | 0.34 | 0.45 | 0.29 | 0.29 | 0.41 | 0.18 | 0.35 | 0.31 | 0.47 | 0.55 | 0.15 | 0.35 |
| pvalue | 0.02 | 0.04 | 0.17 | 0.00 | 0.12 | 0.61 | 0.04 | 0.36 | 0.21 | 0.09 | 0.50 | 0.15 |

### 4.2.4 Transformation of the data

The proportions $p_{ij}$s only take the values in $[0, 1]$. So, the values $d_j = p_{1j} - p_{2j}$ are in $[-1, 1]$, that means that $d_j$ can not be a normal distribution. Perhaps, taking log transformation to $p_{ij}$, then taking the weighted paired t-test to the paired data $\log(p_{1j})$ and $\log(p_{2j})$ can improve the power of the tests. The transformation is particularly suitable for the trials involving large clusters, where simulation studies

suggest that the transformation is safely applied to the trials involving as few as six matched pairs (Donner and Donald, 1987).

Comparing with the previous analysis, I get the similar conclusions.

**Table 4.15. Weighted Paired t-test for AB-UTI/AB after Log Transformation**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.40 | 0.71 | 0.36 | 0.64 | 0.43 | 0.21 | 0.65 | 0.17 | 0.43 | 0.47 | 0.13 | 0.32 |
| Conf.l | -0.06 | 0.08 | -0.12 | 0.05 | -0.16 | -0.45 | 0.22 | -0.42 | -0.30 | -0.47 | -0.38 | -0.10 |
| Conf.h | 0.86 | 1.35 | 0.84 | 1.24 | 1.02 | 0.88 | 1.08 | 0.77 | 1.16 | 1.40 | 0.63 | 0.73 |
| pvalue | 0.08 | 0.03 | 0.12 | 0.04 | 0.13 | 0.49 | 0.01 | 0.53 | 0.21 | 0.28 | 0.58 | 0.12 |

**Table 4.16. Weighted Paired t-test for DDD-UTI/DDD after Log Transformation**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.57 | 0.83 | 0.30 | 0.69 | 0.83 | -0.20 | 0.52 | 0.19 | 0.51 | 0.65 | -0.21 | 0.52 |
| Conf.l | 0.17 | 0.13 | -0.21 | 0.38 | -0.18 | -1.17 | 0.17 | -0.66 | -0.39 | -0.27 | -0.79 | -0.20 |
| Conf.h | 0.98 | 1.52 | 0.81 | 1.01 | 1.84 | 0.78 | 0.87 | 1.05 | 1.42 | 1.56 | 0.37 | 1.25 |
| pvalue | 0.01 | 0.03 | 0.21 | 0.00 | 0.10 | 0.65 | 0.01 | 0.61 | 0.23 | 0.14 | 0.43 | 0.13 |

However, the qqplots (see appendix) do not demonstrate that the power of the tests have been significantly improved after the transformation.

## 4.3 Multiple Comparison Adjusting

I carried out multiple tests for all variables (including the rates and proportions, weighted tests and unweighted tests):

$H_1$ vs $A_1$ with p-value $p_1$,

$H_2$ vs $A_2$ with p-value $p_2$,

suggest that the transformation is safely applied to the trials involving as few as six matched pairs (Donner and Donald, 1987).

Comparing with the previous analysis, I get the similar conclusions.

**Table 4.15. Weighted Paired t-test for AB-UTI/AB after Log Transformation**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.40 | 0.71 | 0.36 | 0.64 | 0.43 | 0.21 | 0.65 | 0.17 | 0.43 | 0.47 | 0.13 | 0.32 |
| Conf.l | -0.06 | 0.08 | -0.12 | 0.05 | -0.16 | -0.45 | 0.22 | -0.42 | -0.30 | -0.47 | -0.38 | -0.10 |
| Conf.h | 0.86 | 1.35 | 0.84 | 1.24 | 1.02 | 0.88 | 1.08 | 0.77 | 1.16 | 1.40 | 0.63 | 0.73 |
| pvalue | 0.08 | 0.03 | 0.12 | 0.04 | 0.13 | 0.49 | 0.01 | 0.53 | 0.21 | 0.28 | 0.58 | 0.12 |

**Table 4.16. Weighted Paired t-test for DDD-UTI/DDD after Log Transformation**

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Est | 0.57 | 0.83 | 0.30 | 0.69 | 0.83 | -0.20 | 0.52 | 0.19 | 0.51 | 0.65 | -0.21 | 0.52 |
| Conf.l | 0.17 | 0.13 | -0.21 | 0.38 | -0.18 | -1.17 | 0.17 | -0.66 | -0.39 | -0.27 | -0.79 | -0.20 |
| Conf.h | 0.98 | 1.52 | 0.81 | 1.01 | 1.84 | 0.78 | 0.87 | 1.05 | 1.42 | 1.56 | 0.37 | 1.25 |
| pvalue | 0.01 | 0.03 | 0.21 | 0.00 | 0.10 | 0.65 | 0.01 | 0.61 | 0.23 | 0.14 | 0.43 | 0.13 |

However, the qqplots (see appendix) do not demonstrate that the power of the tests have been significantly improved after the transformation.

## 4.3 Multiple Comparison Adjusting

I carried out multiple tests for all variables (including the rates and proportions, weighted tests and unweighted tests):

$H_1$ vs $A_1$ with p-value $p_1$,

$H_2$ vs $A_2$ with p-value $p_2$,



and $k = 1, 2, \cdots, 12$ denote the months. $d_{jk} = r_{1jk} - r_{2jk}$ denote the paired difference

of the rates between the two groups for jth cluster and kth month.

Let $t_k$s be the months and $\bar{d}_k$s be the means of $d_{jk}$s, then the linear model is given by

$$\bar{d}_k = \beta_0 + \beta_1 t_k + \varepsilon_k$$

where $\varepsilon_k$ has normal distribution $N(0, \sigma^2)$.

The plots of AB, AB-UTI, DDD, DDD-UTI and Culture show that there are no significant straight line relationships between the time and the means of the differences between the two groups. When using straight line models to fit them, we see that the differences are not significant for AB, DDD and Culture through the year, and there are no significantly decreasing or increasing trends with the time for AB, DDD, DDD-UTI and Culture. However, the differences between the groups are significant for AB-UTI and DDD-UTI through the year. There is a significant decreasing trend with time for AB-UTI. The qqplots of the residuals indicate that the assumptions that the residuals are normally distributed is acceptable.

**Table 4.17. Regression**

| | AB | | AB-UTI | | DDD | | DDD-UTI | | Culture | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Int | Mon | Int | Mon | Int | Mon | Int | Mon | Int | Mon |
| Estimate | 0.14 | -0.06 | 0.70 | -0.05 | -0.99 | 0.21 | 4.30 | -0.27 | 0.17 | 0.04 |
| Std. Error | 0.38 | 0.05 | 0.16 | 0.02 | 3.20 | 0.43 | 1.71 | 0.23 | 0.27 | 0.04 |
| t value | 0.36 | -1.10 | 4.29 | -2.19 | -0.31 | 0.48 | 2.51 | -1.16 | 0.63 | 1.09 |
| Pr(> \|t\|) | 0.72 | 0.30 | 0.00 | 0.05 | 0.76 | 0.64 | 0.03 | 0.27 | 0.54 | 0.30 |

where "Int" and "Mon" denote the intercept and coefficient of each regression straight line.

