A new linear model-based approach for inferences about the mean area under the curve

Gregory E. Wilding[1*] , Rameela Chandrasekhar[1] and Alan D. Hutson[1]

[1]*Department of Biostatistics, University at Buffalo, 3435 Main Street, Buffalo NY 14214.*

June 27, 2011

Abstract

Outcome versus time data is commonly encountered in biomedical and clinical research. A common strategy adopted in analyzing such longitudinal data is to condense the repeated measurements on each individual into a single summary statistic such as the area under the curve (AUC). Standard parametric or non-parametric methods are then applied to perform inferences on the conditional AUC distribution. Disadvantages of this approach include the disregard of the within-subject variation in the longitudinal profile. We propose a general linear model approach, accounting for the within-subject variance, for estimation and hypothesis tests about the mean areas. Inferential properties of our approach are compared to those from standard methods of analysis using Monte Carlo simulation studies. The impact of missing data, within-subject heterogeneity and homogeneity of variance are also evaluated. A real working example is used to illustrate our approach. It is seen that the proposed approach is associated with a significant power advantage over traditional methods, especially when missing data is encountered.

Keywords: Area under the curve, trapezoidal rule, longitudinal data.

---

*Corresponding author: Department of Biostatistics, University at Buffalo, 3435 Main Street, Buffalo NY 14214. E-mail: gwilding@buffalo.edu

1 Introduction

When an outcome is measured repeatedly over time on the same individual the data is said to be longitudinal in nature. Longitudinal data differs from cross-sectional data in that within-subject data points are in general not independent. Hence statistical methods for analysis of such data, which assume such independence, are not valid. As remarked by Fitzmaurice et al. [1], a common strategy adopted in analyzing such longitudinal data is to condense the repeated measurements on any individual to a single summary statistic, which encapsulates an important feature. Scientific researchers find the use of summary statistics appealing since it offers an integrated approach in representing the subject's overall response and is also easy to interpret. Examples of widely used summary statistics are the area under the curve (AUC), mean or the median response for each individual, the maximum of each individual observation and the time corresponding to the maximum observation.

Analysis of longitudinal data using a summary statistic is also known as response-feature analysis or two-staged analysis. Such an approach transforms the problem from a longitudinal one to one which is cross-sectional in nature, essentially eliminating the within-subject correlation component. When the comparison of groups is the objective, these independent observations can be analyzed using standard methodologies, e.g. the two-sample t-test or the non-parametric wilcoxon rank-sum test. When the sample size is not sufficient to model correlated observations, summary measures analysis is appealing from a statistical perspective [1] although concerns exist regarding the loss of information when ignoring the correlation structure.

The summary statistics approach is found to be sufficiently flexible to accomodate missing values [2]. To deal with missing observations, Everitt and Pickles [2] suggest using the available data to compute the summary statistic, particularly when the proportion of missing observations is small. It should be noted that when using the available data to compute the summary statistic, both the proportion and type of missingness will potentially bias estimates and impact the efficiency of inference procedures.

The relationship between outcome and time can be illustrated using an outcome versus time graph. The area beneath the outcome-time curve is commonly known as the AUC. The AUC is applicable in many longitudinal settings and is particularly useful when the expected trajectories are complicated [3]. Fayers and Machin [4] discuss how when using the AUC, assumptions are made about the linearity of responses, e.g. if a 100-point scale is being analyzed, then a shift in score from 0 to 10 is as important as a shift from 50 to 60. Statistical simplicity aside, investigators are interested in using the AUC due to its scientific/clinical significance. In pharmacokinetics, the AUC is used as a measure of drug exposure or drug clearance from the body, where as in quality of life studies, the area under the curve represents a measure of the overall health of the subject. In AIDS studies evaluating the effectiveness of antiretroviral treatments, the area under the viral load vs. time curve translates to a measure of overall viral burden and acts as a predictor of clinical progression. In nutritional studies, the area under the blood sugar response curve reflects the total rise in

blood glucose levels. A recent article by Clark et al. [5] discusses a randomized cross-over study conducted to evaluate the effects of breakfast meal composition on second meal metabolic responses in adults with type 2 diabetes mellitus. Patients were randomized to two sequence groups, one with a high glycemic load and the other with a low glycemic load. After a standard lunch 4hrs post breakfast, blood plasma concentrations were obtained in regular intervals and were assayed for values of glucose, insulin and free fatty acid. Using a summary statistic such as the AUC has a few drawbacks, some of which are discussed by Fitzmaurice [1]. Analysis using a summary statistic is possible only if the covariates are time-invariant. Another disadvantage is that individuals may have the same AUC, but may have completely different individual profiles.

Various methods for approximating the AUC have been referenced in the literature such as the linear trapezoidal rule, log-linear trapezoidal rule, Lagrange method, spline method, to name a few [6]. Methods for the computation of the AUC also differ in the aspect of the area included, e.g. the total AUC, positive incremental AUC, net incremental AUC and the partial AUC. Advantages and disadvantages of each approach have been examined extensively, refer to Venter et al. [7] and Allison et al. [8] for more details. The total AUC calculates the entire area below the outcome-time curve above zero and is widely used. In this paper, we focus on the total AUC using the most commonly used and easy to implement linear trapezoidal method. Towards this end let Y_{ij} be the observed response for subject i at the jth time point t_j, where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. Assuming all subjects to be observed at the same set of occasions, the total AUC for the ith individual can be computed using the trapezoidal rule as follows. Letting $\mathbf{Y}_i = \begin{pmatrix} Y_{i1} & Y_{i2} & Y_{i3} & \ldots & Y_{im} \end{pmatrix}'$, the AUC for the ith subject can be written as $AUC_i = \mathbf{c}'\mathbf{Y}_i$, where $\mathbf{c} = \begin{pmatrix} c_1 & c_2 & c_3 & \ldots & c_m \end{pmatrix}'$, and

$$
c_j = \begin{cases} \frac{t_{j+1}-t_j}{2}, & j = 1, \\ \frac{t_j-t_{j-1}}{2}, & j = m, \\ \frac{t_{j+1}-t_{j-1}}{2}, & otherwise. \end{cases} \tag{1}
$$

Replacing repeated measures on a single individual with a single metric leads to loss of information due to the fact that within-subject variability fails to be captured. This motivates us to consider a new linear model-based approach that incorporates the characteristics of the individual response profile while still using the mean AUC as the parameter of interest. Though the actual estimate of AUC has the same interpretation as before, we obtain a variance estimate using all the available information. Thus yielding more efficient inference procedures based on the same metric. Estimation of parameters will be approached via the restricted maximum likelihood (REML) method along with the Newton-Raphson iterative algorithm to optimize the likelihood function. Contrasts can be constructed using the estimates obtained at each time point to test null hypotheses regarding the mean AUC. This will be explained in more detail in the next section. The difficulty associated with fostering such an approach lies in choosing the true working

correlation structure. Correct specification of the functional form of the covariance matrix improves the efficiency of the estimates [9]. The AIC criterion has often been chosen to guide the selection of a covariance structure ([10],[11]) and we utilize it in our simulation studies.

The rest of the paper is structured as follows. In Section 2, we develop our proposed method for making inferences on AUC means. Inclusion of covariates is discussed in Section 3. Comparison of inferential properties between our approach and the traditional method is conducted using simulation studies in Section 4. In Section 5, we provide a working example to illustrate the applicability of our approach and compare it with the traditional method. To conclude, we summarize the paper and outline our future research in Section 6.

2   The Proposed Approach

In this section we discuss the relationship between the linear combination of time point specific sample means and the parameter of interest, the model-based method for use in inference on the parameter, and the robustness of the method in the context of missing values.

2.1   Re-expression of the mean AUC as a linear combination of means

Let Y_{ij} be defined as in the previous section with $\mathbf{Y} = \left(Y_{11}, Y_{12}, \ldots, Y_{1m}, \ldots, Y_{nm} \right)'$, and let $\overline{\mathbf{Y}} = \left(\overline{Y}_{\cdot 1}, \overline{Y}_{\cdot 2}, \ldots, \overline{Y}_{\cdot m} \right)'$, where $\overline{Y}_{\cdot j} = \frac{1}{n} \sum_{i=1}^{n} Y_{ij}$. Furthermore let $\mu_j = E(\overline{Y}_{\cdot j})$. The estimate of the mean AUC based on the trapezoidal rule can be written as a linear combination of the mean responses at each time point.

Theorem 1. *The estimate of the mean AUC ($\widehat{\mu}_{AUC}$) obtained using the trapezoidal rule can be re-expressed as a linear combination of the mean outcome values at each sampling time.*

Proof. The mean AUC can be estimated using

$$
\begin{aligned}
\widehat{\mu}_{AUC} &= \frac{1}{n}\mathbf{1}_n'(\mathbf{I}_n \otimes \mathbf{c}')\mathbf{Y} \\
&= \frac{1}{n}(\mathbf{1}_n' \otimes \mathbf{I}_1)(\mathbf{I}_n \otimes \mathbf{c}')\mathbf{Y} \\
&= \frac{1}{n}(\mathbf{1}_n'\mathbf{I}_n) \otimes (\mathbf{I}_1\mathbf{c}')\mathbf{Y} \\
&= \frac{1}{n}(\mathbf{1}_n' \otimes \mathbf{c}')\mathbf{Y} \\
&= \frac{1}{n}(\mathbf{c}' \otimes \mathbf{1}_n')\mathbf{Y} \\
&= \frac{1}{n}(\mathbf{c}'\mathbf{I}_m) \otimes (\mathbf{I}_1\mathbf{1}_n')\mathbf{Y} \\
&= \frac{1}{n}(\mathbf{c}' \otimes \mathbf{I}_1)(\mathbf{I}_m \otimes \mathbf{1}_n')\mathbf{Y} \\
&= \mathbf{c}'\frac{1}{n}(\mathbf{1}_n' \otimes \mathbf{I}_m)\mathbf{Y} = \mathbf{c}'\overline{\mathbf{Y}}
\end{aligned}
\tag{2}
$$

4

where $\mathbf{1}_n$ is a column vector of size n, $\mathbf{I}$ is the identity matrix and $\mathbf{c}$ is as noted in equation (1). Thus the estimator of μ_{AUC} can be expressed as a linear combination of sample means at each time point.