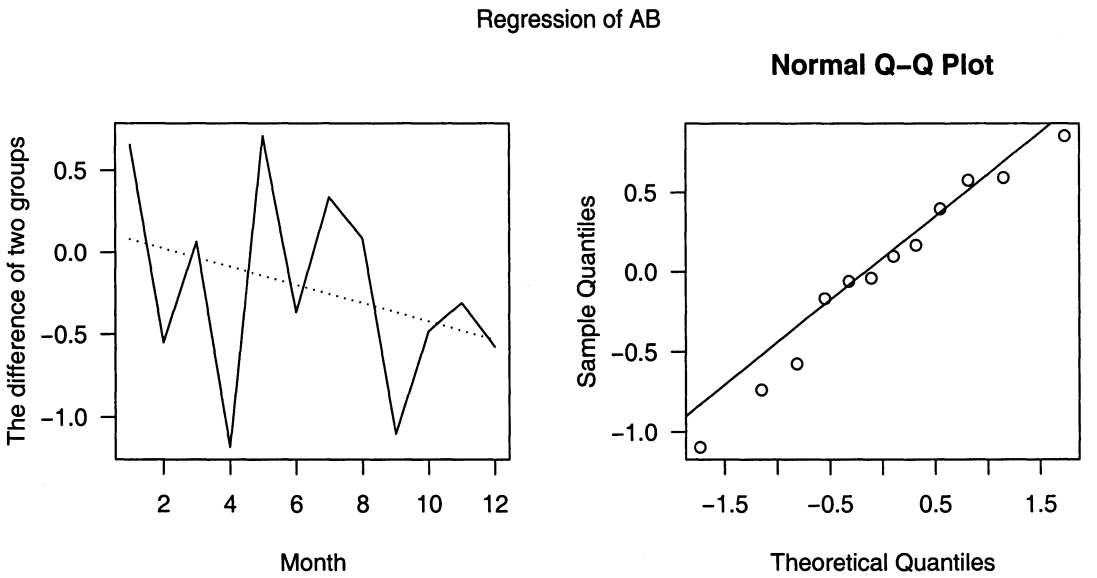


Figure 4.3: Regression plot and qqplot of residuals for AB

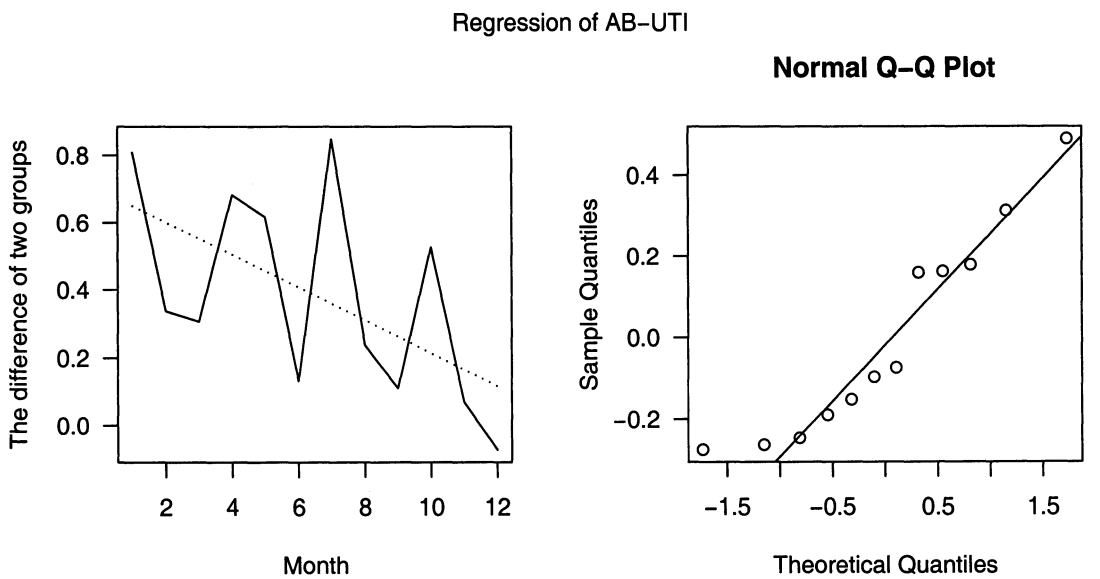


Figure 4.4: Regression plot and qqplot of residuals for AB-UTI

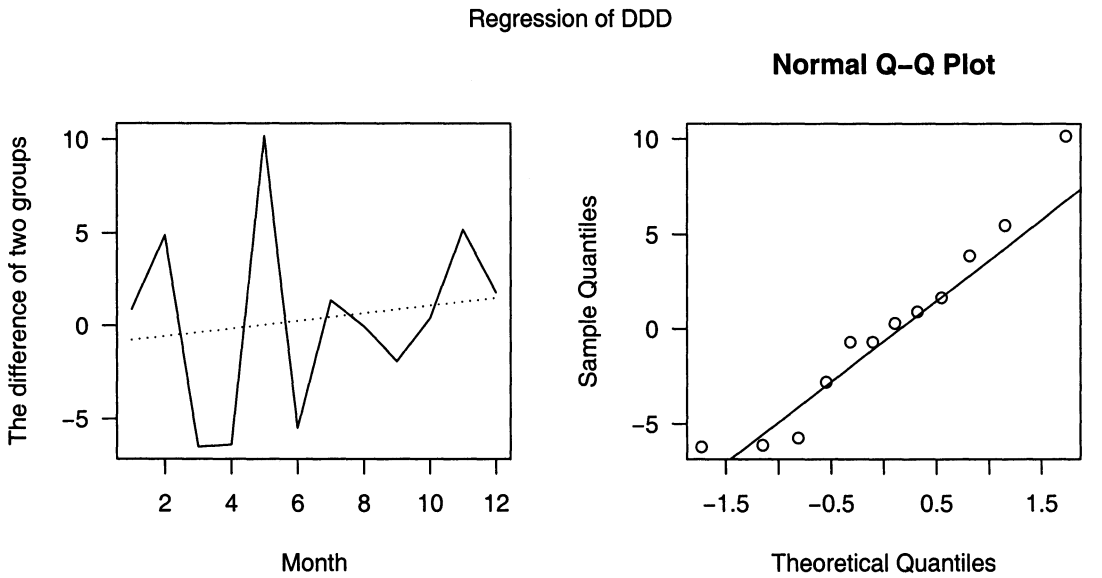


Figure 4.5: Regression plot and qqplot of residuals for DDD

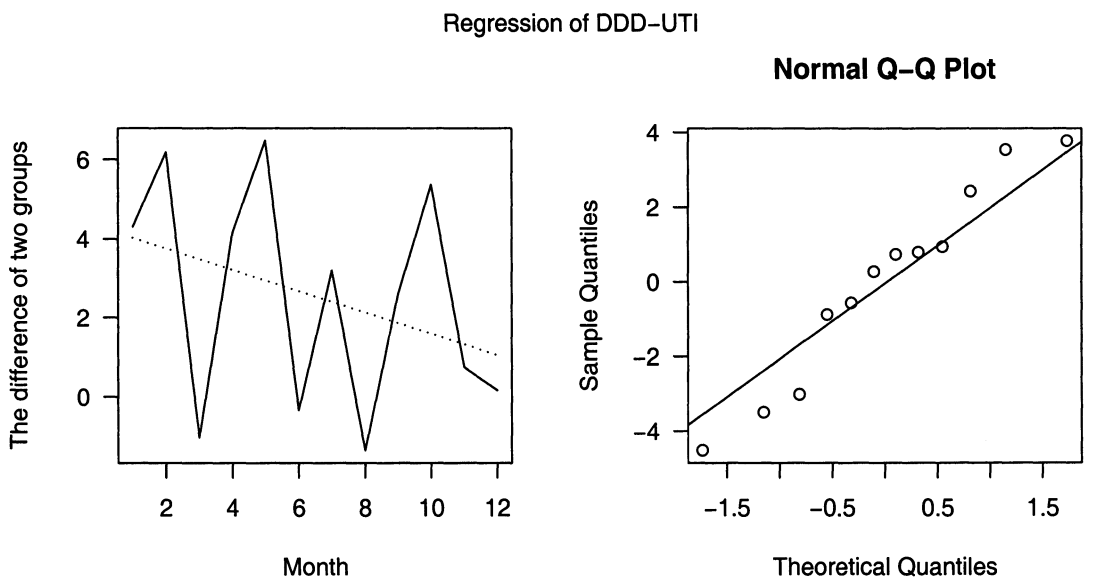


Figure 4.6: Regression plot and qqplot of residuals for DDD-UTI

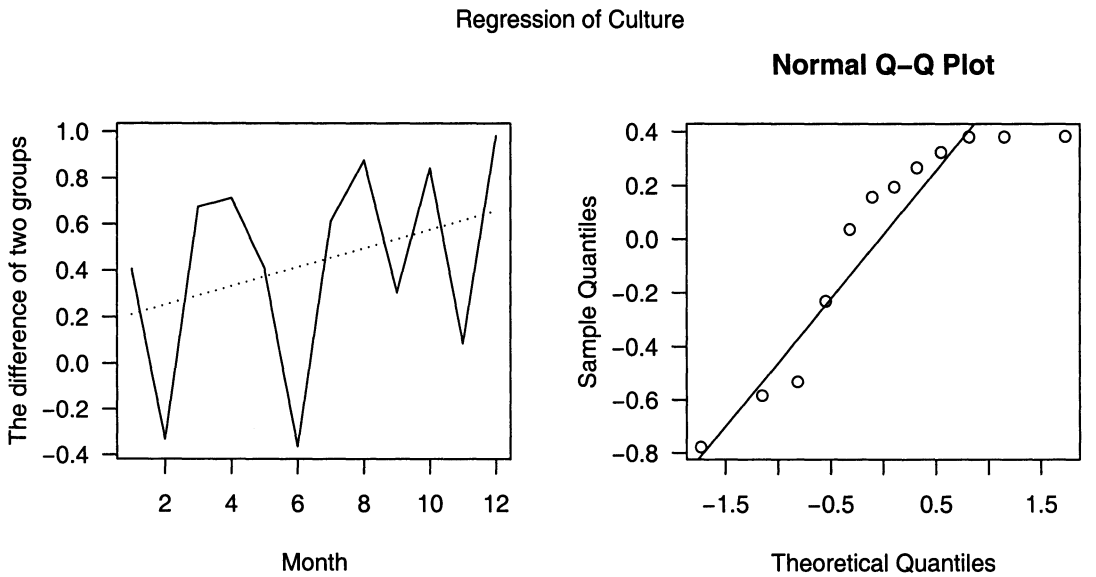


Figure 4.7: Regression plot and qqplot of residuals for Culture

### 4.4.2 Proportion

In this section, I will discuss the relationships between the time and the differences of the groups for the proportions (AB-UTI/AB and DDD-UTI/DDD).

Let $p_{ijk}$ be the proportion (AB-UTI/AB or DDD-UTI/DDD) and $d_{jk} = p_{1jk} - p_{2jk}$ be the paired difference. Then the linear model is

$$\bar{d}_k = \beta_0 + \beta_1 t_k + \varepsilon_k$$

where $\varepsilon_k$ has normal distribution $N(0, \sigma^2)$.

Fitting the data by the straight-line models indicates that the differences between the two groups for both of the proportions are significant. Both of them have a decreasing trend with time. But the trends are not significant statistically. The qqplots of the residuals show that the residuals slightly depart from normality.

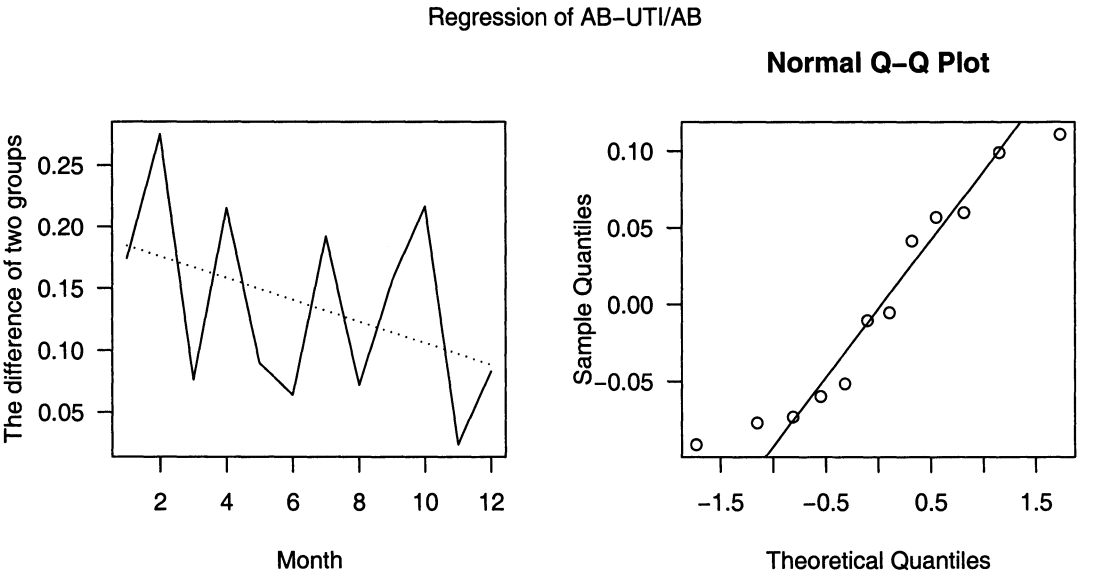


Figure 4.8: Regression plot and qqplot of residuals for AB-UTI/AB

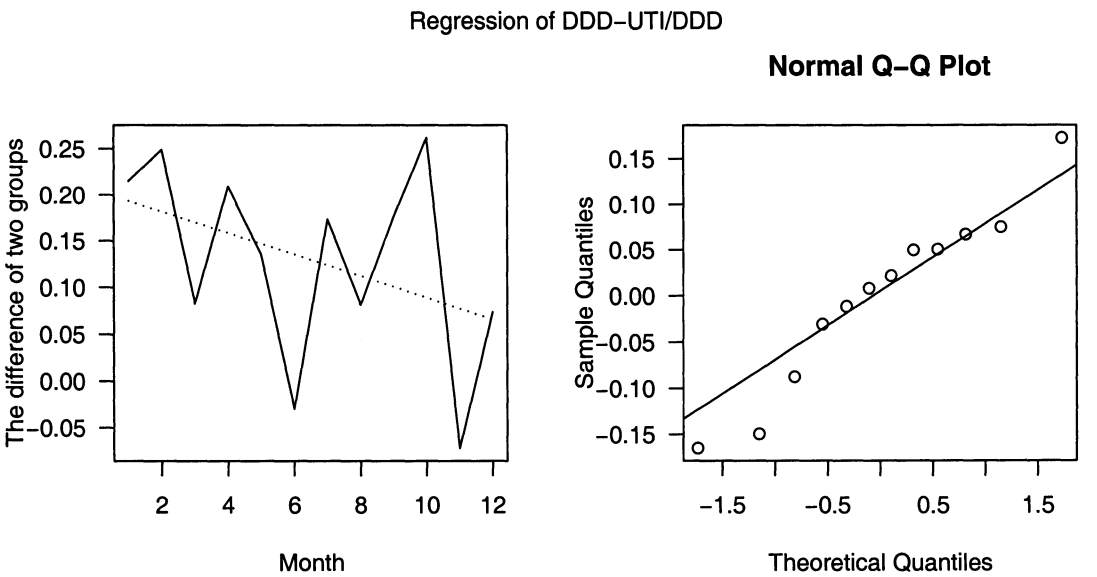


Figure 4.9: Regression plot and qqplot of residuals for DDD-UTI/DDD

**Table 4.18. Regression for Proportion**

| | AB-UTI/AB | | DDD-UTI/DDD | |
|---|---|---|---|---|
| Variable | Int | Mon | Int | Mon |
| Estimate | 0.19 | -0.01 | 0.21 | -0.01 |
| Std. Error | 0.05 | 0.01 | 0.06 | 0.01 |
| t value | 4.18 | -1.39 | 3.27 | -1.36 |
| Pr(> \|t\|) | 0.00 | 0.20 | 0.01 | 0.20 |

### 4.4.3 Weighted Analysis

In the last section, I used regular means of the paired differences between the two groups as the response variables in the straight-line models. However, the analysis is imprecise since the sizes of the clusters are different. In this section, regular means will be replaced by the weighted means in the analysis. Let the weighted means of $d_{jk}$ be $\bar{dw}_{jk} = \frac{\sum_j w_{jk} d_{jk}}{\sum_j w_{jk}}$, then the straight-line model is

$$\bar{dw}_k = \beta_0 + \beta_1 t_k + \varepsilon_k$$

where $\varepsilon_k$ has normal distribution $N(0, \sigma^2)$.