$\square$

Corollary 1. *The estimate of mean AUC has an expected value of $\mathbf{c}'\mu$, where the elements of $\mathbf{c}$ are as noted in (1).*

Proof. Using (2), $E(\hat{\mu}_{AUC}) = E(\mathbf{c}'\overline{\mathbf{Y}}) = \mathbf{c}'\mu$, where $\mu = (\mu_1, \mu_2, \cdots, \mu_m)'$. $\square$

Corollary 2. *The variance of the estimate of mean AUC can be written as a function of variances and covariances associated with the mean outcome values at each time point.*

Proof. Using (2), the variance of the estimate of the mean AUC can be formulated as

$$Var(\hat{\mu}_{AUC}) = \mathbf{c}'Var(\overline{\mathbf{Y}})\mathbf{c} \tag{3}$$

Denoting the variance of $\overline{Y}_{\cdot j}$ by σ_{jj}^2 and the covariance between $\overline{Y}_{\cdot j}$ and $\overline{Y}_{\cdot k}$ by σ_{jk}^2 , (3) becomes

$$Var(\hat{\mu}_{AUC}) = \sum_{j=1}^{m} c_j^2 \sigma_{jj}^2 + 2 \sum_{j=1}^{m-1} \sum_{k>j}^{m} c_j c_k \sigma_{jk}^2 \tag{4}$$

$\square$

Expressing the mean AUC as the linear combination of the mean values at each sample time point leads us to the consideration of the use of alternative methods for inference on the mean AUC. Our proposed method is intended to be flexible in that it can be extended to various experimental designs, accomodate confounding variables and controls for baseline values in randomized settings. It also allows the covariates to vary over clusters and response to vary over clusters after controlling for covariates.

2.2 Linear model-based inferences

We now focus our attention to simple parallel group studies with a goal of testing equality of group means. In the experiments of interest, units are randomized to r groups and outcomes are obtained at prespecified time points. A null hypothesis of interest is

$$H_0 : \mu_{AUC,k} = \mu_{AUC,k'} , \ k \neq k' . \tag{5}$$

Let Y_{ijk} be the response obtained from the ith individual at the jth time point in the kth group ($i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$ and $k = 1, 2, \ldots, r$). Furthermore let $\mathbf{Y} = \left(\mathbf{Y}_1', \mathbf{Y}_2', \ldots, \mathbf{Y}_r'\right)'$, where $\mathbf{Y}_k = \left(Y_{i1k}, Y_{i2k}, \ldots, Y_{imk}\right)'$ is a column vector of size nm. It follows that $E(\mathbf{Y}) = \theta$, where $\theta = \mu \otimes \mathbf{1}_{nm}$.

5

We may then specify a linear relationship of the form $\mathbf{Y} = \mathbf{W}\boldsymbol{\mu} + \boldsymbol{\epsilon}$, where $\mathbf{W} = \mathbf{1}_r \otimes \mathbf{1}_n \otimes \mathbf{I}_m$. Now letting $\boldsymbol{\beta} = \mathbf{M}\boldsymbol{\mu}$, where $\mathbf{M}$ is a non-singular matrix, we may assume a linear model of the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{6}$$

where $\mathbf{Y}$ is a (nmr) column vector of raw outcome values obtained for each individual at each time point, $\mathbf{X} = \mathbf{W}\mathbf{M}^{-1}$ is the design matrix for fixed effects, $\boldsymbol{\beta}$ is a $[r + m + (r \text{ x } m)]$ column vector of unknown fixed effects parameter vector representing group membership, categorical time points and the interaction between treatment group and time, $E(\boldsymbol{\epsilon}) = 0$ and $V(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}$. The expectation and variance of $\mathbf{Y}$ can thus be written as, $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ and $V(\mathbf{Y}) = \boldsymbol{\Sigma}$, where the matrix $\boldsymbol{\Sigma}$ is a block diagonal matrix with the within-subject covariance structure representing blocks on the main diagonal and zeros elsewhere.

Definition 1. *Under model* (6), *the population mean AUC of the rth treatment group is $\mu_{AUC,r} = g(\boldsymbol{c})\boldsymbol{\beta} = \boldsymbol{d}'\boldsymbol{\beta}$, where $\boldsymbol{c}$ is as defined in* (1) *and $g(c) = \boldsymbol{c}' \begin{bmatrix} 1_m & I_m & I_m \end{bmatrix}$ for the parameter levels of interest.*

When the parameterization is such that the model is full rank, $\mu_{AUC} = \mathbf{d}'\boldsymbol{\beta}$. For the less than full rank parameterization of the model, $\mu_{AUC} = \mathbf{d}'A\boldsymbol{\beta} \neq \mathbf{d}'\boldsymbol{\beta}$, where $A = G^-G$ is a matrix and G^- is the generalized inverse such that it satisfies the condition $GG^-G = G$. Then the estimate of AUC obtained is not unique and depends on the particular generalized inverse used. However, inferences are based on the reparameterized full rank model ensuring a unique solution. For a general linear model of the form in equation (6), where $\boldsymbol{\beta}$ consists of parameters representing categorical predictors, the estimate $\widehat{\boldsymbol{\beta}}$ consists of sample means of observations from their respective predictor populations. Using Theorem 1, we propose estimating the mean area of the rth treatment group using $\mathbf{d}'\widehat{\boldsymbol{\beta}}$. Taking advantage of the fact that the mean AUC can be written as a function of the mean outcome values as seen in Theorem 1, we can implement a repeated measures model using the raw outcome means as responses. Utilizing the vector of coefficients $\mathbf{c}$ defined in equation (1), appropriate contrasts can be constructed to test for the hypothesis of equality of mean AUC's.

Theorem 2. *Assume model* (6) *and furthermore assume $\epsilon \sim N(0, \boldsymbol{\Sigma})$. Then the estimate of the mean area in group r, $\widehat{\mu}_{AUC,r} = \boldsymbol{d}'\widehat{\boldsymbol{\beta}}$. The distribution is given by $\boldsymbol{d}'\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{d}'\boldsymbol{\beta}, \boldsymbol{d}'Var(\widehat{\boldsymbol{\beta}})\boldsymbol{d})$.*

Proof. Under the assumption $\epsilon \sim N(0, \boldsymbol{\Sigma})$, it follows that $\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, Var(\widehat{\boldsymbol{\beta}}))$. Then according to multivariate theory, the linear combination of the estimate, $\mathbf{d}'\widehat{\boldsymbol{\beta}}$ is also normally distributed with $\mathrm{E}(\widehat{\mu}_{AUC,r}) = \mathbf{d}'\boldsymbol{\beta} = \mu_{AUC,r}$ and $\mathrm{Var}(\widehat{\mu}_{AUC,r}) = \mathbf{d}'Var(\widehat{\boldsymbol{\beta}})\mathbf{d}$. $\qquad\square$

Remark 1. *Under the relaxed assumption $\epsilon \sim G(0, \boldsymbol{\Sigma})$, where G represents a general distribution, the distribution of $\widehat{\boldsymbol{\beta}}$ can be approximated by the normal distribution under mild regularity conditions. Thus the estimate $\boldsymbol{d}'\widehat{\boldsymbol{\beta}}$ is also asymptotically normally distributed.*

Based on the above distributional assumptions, hypothesis tests about the mean areas can be conducted.

Theorem 3. *If the estimate of the mean AUC for group k and group k' is represented using $\widehat{\mu}_{AUC,k}$ and $\widehat{\mu}_{AUC,k'}$ respectively, then the difference in the mean areas of the two treatment groups ($\mu_{AUC,k-k'}$) can be estimated as $\widehat{\mu}_{AUC,k-k'} \sim N(\mu_{AUC,k} - \mu_{AUC,k'}, Var(\widehat{\mu}_{AUC,k} - \widehat{\mu}_{AUC,k'}))$.*

Proof.

$$
\begin{aligned}
E(\widehat{\mu}_{AUC,k-k'}) &= E(\widehat{\mu}_{AUC,k} - \widehat{\mu}_{AUC,k'}) \\
&= E(\widehat{\mu}_{AUC,k}) - E(\widehat{\mu}_{AUC,k'}) \\
&= \mu_{AUC,k} - \mu_{AUC,k'}.
\end{aligned}
$$

$$
V(\widehat{\mu}_{AUC,k-k'}) = Var(\widehat{\mu}_{AUC,k} - \widehat{\mu}_{AUC,k'}).
$$

$\square$

Our goal to test for equality of AUC means is defined as $H_0 : \mu_{AUC,k-k'} = 0$ vs. $H_0 : \mu_{AUC,k-k'} \neq 0$. Using Theorems 1-3, the null hypothesis can be expressed in terms of the vector of coefficients $\mathbf{c}$ and the parameter vector $\boldsymbol{\beta}$.

Estimation of $\boldsymbol{\beta}$ involves the estimation of the unknown variance and covariance components in $\boldsymbol{\Sigma}$, which is approached via the restricted maximum likelihood approach (REML) along with an iterative algorithm to optimize the likelihood function. Though likelihood ratio tests for fixed-effect parameters are not appropriate using REML, it offers several advantages over the maximum likelihood (ML) approach, one of them being the elimination of bias in the estimation of variance parameters for small samples. See Verbeke and Molenberghs [12], Pinheiro and Bates (2000) [13] and Morrel (1998) [14] for more details. After the estimation of variance components, $\widehat{\boldsymbol{\Sigma}}$ is then used in the estimation of the fixed effect $\boldsymbol{\beta}$. The estimate of $\boldsymbol{\beta}$,

$$
\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{Y} \tag{7}
$$

is obtained using the generalized least squares methodology [15]. The estimate of the variance of $\widehat{\boldsymbol{\beta}}$ is given by

$$
\widehat{V(\widehat{\boldsymbol{\beta}})} = (\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1} \tag{8}
$$

In the event of a less than full rank matrix, estimation of $\boldsymbol{\beta}$ is performed using the generalized inverse denoted by $(.)^-$, where $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^-\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{Y}$. The estimate of mean AUC is then denoted by $\widehat{\mu}^*_{AUC} = \mathbf{d}A\boldsymbol{\beta}$, which depends on the choice of the generalized inverse A. When the variance-covariance matrix is not of full rank, estimation is performed using the generalized inverse. The test statistic is given by

$$
F = \frac{(\mathbf{d}'\widehat{\boldsymbol{\beta}})'[\mathbf{d}'(\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{d}]^{-1}(\mathbf{d}'\widehat{\boldsymbol{\beta}})}{rank(\mathbf{d})} \sim F_{(rank(\mathbf{d}),df)}.
$$

Though the test statistics asymptotically follow a chi-squared distribution, in practice statistical inferences are based on the F distribution with the denominator degrees of freedom specified using various methods such as the generalized Satterthwaite approximation [16] and the Kenward-Roger approximation (KR) [17] to name a few. The KR method to estimate the degrees of freedom improves the estimate for $\boldsymbol{\Sigma}$ and uses a modified test statistic F^* to account for small sample bias and variability in $\widehat{\boldsymbol{\Sigma}}$ [18]. Refer to Kenward and Roger [17] for further details. For a detailed outline of linear models and their implementation, refer to Littell [19], Diggle et al. [20] and Verbeke and Molenberghs [12].

2.3 Missing observations

In clinical research the presence of missing data is frequent. Missing data may arise due to administrative errors, drop-outs, missed appointments, etc. Missing data, especially if encountered in large amounts can lead to substantially biased inferences. Following Rubin's classification scheme ([21], [22]), the missing mechanism may be categorized into Missing At Random (MAR), Missing Completely at Random (MCAR) and Not Missing At Random (NMAR). MCAR occurs when missingness is unrelated to any of the data values, missing or observed. MAR occurs when missingness is related to values that have been observed. NMAR occurs when the cause of missingness is related to unobserved data and is considered to be a case of non-ignorable missingness.

The impact of missing data in the assessment of mean AUC in the context of bioequivalence has been discussed by Chow and Liu [23] and Donner et al. [24]. Though no concrete solution has been offered, Donner et al. [24] demonstrates how incomplete profiles impact bias and variance of the estimated AUC and hence the assessment of bioequivalence. The varying influence of different missingness patterns have also been discussed. Chow and Liu [23] note how the presence of missing values between two end sampling points have little effect on the comparison of bioavailability. However a substantial bias is seen when missing values occur at the two end sampling points.

Under MCAR assumption, the simplest strategy to deal with missing values for the proposed test for equality of AUC means is to discard those instances. To test for equality, Spritzler et al. [25] proposes an estimator for the mean AUC using indicator functions for missing observations. When only few measurements are missing, another strategy is to utilize imputation methods. Commonly, the last observation carried forward (LOCF) method is applied. In our study, we investigate the influence of missing observations on the inferential properties of the test.

Remark 2. *Allison [26] notes how under the assumption of MCAR, parameter estimates obtained using available data are consistent. Thus under the assumption of MCAR, there may be a loss of efficiency and power, but our model parameter estimates (6) and hence the estimate of the mean AUC obtained using our proposed approach is unbiased, i.e. $E(\boldsymbol{d}'\hat{\boldsymbol{\beta}}) = \boldsymbol{d}'\boldsymbol{\beta}$.*

3   Inclusion of covariates

In the non-randomized setting, differences observed in outcomes between treatment groups cannot always be attributed solely to the intervention applied, but may also be explained by the inherent variability among the groups caused by confounding variables, e.g. gender associated differences in the plasma drug concentrations of various antiretroviral therapies [27]. In a randomized setting, randomization aims to dissolve imbalances in confounding variables between treatment groups thus rendering the groups comparable. Hence any observed difference in responses among groups can be attributed as being so due to the intervention. However in practice, covariate imbalances are witnessed even with randomization. In a randomized setting, covariates perceived to influence the outcome can be identified prior to the study and accounted for in the analysis allowing for chance imbalances and efficiency gains. For example, analysis of a continuous outcome using ANCOVA reduces error variability resulting in more precise estimates of the treatment effect. Apart from comparability of groups, treatment differences as a function of covariates may also be of interest.

Generally, confounding variables are accounted for by their inclusion in statistical models. Measurable independent subject factors, categorical or continuous in nature, can be included in the model as illustrated below. Borrowing the notation used in Section 2, we can formulate the model as

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \tag{9}$$

where $\mathbf{Y}$ and $\boldsymbol{\epsilon}$ are defined as before, $\mathbf{X}_1$ is the design matrix for fixed effects treatment, time and their interaction, $\boldsymbol{\beta}_1$ is its corresponding unknown fixed effects parameter vector, $\mathbf{X}_2$ is the design matrix for other between-subject possibly time-invariant subject factors such as gender, age etc., and $\boldsymbol{\beta}_2$ is its corresponding unknown fixed effects parameter vector. Constraining all the fixed effects to a single vector of parameters $\boldsymbol{\beta} = (\boldsymbol{\beta}_1' \ \boldsymbol{\beta}_2')'$ and $\mathbf{X} = (\mathbf{X}_1 \ \mathbf{X}_2)$, the model can be reformulated as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{10}$$

Inclusion of covariates in the model allow us to predict the mean response for a subject associated with a particular set of covariate values, $\mathbf{X}_h$, and the estimated conditional mean in this case may be written as $\hat{Y}_{\cdot j(X_h)} = X_h\hat{\beta}$. Clinically relevant or study motivated values of covariates can be plugged in to acquire the predicted AUC for a specific class of subjects. Making use of the vector of adjusted conditional means, $\widehat{\mathbf{Y}}_h = \left(\widehat{Y}_{1,X_h}, \widehat{Y}_{2,X_h}, \ldots, \widehat{Y}_{m,X_h}\right)'$, we define the adjusted mean AUC in group i to be $\mu_{AUC(X_h),r} = \mathbf{d}'\widehat{\mathbf{Y}}_h$. In the comparison of two groups, null hypotheses of interest may then be of the form,

$$H_0 : \mu_{AUC(X_h),1} = \mu_{AUC(X_h),2} \ , \tag{11}$$

where $\mu_{AUC(X_h),1}$ and $\mu_{AUC(X_h),2}$ are the adjusted mean AUCs in groups 1 and 2, respectively.

4   Simulation Study

For the purpose of examining the properties of the proposed approach in testing the equality of mean AUCs, we conducted a simulation study. We generated outcomes following the mean vector $\boldsymbol{\mu}_1 = (2, 2.5, 3, 3, 2.5, 2)'$ measured at six time points $\mathbf{t} = (1, 2, 3, 4, 5, 6)'$. Using the Cholesky approach detailed by Ripley [28], we generated a multivariate vector with mean $\boldsymbol{\mu}_1$ with correlation modeled under the homogeneous or heterogeneous first-order autoregressive structure (AR(1) or ARH(1)) structure depending on whether variance homogeneity or heterogeneity was assumed. To take into account variance heterogeneity within subjects, simulations were also performed with the coefficient of variation (cv) measure ranging from 0.1 to 0.5. For example, for a cv equal to 0.2, the corresponding variance for the six time points would be equal to $(0.4, 0.5, 0.6, 0.6, 0.5, 0.4)'$. In less than 2% of the cases, simulations involving a high coefficient of variation generated negative outcome values which were set to zero. Under the homogeneous variance assumption, variance σ^2 was set to be equal to 0.1. The within-subject correlation parameter ρ was allowed to range from 0.3 to 0.7.

For each iteration, $n = 30$ random samples were randomized equally to two groups with mean vectors μ_1 and $\boldsymbol{\mu}_2 = \gamma\boldsymbol{\mu}_1$, where the coefficient γ ranged from 1.00 to 1.20 ($\gamma = 1.00$ represents the null scenario). Simulations were performed for 10,000 iterations for each combination of ρ and γ. A repeated measures model (PROC MIXED in SAS) with covariates group, time and their interaction were fit with appropriate contrasts constructed to test for equality of areas. Inferential properties of our approach such as power, actual Type I error rate, coverage and the bias and standard error of estimates were evaluated. Covariance structures such as compound symmetric (CS), unstructured (UN), and the true covariance structure were assumed. Properties of the repeated measures model selected using the Akaike's Information Criterion (AIC) [29] were also studied. Although the appropriate choice of the denominator degrees of freedom is debatable, good small sample properties of the Kenward-Roger degrees of freedom dictated its use in our analysis [30]. For the sake of comparison, results based on the traditional approach being the standard two-sample t-tests are included.

Our results demonstrated comparable properties for all methods. Among all the methods, the model assuming the compound symmetric structure exhibited inflated false positive rates. A decrease in power was also seen with an increase in the coefficient of variation for heterogeneous models. Estimates of the difference in areas were observed to be unbiased with comparable standard errors and both approaches maintained coverage probability.

To evaluate the effect of missing values, outcomes at each time point were generated under a constant missing probability, p. Simulations were performed for p ranging from 0% to 30%. Various combinations

of ρ and p under the homogeneous and heterogeneous variance assumptions yielded a total of 60 simulation scenarios. Simulation results obtained under the homogeneous and heterogeneous variance assumption are summarized in Tables 1 - 4. Power curves for the different methods have been graphically illustrated in Figures 1 - 4. Although power curves of all the methods were comparable when complete data was available, it can be noted that under both the AR(1) and ARH(1) assumption, an increase in p translated to an increase in distance between power curves of the t-test and the proposed approach. Though the t-test maintained coverage and Type I error in the presence of missing data, the estimates were found to be slightly biased with larger standard errors and wider confidence intervals. The bias, standard error and confidence interval width were seen to increase with increase in missingness.