According to the analysis in section 3.5 and section 4.2.3, I choose the same weights as before, which are the reciprocal of $r_{1j}/t_{1j} + r_{2j}/t_{2j}$ for AB, AB-UTI and Culture and the reciprocal of $1/n_{1j} + 1/n_{2j}$ for others.

All of the weighted regression plots show that there are no significant linear relationships between the time and the means of the difference of the two groups. Compared to unweighted cases, the effects of the differences between the two groups throughout the year have no significant change except for Culture, which turns out to be significant statistically. There are no significantly decreasing or increasing trends with the time for all variables (including AB, AB-UTI, DDD, DDD-UTI, Culture,

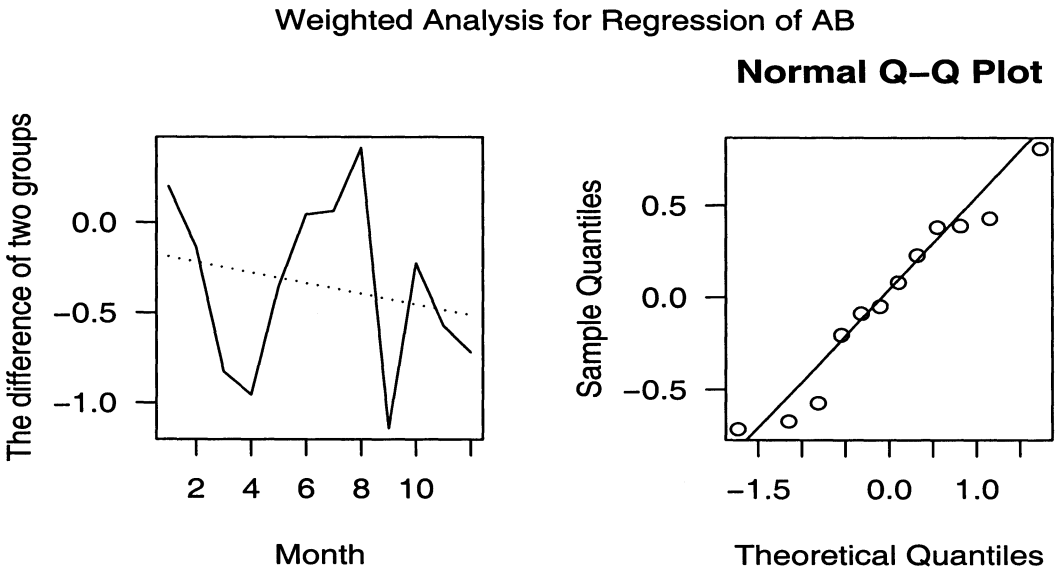


Figure 4.10: Weighted regression plot and qqplot of residuals for AB

AB-UTI/AB and DDD-UTI/DDD). The qqplots of the residuals also indicate that the assumptions of the residuals are normal distributions are appropriate except for the DDD-UTI/DDD.

**Table 4.19. Weighted Analysis for Regression**

| | AB | | AB-UTI | | DDD | | DDD-UTI | | Culture | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Int | Mon | Int | Mon | Int | Mon | Int | Mon | Int | Mon |
| Estimate | -0.16 | -0.03 | 0.33 | -0.01 | -1.09 | 0.46 | 3.38 | -0.06 | 0.44 | 0.01 |
| Std. Error | 0.31 | 0.04 | 0.15 | 0.02 | 2.61 | 0.35 | 1.58 | 0.21 | 0.20 | 0.03 |
| t value | -0.51 | -0.70 | 2.14 | -0.41 | -0.42 | 1.30 | 2.15 | -0.28 | 2.22 | 0.20 |
| $Pr(>\|t\|)$ | 0.62 | 0.50 | 0.06 | 0.69 | 0.68 | 0.22 | 0.06 | 0.79 | 0.05 | 0.85 |

**Table 4.20. Weighted Regression for Proportions**

| | AB-UTI/AB | | DDD-UTI/DDD | |
|---|---|---|---|---|
| Variable | Int | Mon | Int | Mon |
| Estimate | 0.19 | -0.01 | 0.17 | 0.00 |
| Std. Error | 0.06 | 0.01 | 0.04 | 0.01 |
| t value | 3.02 | -0.96 | 4.41 | -0.77 |
| $Pr(>\|t\|)$ | 0.01 | 0.36 | 0.00 | 0.46 |