A few other features of this simulation study are worth mentioning. Convergence criteria was met satisfactorily in all but UN models, with the lack of convergence increasing with p. The proportion of negative observations generated in our simulations for each scenario ranged from 0 - 2.5%, the maximum being observed under $cv = 0.5$. Inflated Type 1 error was observed in CS models in the presence of heterogeneous variances and a decrease in power was noted with increasing correlation. The power curve obtained using the model selected by the AIC did not appear to be visually different from the model following the true covariance structure. Using the repeated measures model has its disadvantages in that it requires the specification of the covariance structure to be used. Our simulations have demonstrated that in such cases, the model selection can be guided by the AIC without experiencing a loss of power. Our simulations also demonstrate that when compared with the two sample t-test, the proposed repeated measures analysis leads to a more powerful test for equality of area under the curves for unbalanced data.

5   Application: Acupuncture Study

Consider a randomized controlled trial evaluating a short course of traditional acupuncture compared with usual care for persistent non-specific low back pain conducted by three private acupuncture clinics and 18 general practices in York, England [31]. The trial enrolled 241 adults aged 18-65 with non-specific low back pain of 4-52 weeks duration. The intervention consisted of 10 individualised acupuncture treatments from one of six qualified acupuncturists or usual care only. The primary outcome variable was the bodily pain dimension of the SF-36 health status questionnaire, measured at 12 and 24 months. Secondary outcomes consisted of quality of life scores obtained using the remaining dimensions of the SF-36 administered at baseline and at 3, 12, and 24 months. Walters [32] considered summarizing the SF-36 bodily pain scores over the 2-year follow-up using the AUC based on the 173 patients who completed all four assessments. Individual profile plots, histograms of the distribution of the AUC, and mean plots for each treatment group are presented in Figures 5 and 6. The variable SF-36 is scored on a 0 to 100 scale where 100 implies no pain and a score of 0 indicates the most severe score.

The proposed approach discussed in the manuscript was applied and was based on a model which included parameter effects for treatment, time, treated as a categorical variable, and their interaction. From the fitted model estimated mean scores at the four time points 0, 3, 12 and 24 months was obtained and the mean AUC for each intervention group was estimated using contrasts specified in Table 5. The within-subject covariance was modelled using the structures AR(1), UN and CS. Analysis was considered for the complete data available as well as for completers, where a completer is defined as a subject with all four assessment values available. For comparison, the traditional approach where the AUC is calculated using the trapezoidal method for every individual and groups are then compared using a two sample t-test was applied.

The results from our analysis using the complete data and using only patients that completed all four assessments are provided in Table 6 and 7, respectively. Analysis of complete data suggested a significant difference in mean areas favouring the acupuncture group. The AIC statistic computed revealed the UN model to be the best fit. Using only data from completers, the UN and t-test methods demonstrated only marginal significance.

To illustrate the flexibility of our proposed approach, the differential effect of patient age and its interaction with the treatment group was investigated. The model with the additional terms age and its interaction with treatment was fit and the vector of estimated conditional means at each time point were obtained. The adjusted mean AUC for each intervention as a function of age appears in Figure 7. A higher estimated mean AUC was revealed for the acupuncture group with the AUC decreasing with increase in age. It is also observed that the difference in mean AUCs between the usual care and the acupuncture group decreases with increase in the covariate age.

6 Discussion and Conclusion

In this paper we introduced a general linear model approach that encapsulates the entire subject profile to make inferences on the group AUC means. Inferential properties of the test of group means under the proposed and the standard approach were evaluated using simulation studies. For simplicity, our simulations were conducted by generating a hypothetical means model assuming a variety of covariance structures. Little difference was noted between in the inferential properties of the two approaches when the complete data was observed. However, our approach was seen to produce a significant power advantage over traditional methods in conjunction with missing data, with the distance between power curves widening with increased missingness. In the presence of missingness, though the traditional approach maintained coverage consistently, it was observed to produce biased estimates with larger standard errors and wider confidence intervals.

Our work can be extended by considering other distinct missingness profiles. When the distribution of the test statistic under the null is suspect, a permutation version of our approach can also be implemented.

For extremely small sample sizes, a bootstrap version can be considered. The proposed approach can be extended to the assessment of bioequivalence and observations encountering limit of detection. Research including the above mentioned is currently underway.

References

[1] Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Wiley, 2004.

[2] Everitt B, Pickles A. *Statistical aspects of the design and analysis of clinical trials*. World Scientific Publishing Company, 2004.

[3] Fairclough DL. *Design and analysis of quality of life studies in clinical trials*. Chapman & Hall/CRC Interdisciplinary Statistics, 2010.

[4] Fayers P, Machin D. *Quality of Life: The Assessment, Analysis and Interpretation of Patient-reported Outcomes*. Wiley, 2007.

[5] Clark CA, Gardiner J, McBurney MI, Anderson S, Weatherspoon LJ, Henry DN, Hord NG. Effects of breakfast meal composition on second meal metabolic responses in adults with type 2 diabetes mellitus. *European Journal of Clinical Nutrition* May 2006; **60**(9):1122–1129, doi:10.1038/sj.ejcn.1602427. URL http://dx.doi.org/10.1038/sj.ejcn.1602427.

[6] Yeh KC, Kwan KC. A comparison of numerical integrating algorithms by trapezoidal, Lagrange, and spline approximation. *Journal of Pharmacokinetics and Pharmacodynamics* Feb 1978; **6**(1):79–98, doi:10.1007/BF01066064. URL http://dx.doi.org/10.1007/BF01066064.

[7] Venter CS, Slabber M, Vorster HH. Labelling of foods for glycaemic index: advantages and problems. *South African Journal of Clinical Nutrition* 2003; **16**(4):118–126.

[8] Allison DB, Paultre F, Maggio C, Mezzitis N, Pi-Sunyer FX. The use of areas under curves in diabetes research. *Diabetes Care* Feb 1995; **18**(2):245–250, doi:10.2337/diacare.18.2.245. URL http://dx.doi.org/10.2337/diacare.18.2.245.

[9] Lu K, Mehrotra DV. Specification of covariance structure in longitudinal data analysis for randomized clinical trials. *Statistics in medicine* Feb 2010; **29**(4):474–488, doi:10.1002/sim.3820. URL http://dx.doi.org/10.1002/sim.3820.

[10] Ferron J, Dailey R, Yi Q. Effects of Misspecifying the First-Level Error Structure in Two-Level Models of Change. *Multivariate Behavioral Research* 2002; **37**(3):379–403, doi:10.1207/S15327906MBR3703_4. URL http://dx.doi.org/10.1207/S15327906MBR3703_4.

[11] Keselman HJ, Algina J, Kowalchuk RK, Wolfinger RD. A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics - Simulation and Computation* 1998; **27**(3):591–604, doi:10.1080/03610919808813497. URL http://dx.doi.org/10.1080/03610919808813497.

[12] Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data.* Springer Verlag (June 1, 2000), 2000.

[13] Pinheiro JC, Bates DM. *Mixed-Effects Models in S and S-PLUS.* Springer (April 2, 2002), 2000.

[14] Morrell CH. Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics* Dec 1998; **54**(4):1560–1568. URL `http://view.ncbi.nlm.nih.gov/pubmed/9883552`.

[15] Graybill FA. *Theory and Application of the Linear Model.* Duxbury Press; 1 edition (March 27, 2000), 2000.

[16] Fai AHT, Cornelius PL. Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of Statistical Computation and Simulation* 1996; **54**(4):363–378, doi:10.1080/00949659608811740. URL `http://dx.doi.org/10.1080/00949659608811740`.

[17] Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* Sep 1997; **53**(3):983–997. URL `http://view.ncbi.nlm.nih.gov/pubmed/9333350`.

[18] Schaalje GB, McBride JB, Fellingham GW. Approximations to Distributions of Test Statistics in Complex Mixed Linear Models Using SAS Proc MIXED. *Proceedings of the SAS Users Group International 26th Anunual Conference*, 2001.

[19] Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O. *SAS system for mixed models.* SAS Publishing; 2nd edition (February 21, 2006), 2006.

[20] Diggle PJ, Heagerty P, Liang KY, Zeger SL. *Analysis of Longitudinal Data.* Oxford University Press, USA; 2 edition (August 15, 2002), 2002.

[21] Little R, Rubin D. *Statistical Analysis with Missing Data.* New York: Wiley, 1987. URL `http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471183865.html`.

[22] Rubin DB. Inference and missing data. *Biometrika* Dec 1976; **63**(3):581–592, doi:10.1093/biomet/63.3.581. URL `http://dx.doi.org/10.1093/biomet/63.3.581`.

[23] Chow SC, Liu JP. *Design and analysis of bioavailability and bioequivalence studies.* CRC Press, 2008.

[24] Donner A, Hauck WW, Zou G. The impact of missing values in the concentration-time curve on the assessment of bioequivalence. *Pharmaceutical Statistics* 2005; **4**(2):91–99, doi:10.1002/pst.156. URL `http://dx.doi.org/10.1002/pst.156`.

[25] Spritzler J, DeGruttola VG, Pei L. Two-Sample Tests of Area-Under-the-Curve in the Presence of Missing Data. *The International Journal of Biostatistics* 2008; **4**(1). URL `http://www.bepress.com/ijb/vol4/iss1/1/`.

[26] Allison PD. *Missing Data (Quantitative Applications in the Social Sciences)*. Sage Publications, Inc, 2002.

[27] Ofotokun I, Chuck S, Hitti J. Antiretroviral pharmacokinetic profile: A review of sex differences. *Gender Medicine* Jun 2007; **4**(2):106–119, doi:10.1016/S1550-8579(07)80025-8. URL `http://dx.doi.org/10.1016/S1550-8579(07)80025-8`.

[28] Ripley BD. *Stochastic Simulation*. Wiley-Interscience; 1 edition (March 10, 2006), 1987.

[29] Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. *International Symposium on Information Theory, 2 nd, Tsahkadsor, Armenian SSR*, 1973; 267–281.

[30] Gomez EV, Schaalje GB, Fellingham GW. Performance of the KenwardRoger Method when the Covariance Structure is Selected Using AIC and BIC. *Communications in Statistics - Simulation and Computation* 2005; **34**(2):377–392, doi:10.1081/SAC-200055719. URL `http://dx.doi.org/10.1081/SAC-200055719`.

[31] Thomas KJ, MacPherson H, Thorpe L, Brazier J, Fitter M, Campbell MJ, Roman M, Walters SJ, Nicholl J. Randomised controlled trial of a short course of traditional acupuncture compared with usual care for persistent non-specific low back pain. *BMJ* Sep 2006; **333**(7569):623+, doi:10.1136/bmj.38878.907361.7C. URL `http://dx.doi.org/10.1136/bmj.38878.907361.7C`.

[32] Walters SJ. *Quality of Life Outcomes in Clinical Trials and Health-Care Evaluation: A Practical Guide to analysis and interpretation*. Wiley; 1 edition (November 23, 2009), 2009.
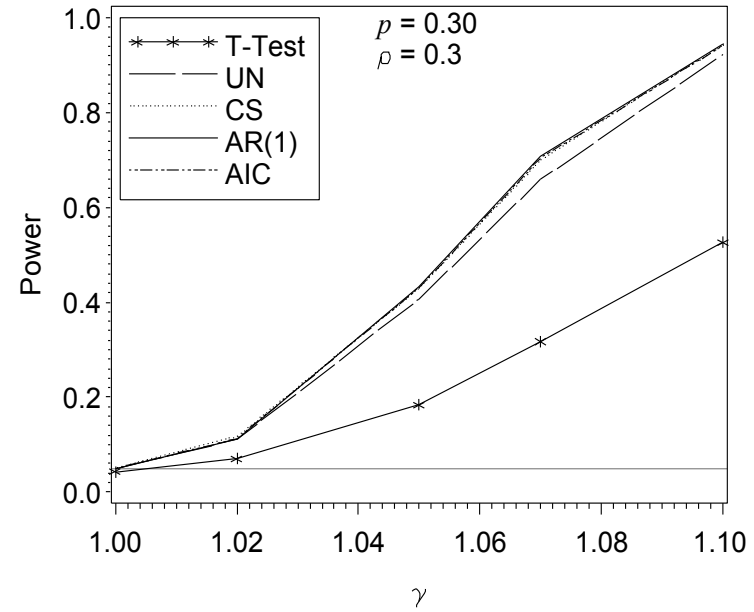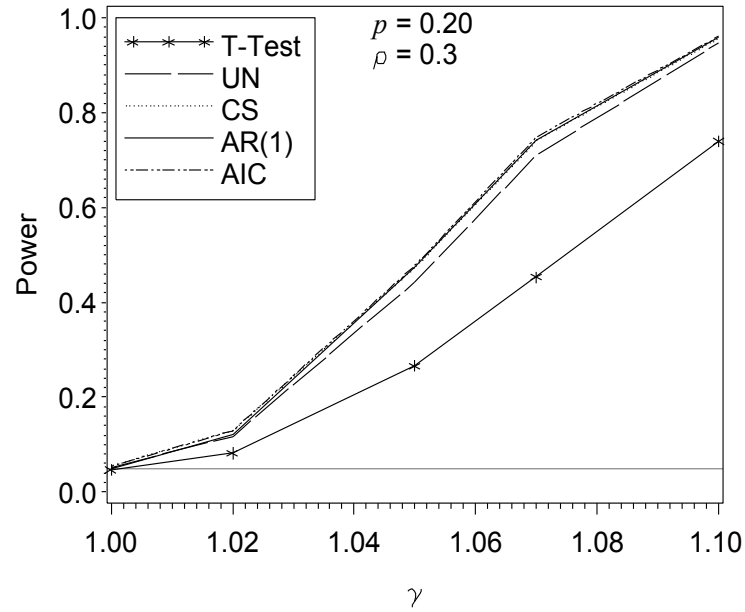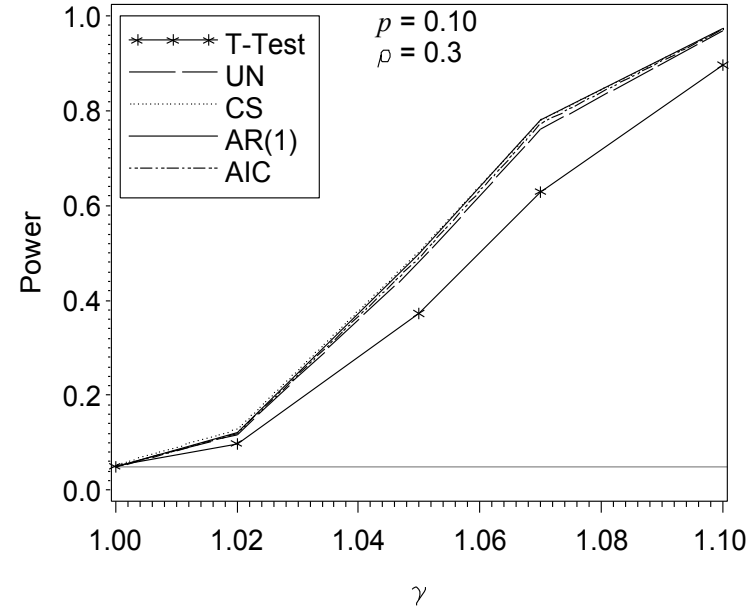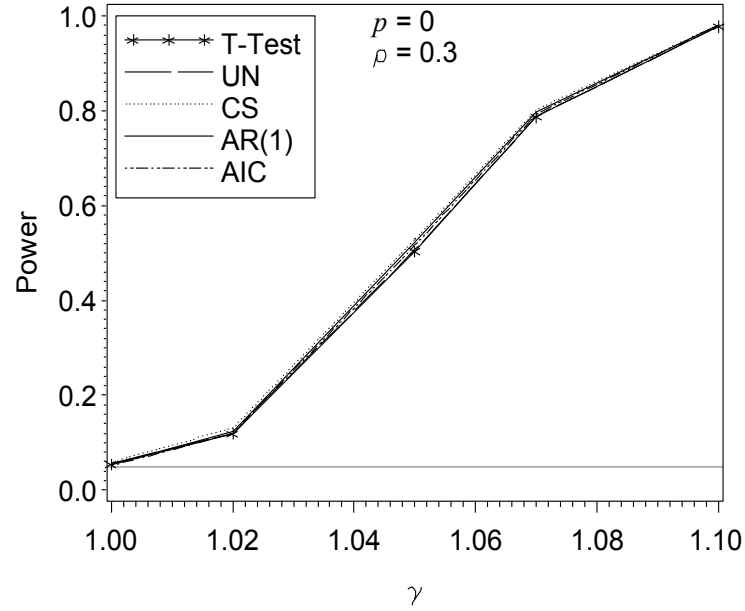
Figure 1: Comparison of power curves for tests for equality of area under the curves. Data was generated under the AR(1) assumption with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$, $\rho = 0.3$ and $\sigma^2 = 0.1$. Simulations were performed for 10,000 iterations.