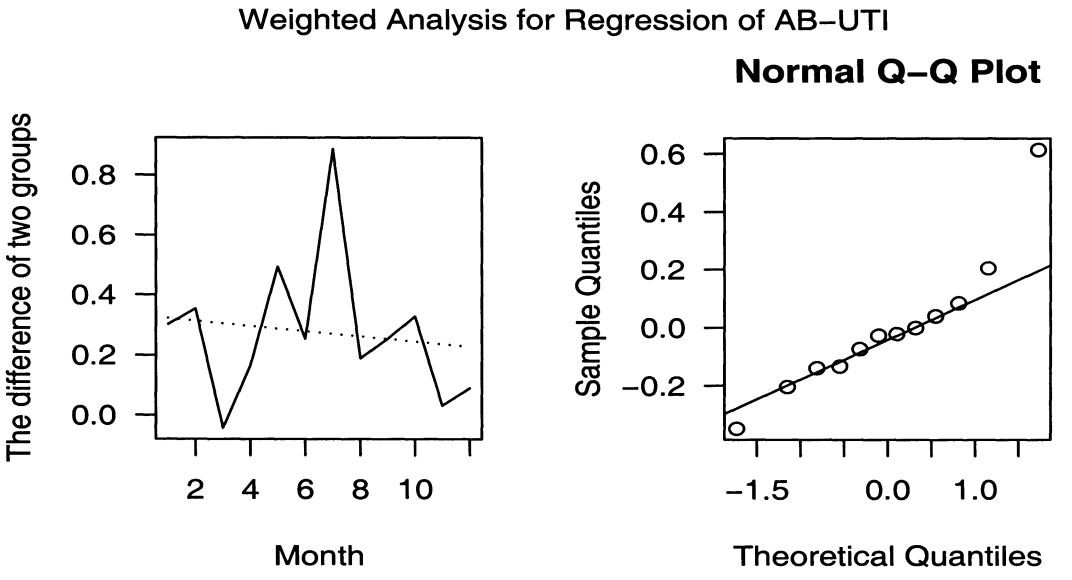


Figure 4.11: Weighted regression plot and qqplot of residuals for AB-UTI

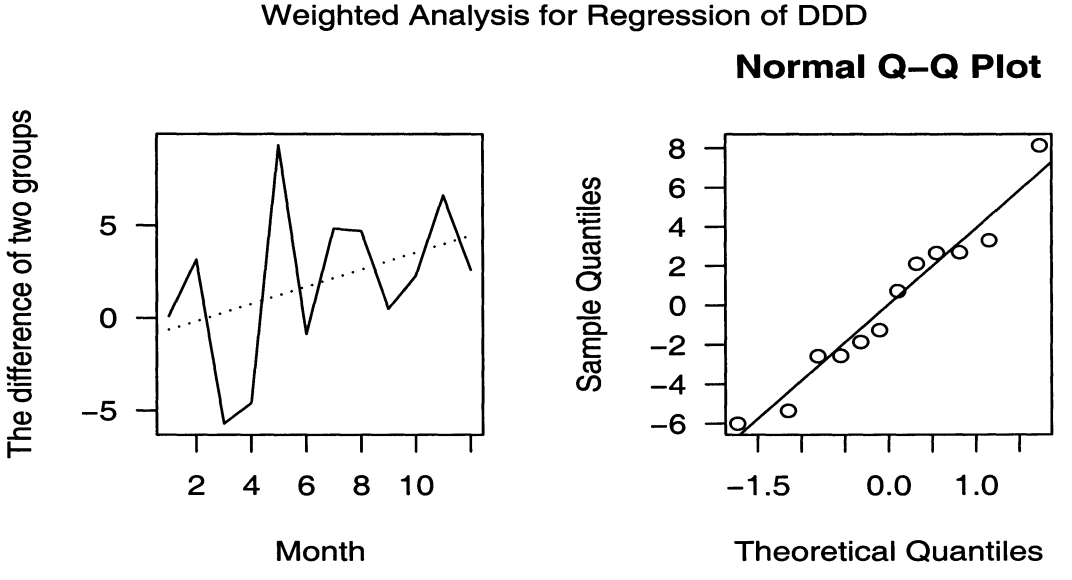


Figure 4.12: Weighted regression plot and qqplot of residuals for DDD

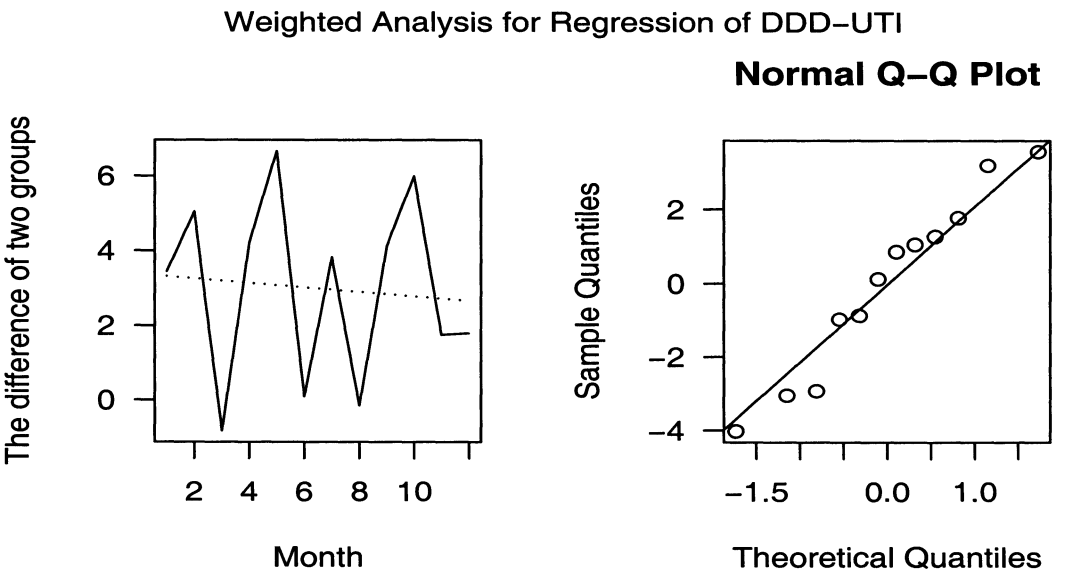


Figure 4.13: Weighted regression plot and qqplot of residuals for DDD-UTI

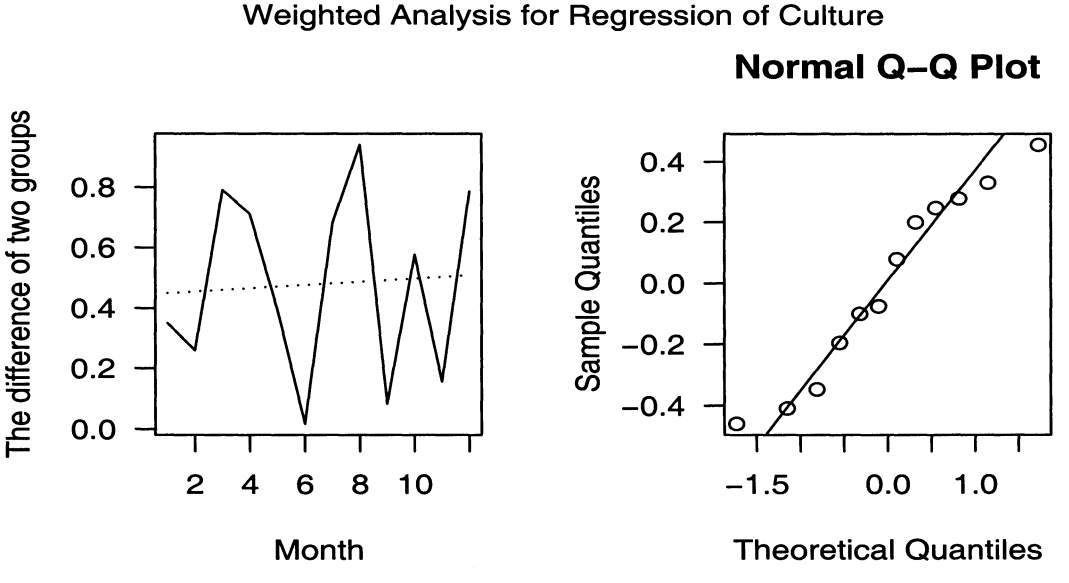


Figure 4.14: Weighted regression plot and qqplot of residuals for Culture

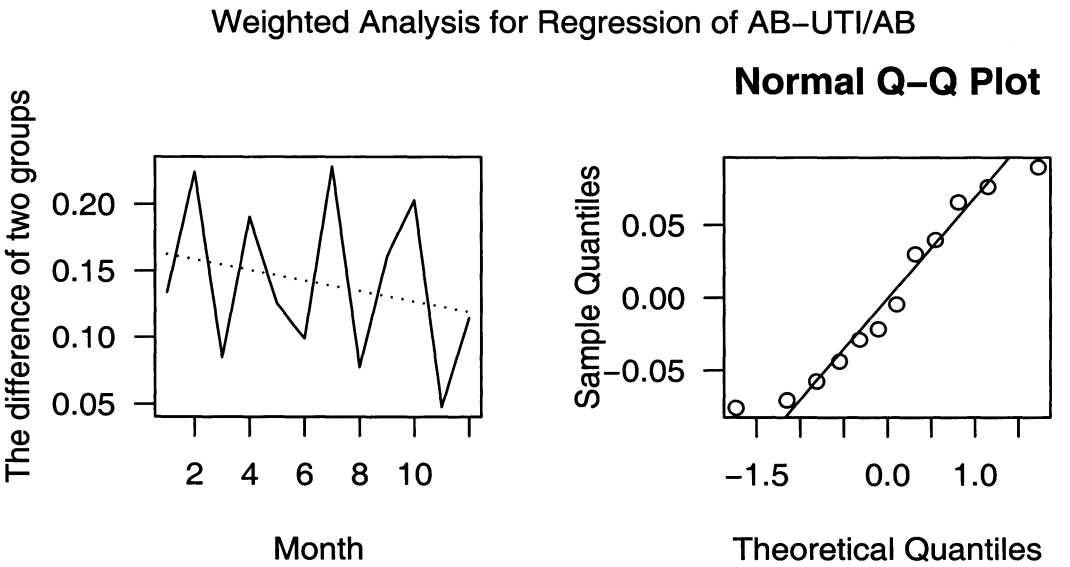


Figure 4.15: Weighted regression plot and qqplot of residuals for AB-UTI/AB

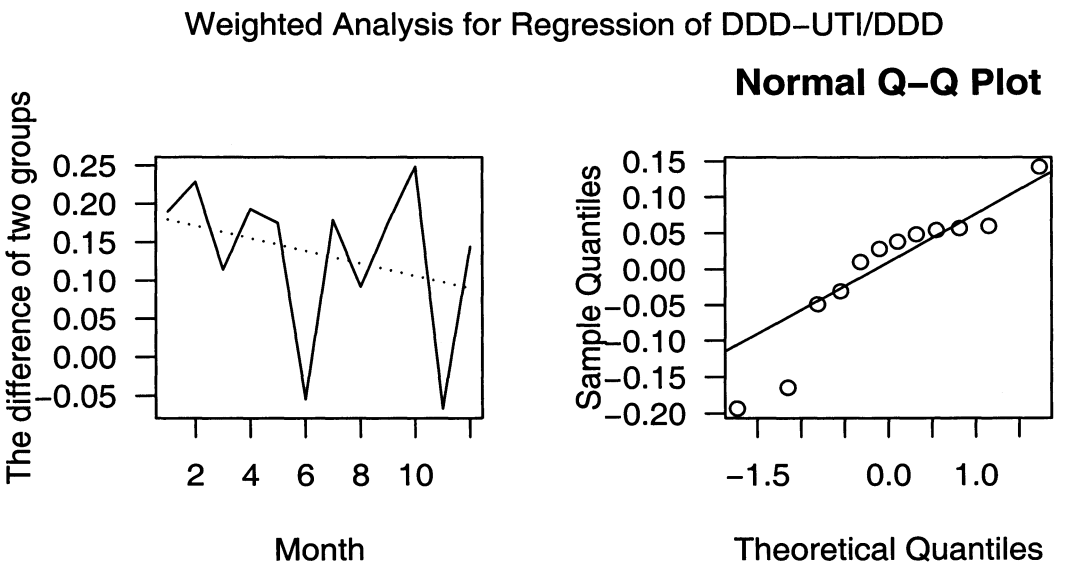


Figure 4.16: Weighted regression plot and qqplot of residuals for DDD-UTI/DDD

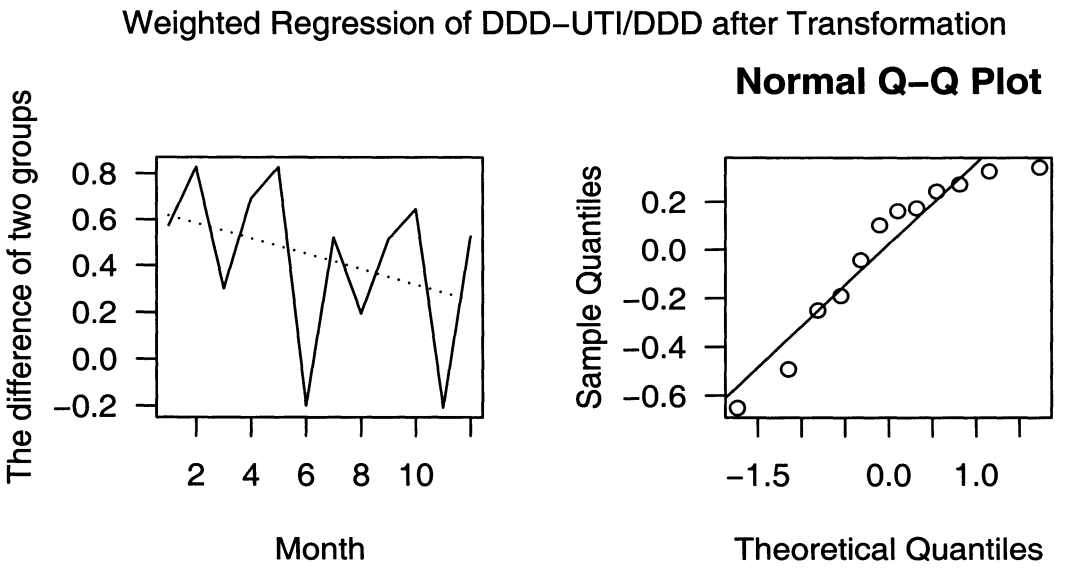


Figure 4.17: Weighted regression plot and qqplot of residuals for DDD-UTI/DDD after log transformation

Since the $p_{ijk}$ for the proportions (DDD-UTI/DDD and AB-UTI/DDD) take values in $[0, 1]$, $d_{jk} = \log(p_{1jk}) - \log(p_{2jk})$ is in $(-\infty, +\infty)$. Perhaps a transformation prior to the regressions would improve the precision.

A similar conclusion is reached after log transformation for DDD-UTI/DDD. The qqplot for residuals (Figure 4.17) shows that the precision of the regression for DDD-UTI/DDD has been improved by the transformation. However, the transformation does not change too much for the qqplot of residuals for AB-UTI/AB.

**Table 4.21. Weighted Regression for DDD-UTI/DDD after Trans**

| | Estimate | Std. Error | t value | Pr(> \|t\|) |
|---|---|---|---|---|
| Int | 0.65 | 0.21 | 3.06 | 0.01 |
| Mon | -0.03 | 0.03 | -1.15 | 0.28 |

# Chapter 5

# Conclusions

## 5.1 Main Results

The two-side paired t-tests for the rates of antibiotic courses and defined daily dosage indicate that the algorithms do not reduce the antibiotic use in nearly half of the 12 months. Even for the months in which fewer antibiotics were used for the intervention group, the differences are not statistically significant (p-values are larger than 0.4). Thus, the analyse demonstrate that the algorithms cannot reduce the overall antibiotic use significantly in LTCFs.

The two-side paired t-tests for the rates of urine cultures and antibiotic use for urinary infections (including antibiotic courses and defined daily dosage) show that the rates are reduced in most months. However, the reductions are not significant at the level of $\alpha = 0.05$. If we choose the level of $\alpha = 0.1$ as the criterion, the differences are significant for urine cultures in 4th and 10th months, are significant for antibiotic

courses for urinary infections in 1st month and are significant for antibiotic defined daily dosage for urinary infections in 1st, 4th, 5th and 10th months.

The algorithms did not significantly reduce the antibiotic use for urinary infections and overall antibiotic use. How about the proportions of antibiotic use for urinary infections in overall antibiotic use? The two-side paired t-tests for the proportions (including the antibiotic courses and defined daily dosage) illustrate that the algorithms reduce almost all of the proportions except for the proportions of antibiotic daily dosage in 6th and 11th months. The reductions are significant at the level of $\alpha = 0.1$ for these two variables in 1st, 2nd, 4th and 7th months and are significant for daily dosage in 10th month. They are also significant at the level of $\alpha = 0.05$ for both variables in 1st, 2nd and 4th months.

The above analysis is based on the assumption that the difference for each variable comes from a normal distribution with common variance. Although empirical research suggests that a paired t-test is fairly robust to departures from the assumption, further analysis should also be conducted. If the assumption of the normal distribution is not satisfied, Wilcoxon signed rank tests can be used for the analysis. By using Wilcoxon signed rank tests, I get similar conclusions for these variables. In this study, the assumption of common variance obviously is violated since the sample sizes of clusters vary from 101 to 378. Weighted cluster-level paired t-tests were used to improve the analysis by giving weights that are related to the sample sizes or variances. Some gain in power of the tests may be achieved using log transformation of the proportions before weighting the paired t-tests. By the weighted analysis (after

log transformation if needed) for all of the variables, estimates of the effect of intervention and the corresponding p-values have not been changed very much. However, the corresponding confidence intervals are much shorter than before. This indicates that the powers of the tests are improved by the weighted tests.

The plots of the differences between the two groups for these variables vs time (study month) indicate that the differences fluctuate with time and there are no significant increasing or decreasing trends. Fitting them by straight-line models shows: antibiotic courses (including overall courses and courses for urinary infections), defined daily dosage for urinary infections and the two proportions (including the courses and daily dosage) have negative slopes. However, the slopes are not significant at the level of $\alpha = 0.1$ except for antibiotic courses for urinary infections. Culture and defined daily dosage have insignificant positive slopes. The intercepts of antibiotic use for urinary infections and the proportions (including courses and dosage) are positive and are significant at the level of $\alpha = 0.05$. This means that the algorithms significantly reduce the antibiotic use for urinary infections and the proportions in the 12-month study.

Similar conclusions are reached by weighting the linear regressions (after transformation if needed) except that the deceasing trend of antibiotic courses for urinary infections changes to be not significant and the intercept of urine culture changes to be significant. Therefore, the weighted regressions indicate that the algorithms also significantly reduce the number of cultures in the 12-month study in addition to antibiotic use for urinary infections and proportions. Furthermore, the loss in power of

the regression analysis is compensated since the assumption of homogeneity of variances is satisfied. And qqplots of the residuals have been improved by the weighted analysis.