Figure 2: Comparison of power curves for tests for equality of area under the curves. Data was generated under the ARH(1) assumption with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)$, $\rho = 0.3$ and $cv = 0.1$. Simulations were performed for 10,000 iterations.
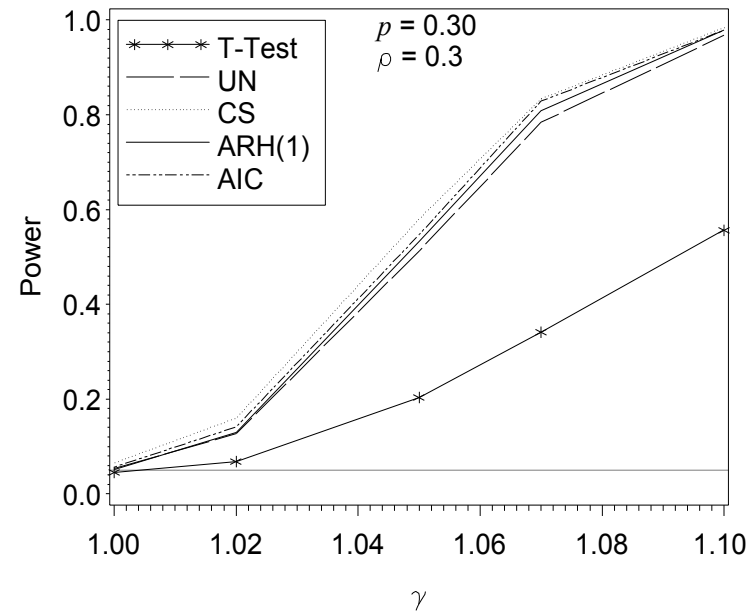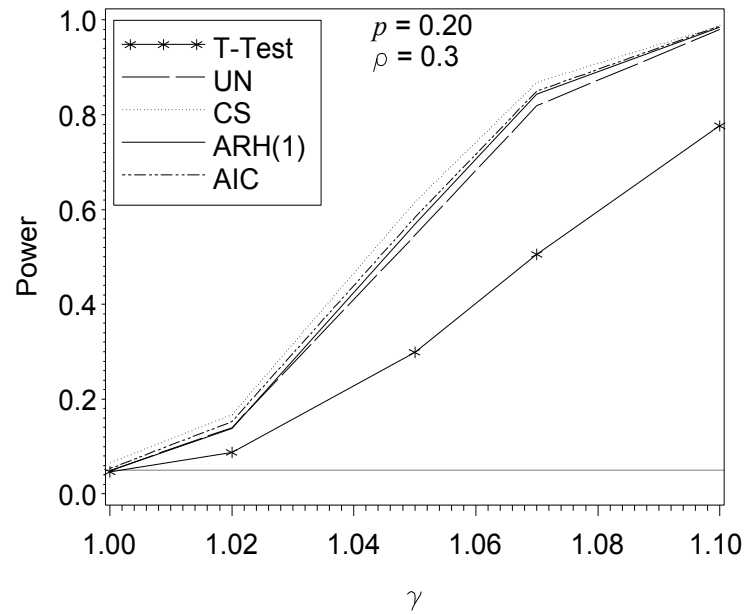
Figure 3: Comparison of power curves for tests for equality of area under the curves. Data was generated under the ARH(1) assumption with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$, $\rho = 0.5$ and $cv = 0.1$. Simulations were performed for 10,000 iterations.
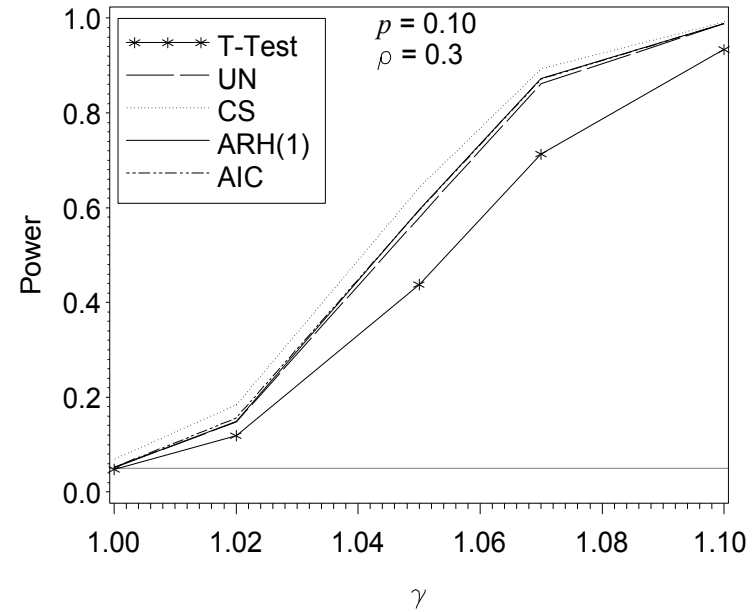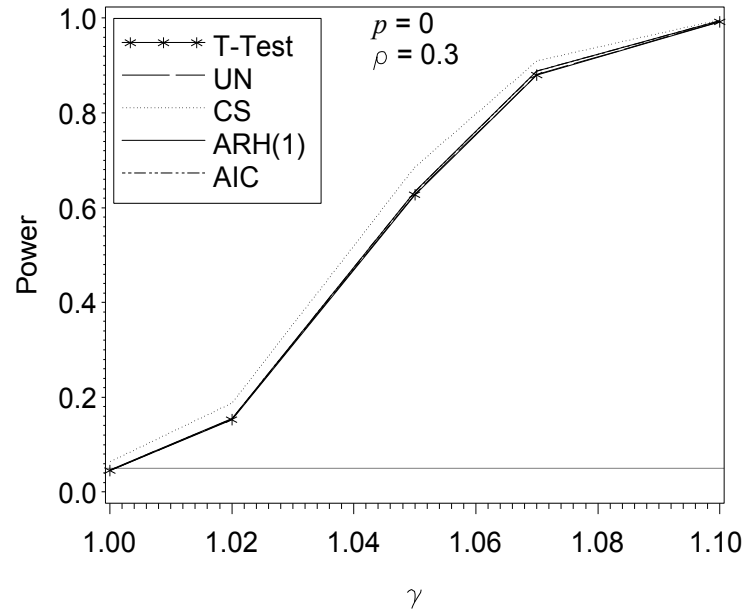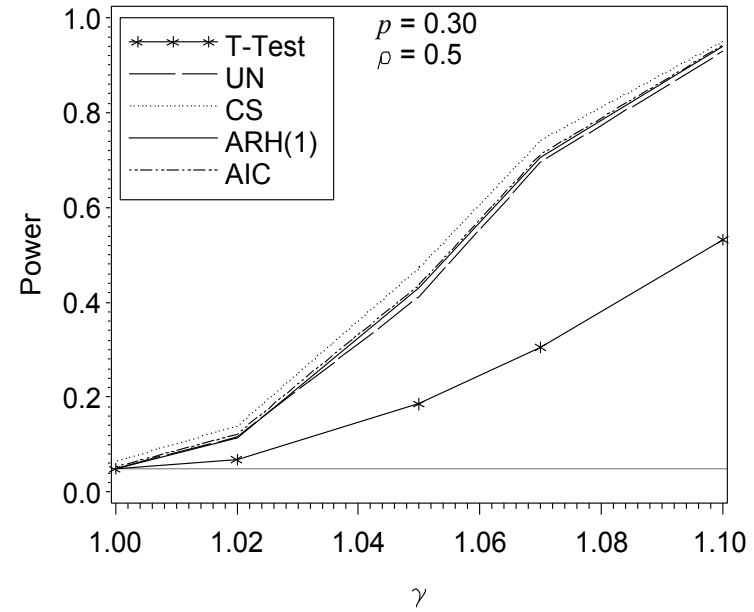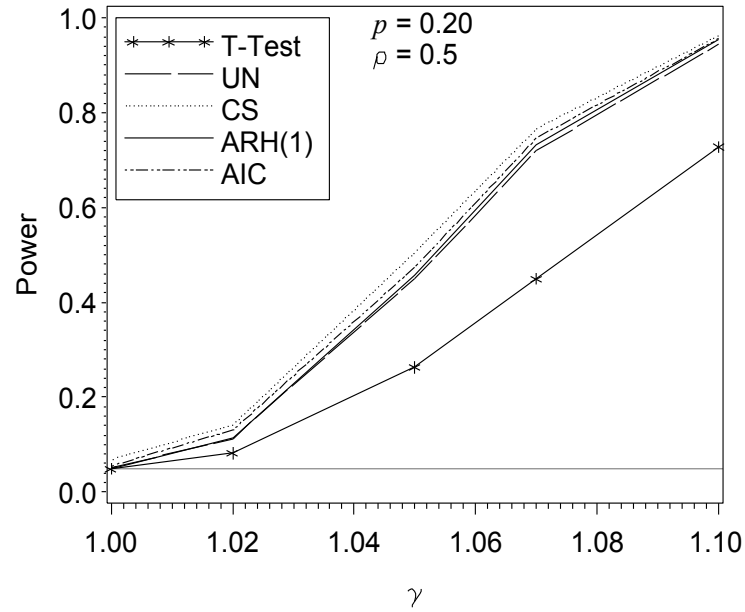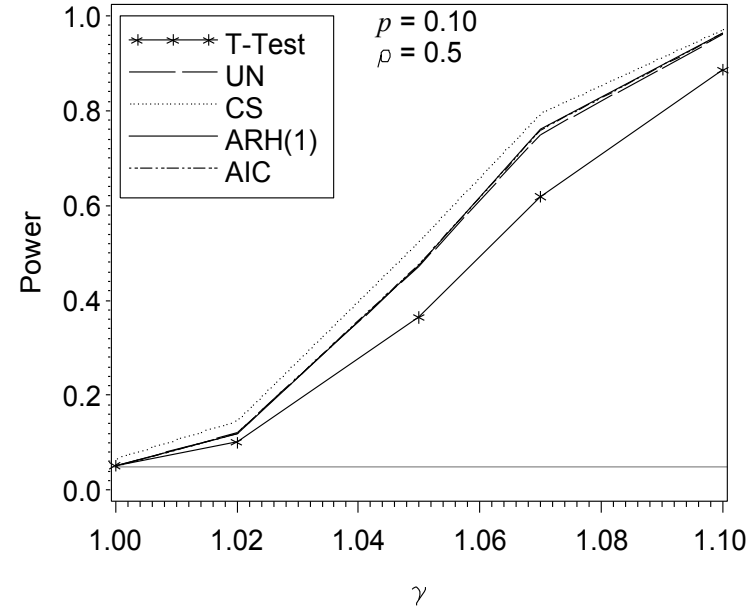
Figure 4: Comparison of power curves for tests for equality of area under the curves. Data was generated under the ARH(1) assumption with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$, $\rho = 0.7$ and $cv = 0.1$. Simulations were performed for 10,000 iterations.
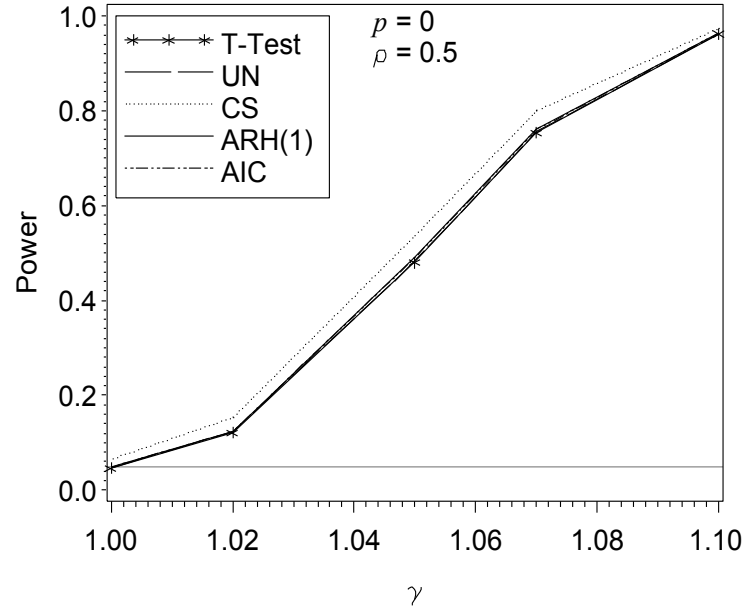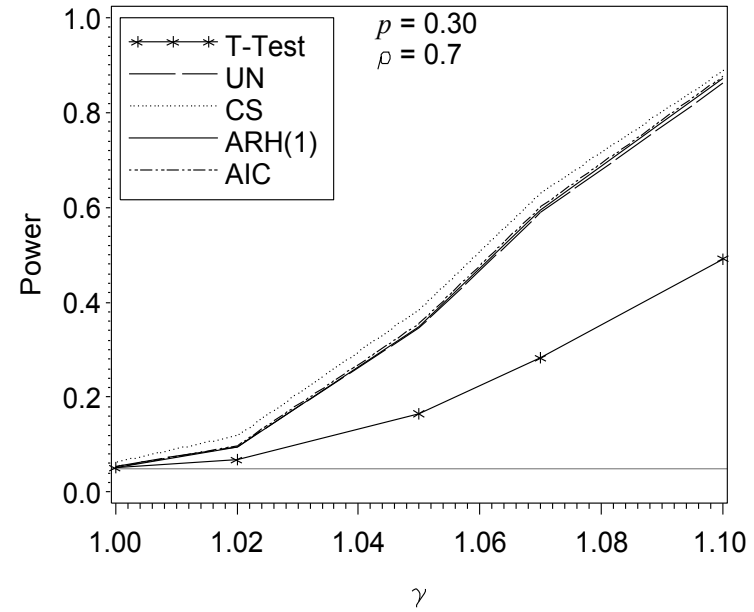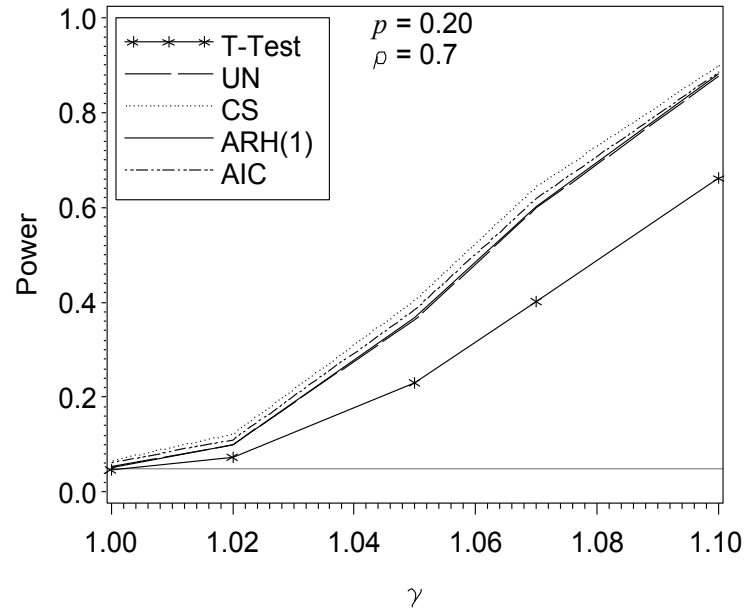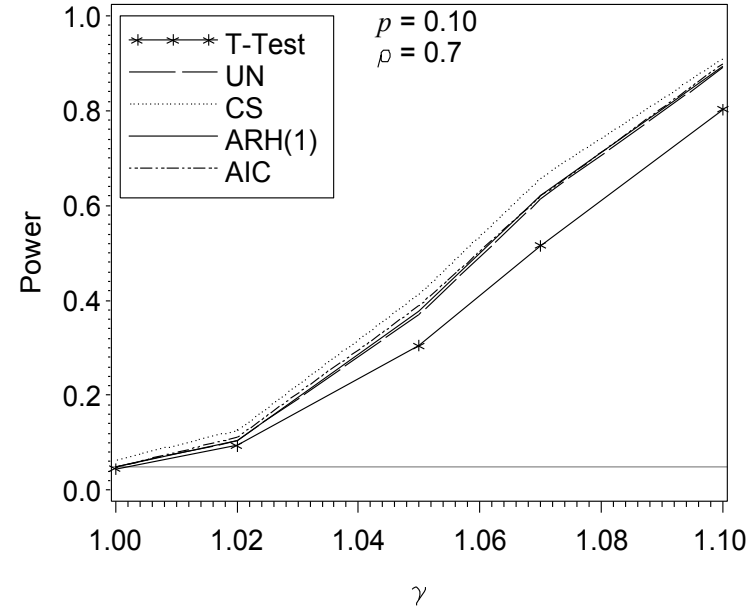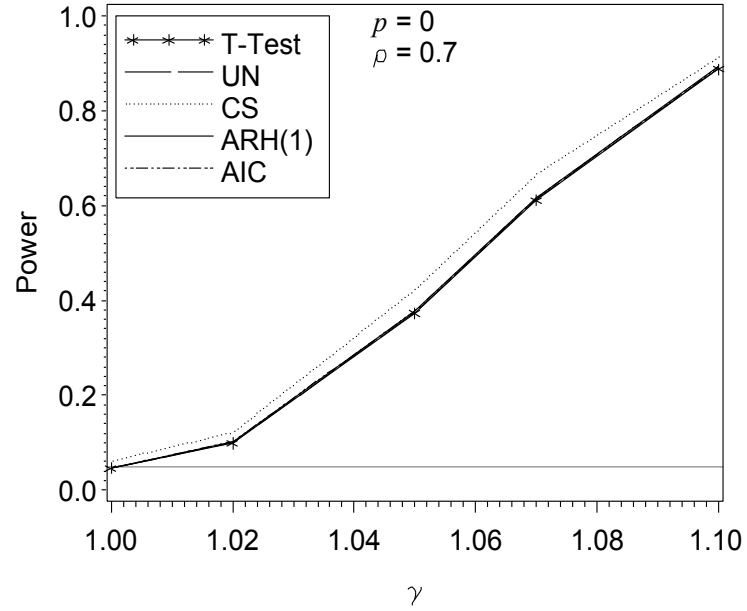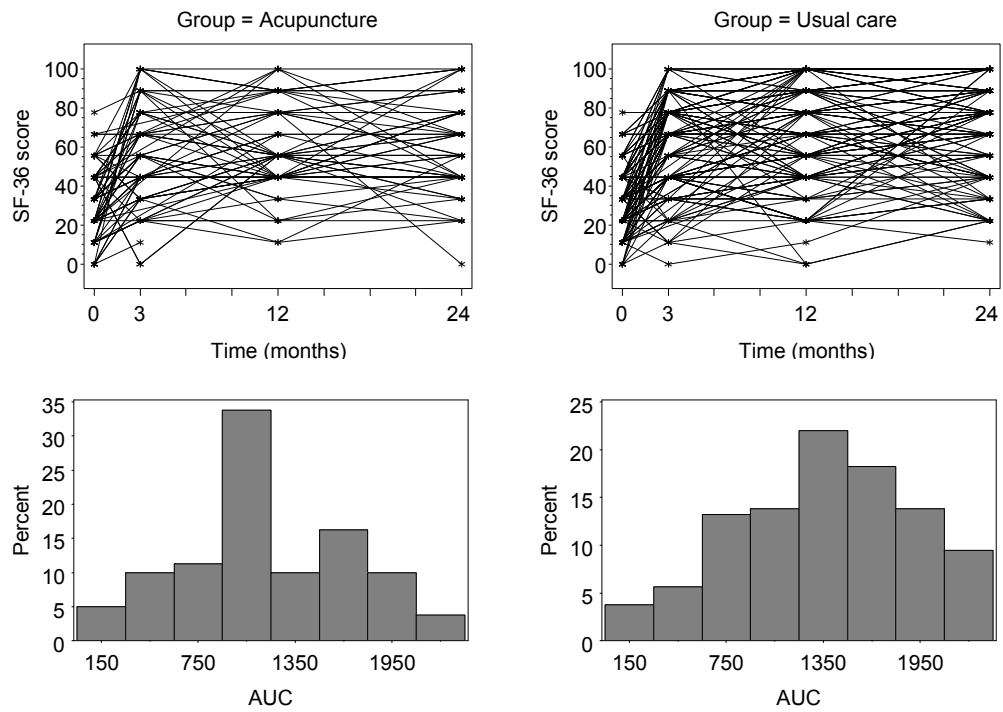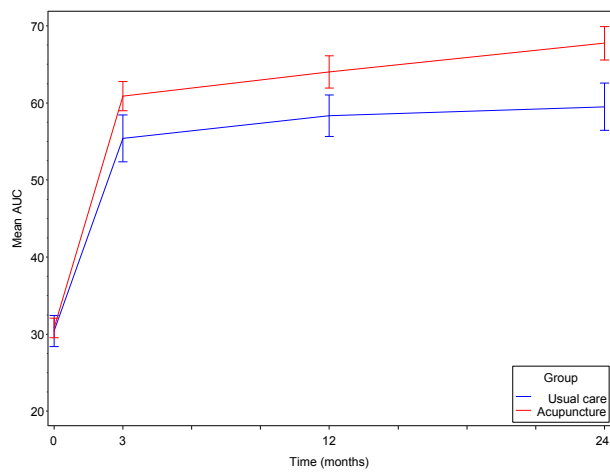
Figure 5: Exploratory plots