## 5.2 Strengths and Possible Limitations

Cluster-level analyses were conducted in this study since the primary questions of interest focused on the randomized unit as a whole rather than on the individual resident. Furthermore, cluster-level analyses remove the problem that the individual data lack statistical independence among observations within a cluster; And the analyses are easier to conduct and explain than individual-level analyses. However, individual-level analyses provide more efficient estimates of the effect of intervention than unweighted analyses when there are many clusters per group, particularly when cluster sizes are highly variable (Donner and Klar (2002) p80). Therefore, weighted cluster-level analyses are conducted to compensate for the lack of power due to the variation of cluster sizes in this study.

Paired t-tests (weighted or unweighted) are conducted to analyze the differences of the intervention group and control group since the nursing homes are pairwise matched and a paired t-test is fairly robust to departures from the normality assumption and the homogeneity of variance assumption. Furthermore, evidence suggests (Brookmeyer and Chen, 1998) that paired t-tests perform well with data arising from community intervention trials even with as few as five pairs, when compared with

Wilcoxon signed rank test, standard Mantel-Haenszel chi-square test and Liang's test, which need more numbers of pairs in the comparison of the proportions of the two groups.

My primary aim in this study was to evaluate whether the clinical algorithms would lead to a reduction in overall antibiotic use, antibiotic use for urinary infections etc. According to the hypothesis, one-side t-tests should be considered. However, I use two-side t-tests in this study. Therefore, the corresponding conclusions are more conservative if evaluating the effect of the intervention by p-values.

In regression analyses (unweighted, weighted or weighted after transformation), the assumption that the residuals are normally and independently distributed with mean zero and constant variance should be checked for assuring the exact tests of the hypothesis. However, due to a small sample size (12) the appearance of a moderate departure from normality can be acceptable when considering fluctuations. Moreover, the analysis of variance and related inferences are robust to the normality assumption (Montgomery, 1997).

Missing data occurred in some clusters. Various imputation techniques can be considered, but none of which correctly accounts for the effect of imputation on the resulting estimates of variance. Therefore, I used the simplest method in this study, removing the subjects with missing data. That leads to loss the information not only from the clusters of the missing data but also from the corresponding matched clusters. Perhaps choosing a more appropriate imputation method would be beneficial in a future research study.

# Bibliography

Abrutyn, E., Mossey, J., Levinson, M., Boscia, J., Pitsakis, P., and Kaye, D. (1991). Epidemiology of asymptomatic bacteriuria in elderly women. *J Am Geriatr Soc*, 39:388–393.

Armitage, P., Berry, G., and Matthews, J. N. S. (2002). *Statistical methods in medical research.* Blackwell Science Ltd, Oxford, Boston, fourth edition.

Bland, J. M. and Kerry, S. M. (1997). Statistics notes: Trials randomised in clusters. *British Medical Journal*, 315:600.

Brookmeyer, R. and Chen, Y.-Q. (1998). Person-time analysis of paired community intervention trials when the number of communities is small. *Stat Med*, 17:2121–2132.

Donner, A. and Donald, A. (1987). Analysis of data arising from a stratified design with cluster as unit of randomization. *Stat in Med*, 6:43–52.

Donner, A. and Klar, N. (2002). *Design and analysis of cluster randomization trials in health research.* Arnold, London.

Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis.* Wiley, New York, NY, second edition.

Gail, M., Mark, S., Carroll, R., Green, S., and Pee, D. (1996). On design considerations and randomization-based inference for community intervention trials. *Stat in Med*, 15:1069–1092.

Korn, E. (1984). The paired t-test. *Applied Statistics*, 33:230–231.

Liang, K., Beaty, T., and Cohen, B. (1986). Application of odds ratio regression models for assessing familial aggregation from case-control studies. *Americal Journal of Epidemiology*, 124:678–683.

Liang, K. (1985). Odds ratio inference with dependent data. *Biometrika*, 72:678–682.

Loeb, M., Brazil, K., Lohfeld, L., and et al (2002). Optimizing antibiotics in residents of nursing homes: protocol of a randomized trial. *BMC Health Services Research*, 2(17). http://www.biomedcentral.com/1472-6963/2/17.

Loeb, M., Simor, A., Landry, L., Walter, S., McArthur, M., Duffy, J., Kwan, D., and McGeer, A. (2001). Antibiotic use in facilities which provide chronic care. *Journal of General Internal Medicine*, 16:376–383.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from restrospective studies of disease. *Journal of the National Cancer Institute*, 22:719–748.

Marubini, E., Correa, M., and Milani, S. (1988). Analysis of dichotomous response variables in teratology. *Biometrical Journal*, 30:965–974.

Montgomery, D. (1997). *Design and analysis of experiments*. John Wiley Sons, New York, NY, fourth edition.

Nicolle, L., Bentley, D., Garibaldi, R., Neuhaus, E., and Smith, P. (1996). Antimicrobial use in long-term-care facilities. *Infect Control Hosp Epidemiol*, 17:119–128.

Nicolle, L., Bjornson, J., Harding, G., and MacDonell, J. (1983). Bacteriuria in elderly institutionalized men. *N Engl J Med*, 309:1420–1425.

Nicolle, L., Mayhew, J., and Bryan, L. (1987). Prospective randomized comparison of therapy and no therapy for asymptomatic bacteriuria in institutionalized women. *Am J Med*, 83:27–33.

Ouslander, J., Schapira, M., Schnelle, J., and et al. (1995). Does eradicating bacteriuria affect the severity of chronic urinary incontinence in nursing home residents. *Ann Inter Med*, 122:749–754.

Strausbaugh, L., Crossley, K., Nurse, B., and Thrupp, L. (1996). Antimicrobial resistance in long-term-care facilities. *Infect Control Hosp Epidemiol*, 17:129–140.

Warren, J., Palumbo, F., Fitterman, L., and Speedie, S. (1991). Incidence and characteristics of antibiotic use in aged nursing home patients. *J Am Geriatr Soc*, 39:963–972.

Woolson, R. F. and Clarke, W. R. (2002). *Statistical methods for the analysis of biomedical data.* A John Wiley Sons, Inc., New York, NY, second edition.

Zimmer, J., Bentley, D., Valenti, W., and Wilson, N. (1986). Systemic antibiotic use in nursing homes. *J Am Geriatr Soc*, 34:703–710.

# Appendix A

# Related S-plus Code and qqplots for t-tests

```
 # The function of plotting the data
  utifun<-function(data,residay=residay,n=0.5, m=0.2) {
  rate<-as.matrix(data)/as.matrix(residay)
  control<-apply(rate[,1:10],1,mean, na.rm=T)
  experiment<-apply(rate[,11:20],1,mean, na.rm=T)
  plot(1:12,control,type="l",xlab="Month",ylab="",
    ylim=c(min(control,experiment)-n,max(control,experiment)+n))
  lines(experiment,lty=2)
  }


# paired t-test function for all variables
  pairt<-function(data,residay){
  rate<-as.matrix(data)/as.matrix(residay)
  t.tes<-matrix(NA,nrow=12,ncol=4)
  w.tes<- matrix(NA,nrow=12,ncol=1)
  for( i in 1:12){
    a<-rate[i,1:10]
    b<-rate[i,11:20]
    t<- t.test(a,b,paired=T)
    w<-wilcox.test(a,b,paired=T)
    t.tes[i,1:4]<-cbind(t$estimate,t$conf.int[1],t$conf.int[2],
      t$p.value)
    w.tes[i,1]<-round(w$p.value,2)
    }
  out<-data.frame(t.tes,w.tes)
  names(out)<-c("est","t.c", "t.h","t.p","w.p")
  out
```

```
  }


# weighted pair t-test function
  tw.test<-function(x,y,weight){
  n<-length(x)
  if((n!= length(y))||(n!=length(weight)))
  stop("'x' 'y' and 'weight' must have the same length when paired=TRUE.")
  d <- x - y
  if((bad.obs <- sum(!(both.ok <- is.finite(d*weight)))) > 0) {
     if(!all(is.finite(x)))
     is.not.finite.warning(x)
     if(!all(is.finite(y)))
     is.not.finite.warning(y)
     if(!all(is.finite(weight)))
     is.not.finite.warning(weight)
     d <- d[both.ok]
     n <- length(d)
     weight<-weight[both.ok]
     }
  est<-sum(d*weight)/sum(weight)
  s<-sqrt(sum(weight*(d-est)^2)/sum(weight))
  t<-sqrt(n-1)*est/s
  pvalue<-2*(1-pt(abs(t),n-1))
  conf.l<-est-qt(0.975,n-1)*s/sqrt(n-1)
  conf.h<-est+qt(0.975,n-1)*s/sqrt(n-1)
  df<-n-1
  result<-cbind(est,conf.l,conf.h,t,pvalue,df)
  return(round(result,2))
  }

# weighted paired t-test for count outcome
  pair.w1<-function(data1,data2){
  rate<-as.matrix(data1)/as.matrix(data2)
  r1<-rate[,1:10]
  r2<-rate[,11:20]
  weight<-1/(r1/as.matrix(data2[,1:10])+r2/as.matrix(data2[,11:20]))
  tw.tes<- NULL par(mfrow=c(4,3))
  for (i in 1:12){
    tw<-tw.test(r1[i,],r2[i,],weight[i,])
    tw.tes<-rbind(tw.tes,tw)
    qqnorm(weight[i,]*(r1[i,]-r2[i,])/sum(weight[i,],na.rm=T))
    qqline(weight[i,]*(r1[i,]-r2[i,])/sum(weight[i,],na.rm=T))
```

```
    }
  data.frame(tw.tes)
  }

# weighted paired t-test for proportion and continuous outcome
  pair.w2<-function(data1,data2, wt=bed[,2]){
  p<-as.matrix(data1)/as.matrix(data2)
  p01<-p[,1:10]
  p02<-p[,11:20]
  weight<-wt[1:10]*wt[11:20]/(wt[1:10]+wt[11:20])
  tw.tes<-NULL
  par(mfrow=c(4,3))
  for (i in 1:12)
   {t<-tw.test(p01[i,],p02[i,],weight)
   tw.tes<-rbind(tw.tes,t)
   qqnorm(weight*(p01[i,]-p02[i,])/sum(weight,na.rm=T))
   qqline(weight*(p01[i,]-p02[i,])/sum(weight,na.rm=T))
   }
  data.frame(tw.tes)
  }

# Regression function
  reg<- function(data1,data2){
  ratio<-as.matrix(data1)/as.matrix(data2)
  d<-ratio[,1:10]-ratio[,11:20]
  dbar<-apply(d,1,mean, na.rm=T)
  x<-data.frame(cbind(c(1:12),dbar))
  names(x)<-c("Months","Diff.group")
  y<-lm(Diff.group~Months,data=x)
  par(mfrow=c(1,2))
  plot(x[,1],x[,2],xlab="Month",ylab="The difference of two groups",type="l")
  lines(x[,1],y$fit,lty=3)
  qqnorm(resid(y))
  qqline(resid(y))
  list(summary(y,cor=F),anova(y))
  }

#weighted mean function
  w.mean<-function (x, w, na.rm = FALSE) {
    if (missing(w))
        w <- rep(1, length(x))
    if (is.integer(w))
        w <- as.numeric(w)
```

```
    if (na.rm) {
        w <- w[i <- !is.na(x*w)]
        x <- x[i]
    }
    sum(x * w)/sum(w)
  }

# Weighted regression for count outcome
  reg.w1<-function(data1,data2){
  ratio<-as.matrix(data1)/as.matrix(data2)
  d<-ratio[,1:10]-ratio[,11:20]
  weight<-1/(ratio[,1:10]/as.matrix(data2[,1:10])
    +ratio[,11:20]/as.matrix(data2[,11:20]))
  dbar<-NULL
  for (i in 1:12){
    dd<-w.mean(d[i,],weight[i,],na.rm=T)
    dbar<-c(dbar,dd)
    }
  x<-data.frame(cbind(c(1:12),dbar))
  names(x)<-c("Months","Diff.group")
  y<-lm(Diff.group~Months,data=x)
  par(mfrow=c(1,2))
  plot(x[,1],x[,2],xlab="Month",ylab="The difference of two groups",type="l")
  lines(x[,1],y$fit,lty=3)
  qqnorm(resid(y))
  qqline(resid(y))
  list(summary(y,cor=F),anova(y))
  }

# Weighted regression for continuous and proportion outcome
  reg.w2<- function(data1,data2,wt=bed[,2]){
  ratio<-as.matrix(data1)/as.matrix(data2)
  weight<-wt[1:10]*wt[11:20]/(wt[1:10]+wt[11:20])
  d<-ratio[,1:10]-ratio[,11:20]
  dbar<-NULL
  for (i in 1:12){
     dd<-w.mean(d[i,],weight,na.rm=T)
     dbar<-c(dbar,dd)
     }
  x<-data.frame(cbind(c(1:12),dbar))
  names(x)<-c("Months","Diff.group")
  y<-lm(Diff.group~Months,data=x)
  par(mfrow=c(1,2))
```

```
plot(x[,1],x[,2],xlab="Month",ylab="The difference of two groups",type="l")
lines(x[,1],y$fit,lty=3)
qqnorm(resid(y))
qqline(resid(y))
list(summary(y,cor=F),anova(y))
}
```

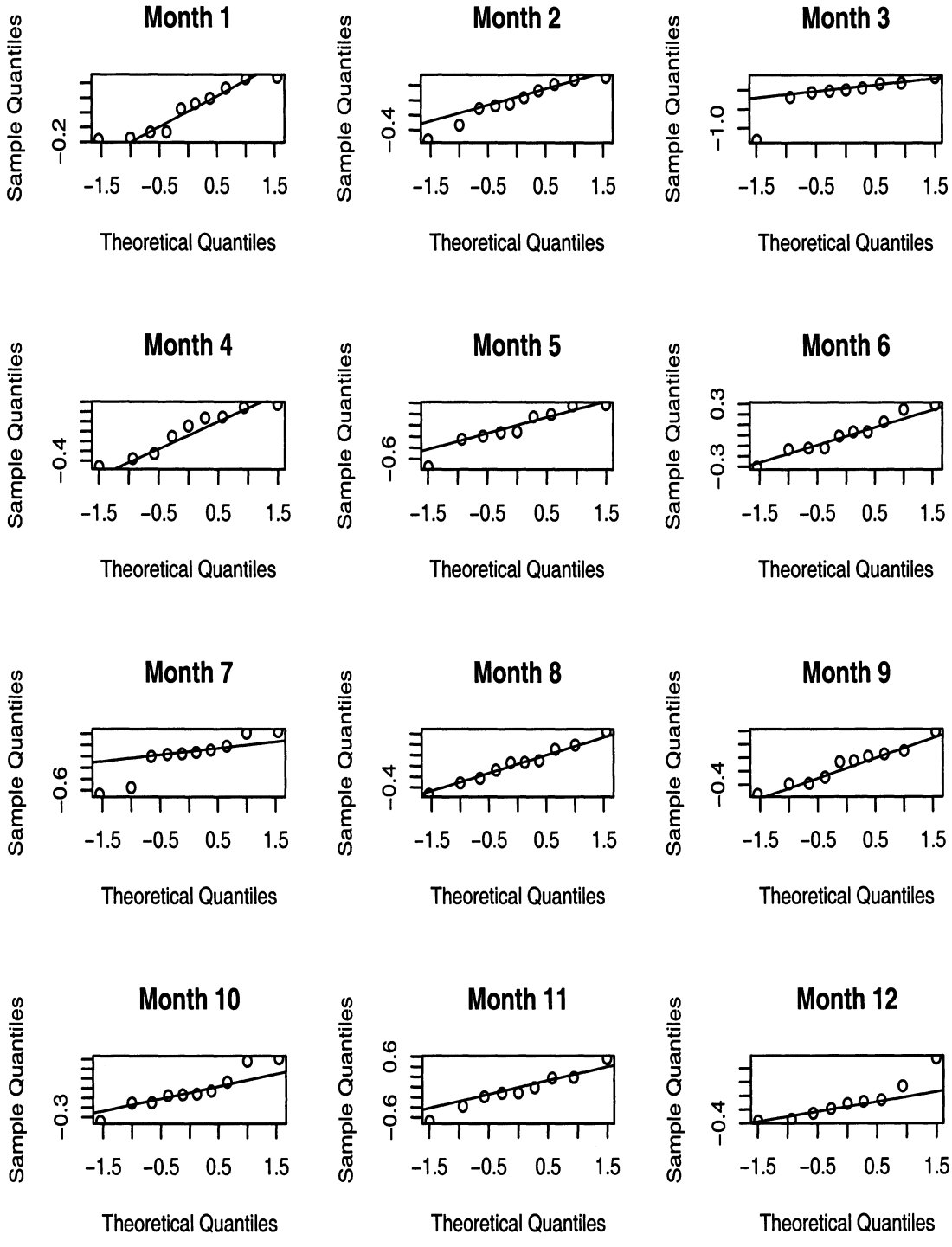


Figure A.1: qqplot of weighted AB for each month

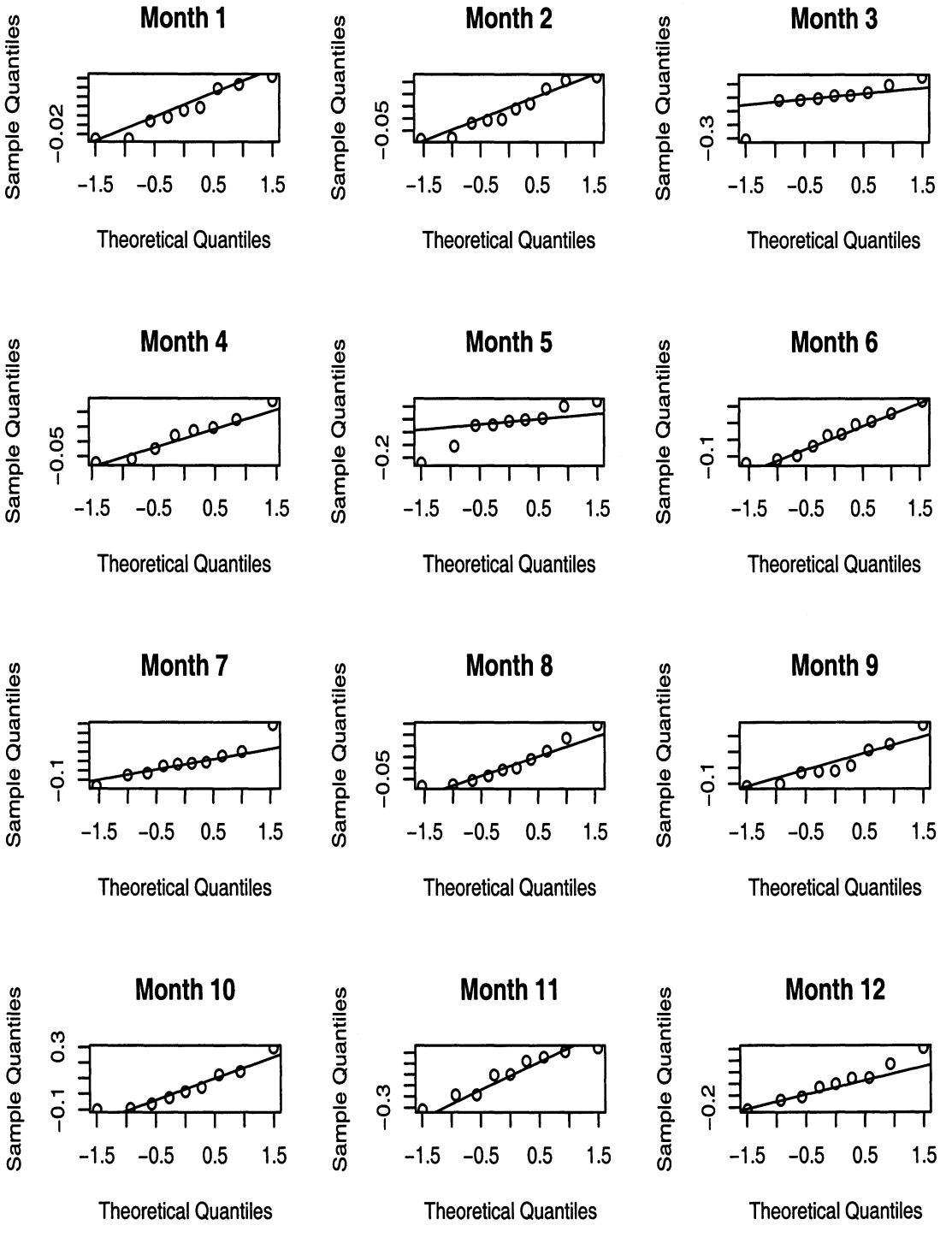


Figure A.2: qqplot of weighted AB-UTI for each month

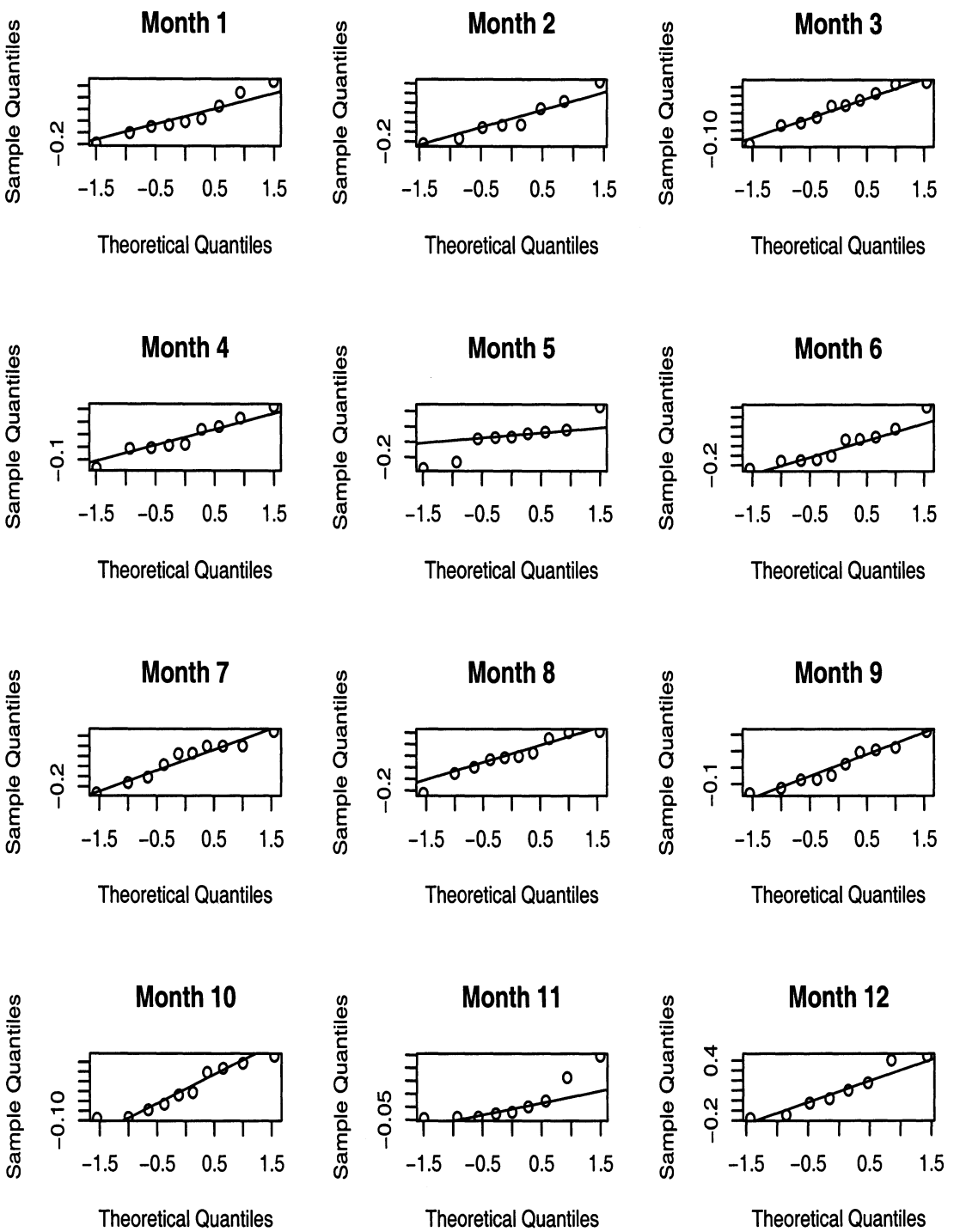


Figure A.3: qqplot of weighted Culture for each month

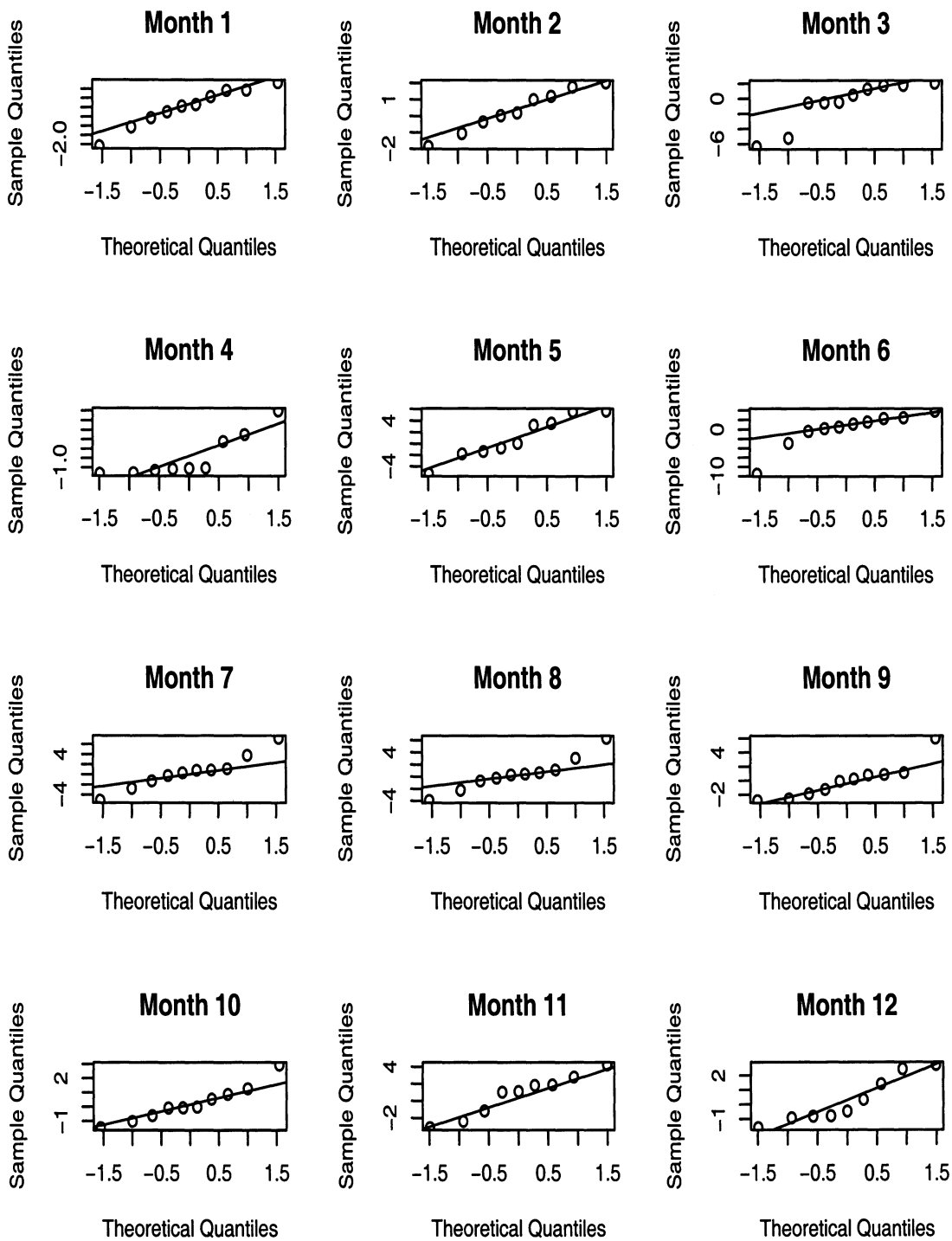


Figure A.4: qqplot of weighted DDD for each month

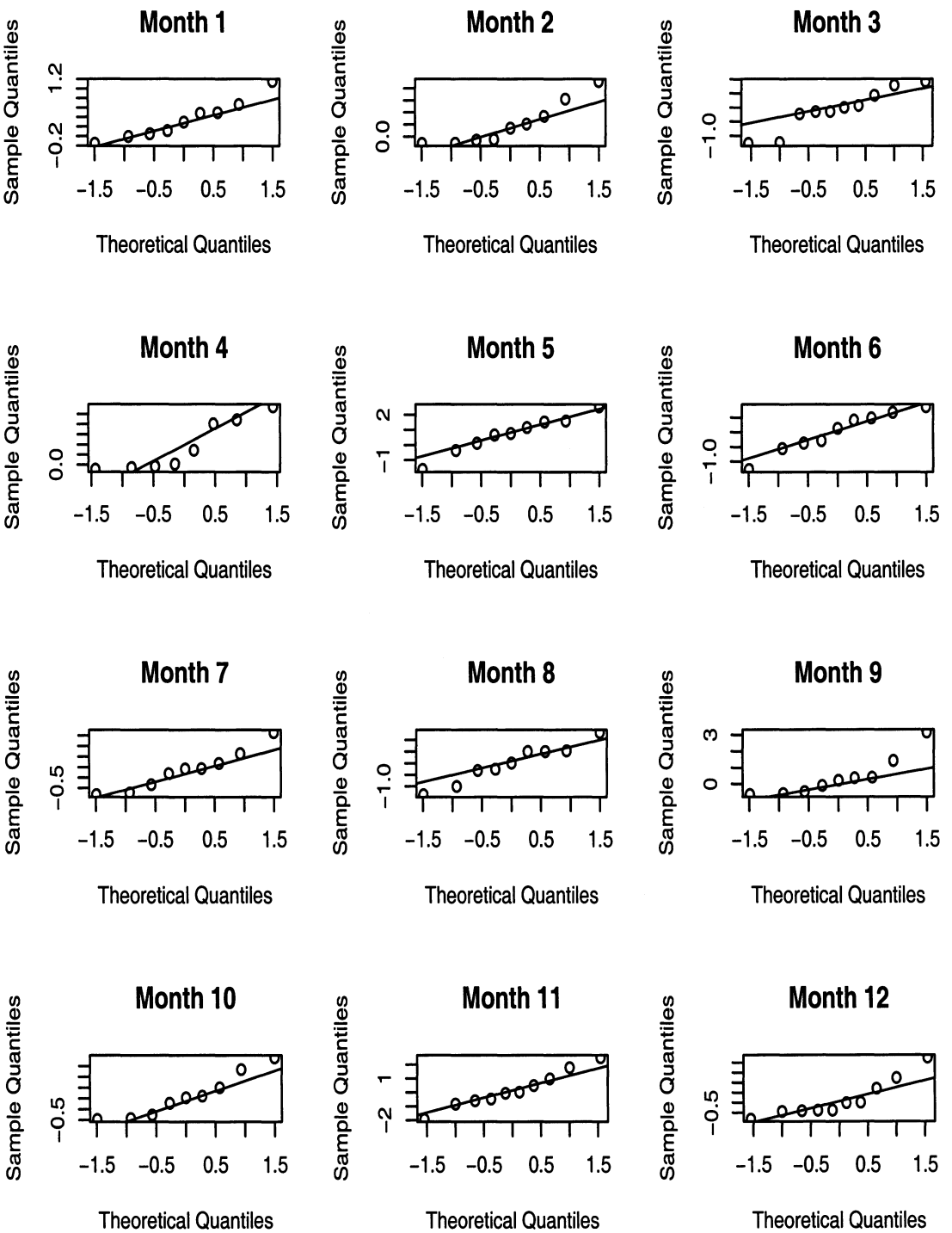


Figure A.5: qqplot of weighted DDD-UTI for each month

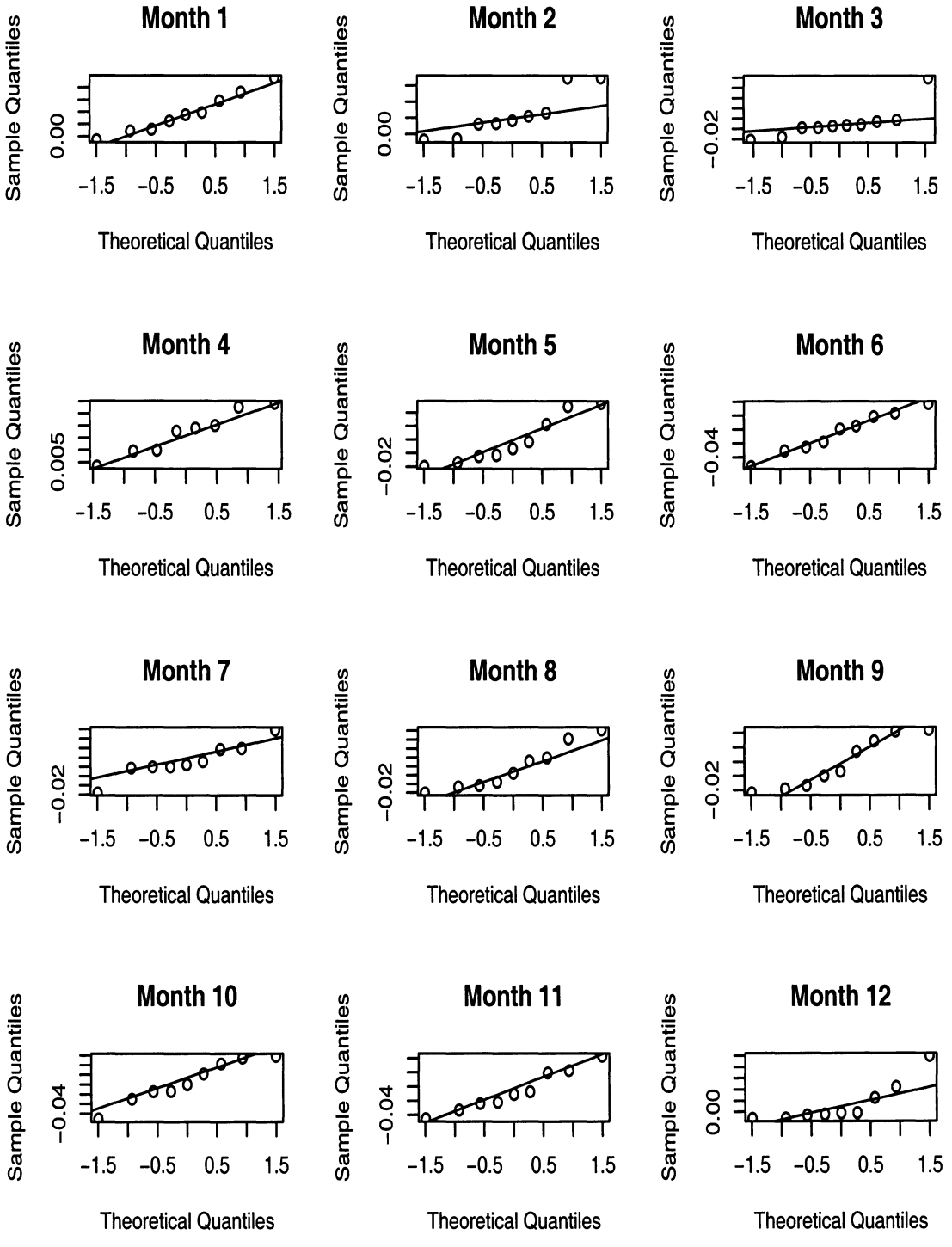


Figure A.6: qqplot of weighted DDD-UTI/DDD for each month

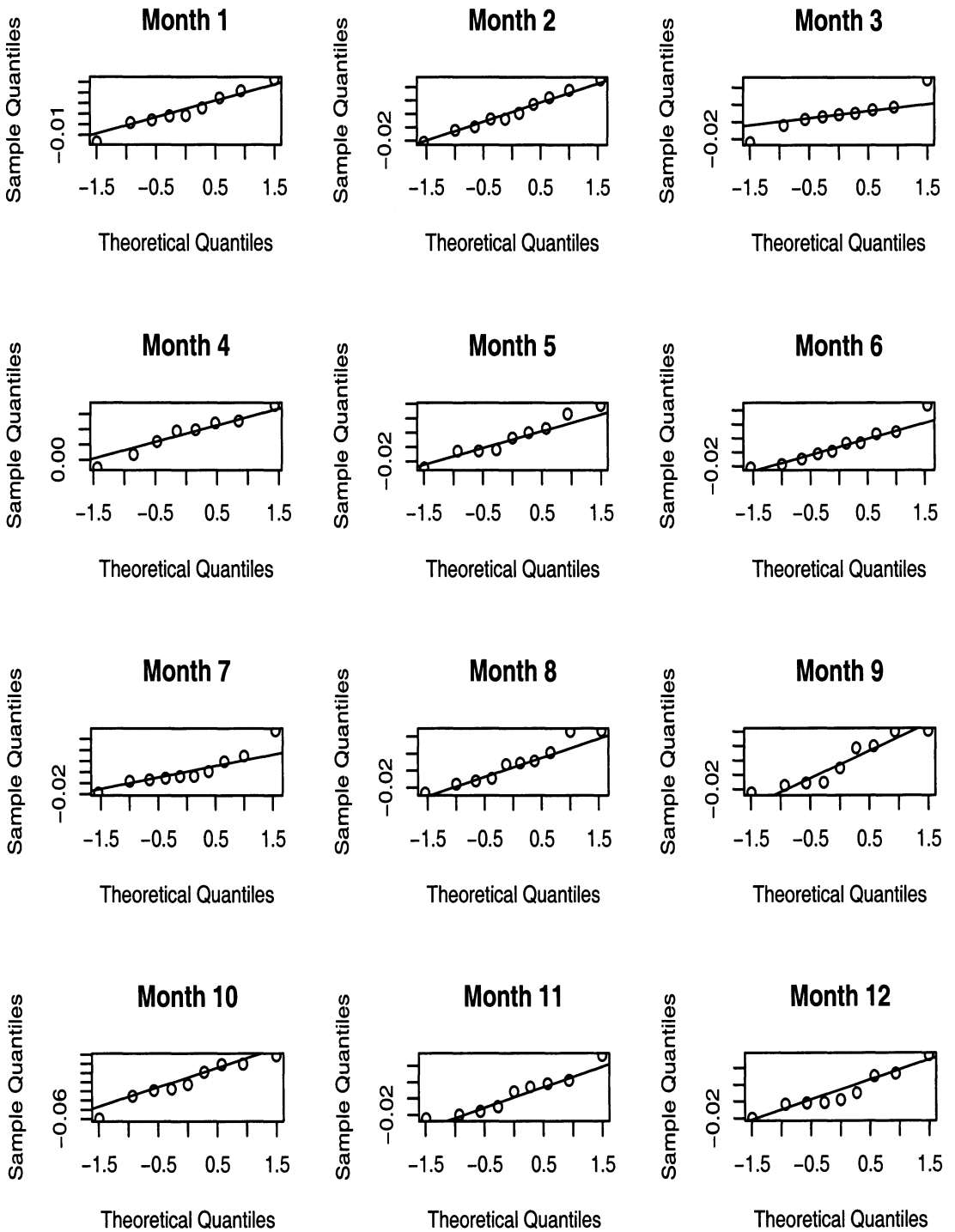


Figure A.7: qqplot of weighted AB-UTI/AB for each month

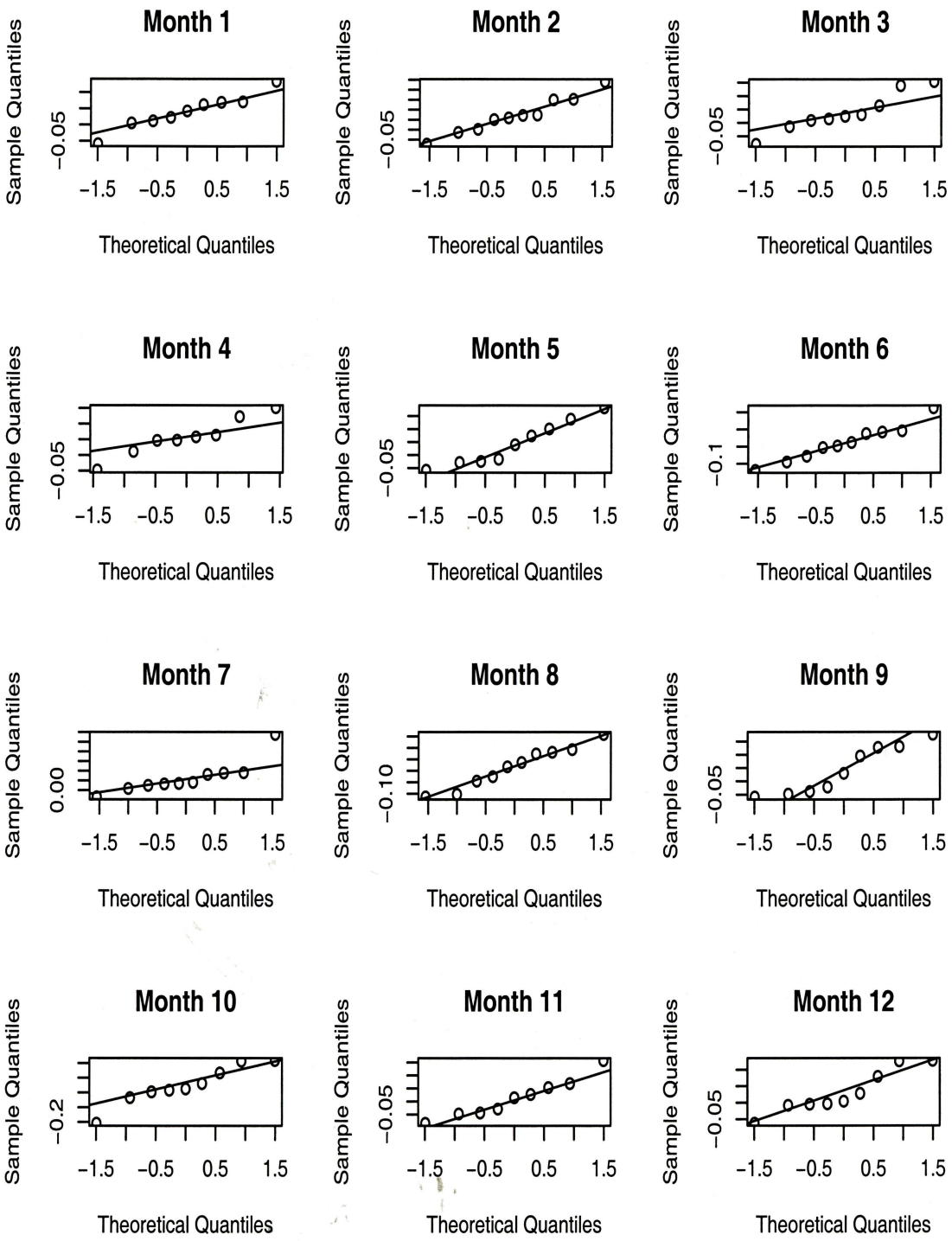


Figure A.8: qqplot of weighted AB-UTI/AB for each month after transformation

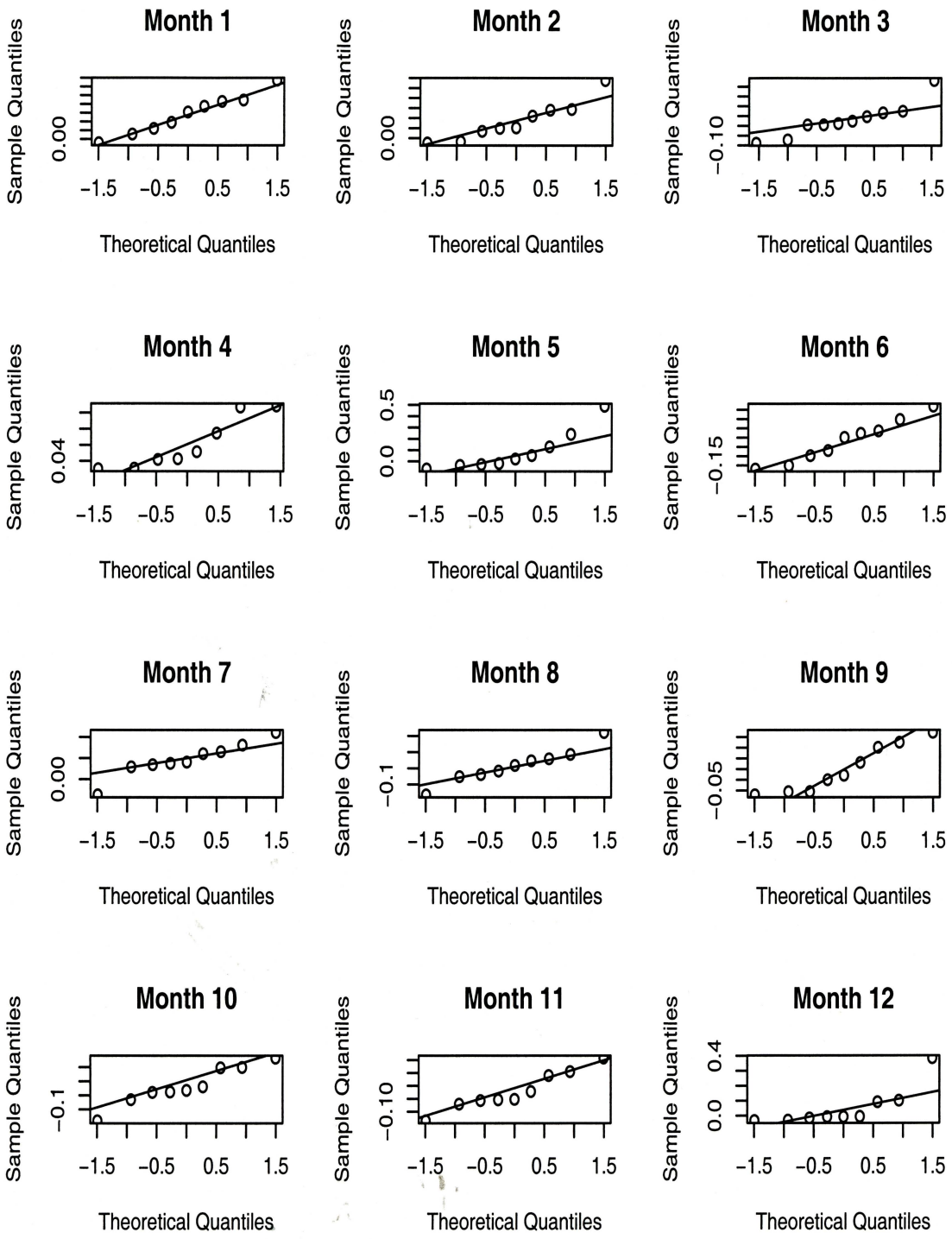


Figure A.9: qqplot of weighted DDD-UTI/DDD for each month after transformation