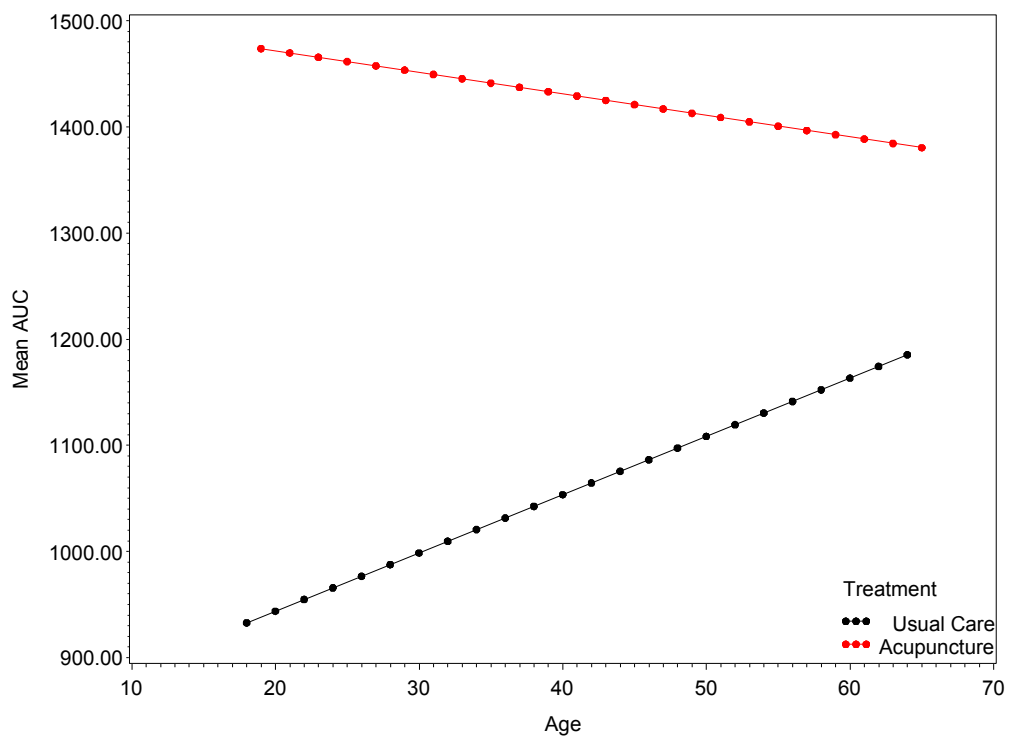

Figure 6: Mean plot of areas

Figure 7: Differential effect of covariate age by treatment group

Table 1: Comparison of Bias, Std, Coverage and CI width for the test of equality of areas under the curve. Data was generated under the AR(1) assumption with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$ and $\sigma^2 = 0.1$. Simulations were performed for 10,000 iterations.

p	γ	Bias (std)				Coverage (CI width)			
		UN	CS	AR(1)	T-Test	UN	CS	AR(1)	T-Test
$\rho = 0.3$									
0	1	0.002(0.315)	0.002(0.307)	0.002(0.315)	0.002(0.315)	0.945(1.289)	0.941(1.253)	0.947(1.269)	0.945(1.289)
	1.02	0.003(0.314)	0.003(0.306)	0.003(0.315)	0.003(0.314)	0.948(1.286)	0.943(1.25)	0.95(1.267)	0.948(1.286)
	1.05	0.003(0.315)	0.003(0.307)	0.003(0.315)	0.003(0.315)	0.951(1.289)	0.944(1.252)	0.95(1.268)	0.951(1.289)
	1.07	0.001(0.315)	0.001(0.307)	0.001(0.315)	0.001(0.315)	0.946(1.289)	0.94(1.252)	0.948(1.269)	0.946(1.289)
	1.1	0.006(0.315)	0.006(0.307)	0.006(0.315)	0.006(0.315)	0.952(1.29)	0.946(1.254)	0.951(1.269)	0.952(1.29)
0.2	1	0(0.342)	0(0.331)	0.001(0.337)	-0.002(0.445)	0.948(1.408)	0.945(1.352)	0.951(1.358)	0.952(1.825)
	1.02	-0.001(0.343)	-0.002(0.332)	-0.002(0.337)	0.01(0.449)	0.947(1.412)	0.94(1.355)	0.946(1.36)	0.953(1.837)
	1.05	-0.002(0.342)	-0.003(0.331)	-0.003(0.336)	0.029(0.451)	0.951(1.409)	0.948(1.354)	0.952(1.358)	0.951(1.849)
	1.07	0.006(0.342)	0.004(0.331)	0.004(0.337)	0.047(0.455)	0.949(1.409)	0.945(1.353)	0.951(1.359)	0.95(1.862)
	1.1	-0.001(0.342)	-0.002(0.331)	-0.002(0.336)	0.062(0.457)	0.947(1.408)	0.943(1.352)	0.946(1.357)	0.947(1.872)
$\rho = 0.7$									
0	1	0.004(0.442)	0.004(0.432)	0.004(0.441)	0.004(0.442)	0.948(1.812)	0.942(1.766)	0.948(1.789)	0.948(1.812)
	1.02	0.003(0.443)	0.003(0.433)	0.003(0.442)	0.003(0.443)	0.949(1.817)	0.944(1.77)	0.949(1.795)	0.949(1.817)
	1.05	0.006(0.443)	0.006(0.432)	0.006(0.442)	0.006(0.443)	0.949(1.815)	0.941(1.77)	0.946(1.794)	0.949(1.815)
	1.07	-0.006(0.443)	-0.006(0.433)	-0.006(0.442)	-0.006(0.443)	0.949(1.817)	0.943(1.771)	0.949(1.793)	0.949(1.817)
	1.1	-0.006(0.442)	-0.006(0.431)	-0.006(0.441)	-0.006(0.442)	0.946(1.811)	0.942(1.765)	0.946(1.789)	0.946(1.811)
0.2	1	-0.001(0.454)	-0.001(0.445)	-0.001(0.45)	-0.012(0.53)	0.95(1.861)	0.944(1.821)	0.946(1.827)	0.951(2.171)
	1.02	0.001(0.454)	0.004(0.444)	0.004(0.449)	0.018(0.532)	0.948(1.862)	0.944(1.815)	0.948(1.823)	0.952(2.181)
	1.05	0.001(0.453)	-0.001(0.444)	0(0.45)	0.03(0.535)	0.948(1.859)	0.943(1.819)	0.949(1.827)	0.949(2.194)
	1.07	0.005(0.454)	0.005(0.445)	0.004(0.45)	0.048(0.538)	0.953(1.861)	0.95(1.822)	0.954(1.829)	0.952(2.203)
	1.1	-0.002(0.454)	-0.002(0.444)	-0.003(0.449)	0.058(0.541)	0.95(1.861)	0.946(1.818)	0.948(1.825)	0.949(2.218)

Table 2: Comparison of Bias, Std, Coverage and CI width for the test of equality of areas under the curve. Data was generated under the ARH(1) assumption with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$, $cv = 0.1$ and $\sigma^2 = 0.1$. Simulations were performed for 10,000 iterations.

p	γ	Bias (std)				Coverage (CI width)			
		UN	CS	ARH(1)	T-Test	UN	CS	ARH(1)	T-Test
$\rho = 0.3$									
0	1	-0.003(0.267)	-0.003(0.248)	-0.003(0.268)	-0.003(0.267)	0.955(1.092)	0.936(1.011)	0.954(1.082)	0.955(1.092)
	1.02	-0.002(0.27)	-0.002(0.25)	-0.002(0.271)	-0.002(0.27)	0.952(1.104)	0.937(1.023)	0.954(1.095)	0.952(1.104)
	1.05	0(0.274)	0(0.255)	0(0.275)	0(0.274)	0.952(1.123)	0.935(1.039)	0.952(1.113)	0.952(1.123)
	1.07	-0.001(0.276)	-0.001(0.257)	-0.001(0.277)	-0.001(0.276)	0.95(1.131)	0.934(1.047)	0.952(1.121)	0.95(1.131)
	1.1	-0.001(0.281)	-0.001(0.261)	-0.001(0.282)	-0.001(0.281)	0.953(1.151)	0.939(1.066)	0.953(1.14)	0.953(1.151)
0.2	1	0.004(0.29)	0.003(0.268)	0.003(0.287)	0.003(0.414)	0.951(1.195)	0.934(1.093)	0.952(1.164)	0.953(1.697)
	1.02	-0.001(0.293)	0(0.27)	0(0.29)	0.011(0.416)	0.947(1.208)	0.93(1.103)	0.948(1.175)	0.952(1.705)
	1.05	0.003(0.297)	0.002(0.274)	0.002(0.294)	0.032(0.422)	0.945(1.222)	0.929(1.118)	0.949(1.192)	0.952(1.731)
	1.07	0.001(0.3)	0(0.276)	0(0.296)	0.045(0.428)	0.952(1.236)	0.935(1.129)	0.952(1.202)	0.951(1.751)
	1.1	0(0.305)	0.001(0.281)	0(0.302)	0.063(0.434)	0.951(1.256)	0.932(1.148)	0.95(1.223)	0.948(1.779)
$\rho = 0.7$									
0	1	0.002(0.37)	0.002(0.348)	0.002(0.371)	0.002(0.37)	0.953(1.518)	0.939(1.422)	0.953(1.511)	0.953(1.518)
	1.02	0(0.374)	0(0.351)	0(0.375)	0(0.374)	0.951(1.533)	0.938(1.437)	0.951(1.527)	0.951(1.533)
	1.05	-0.003(0.381)	-0.003(0.358)	-0.003(0.381)	-0.003(0.381)	0.952(1.562)	0.937(1.464)	0.95(1.555)	0.952(1.562)
	1.07	0.003(0.383)	0.003(0.36)	0.003(0.384)	0.003(0.383)	0.948(1.571)	0.934(1.472)	0.95(1.565)	0.948(1.571)
	1.1	0(0.39)	0(0.366)	0(0.39)	0(0.39)	0.952(1.599)	0.938(1.499)	0.953(1.591)	0.952(1.599)
0.2	1	-0.002(0.379)	-0.001(0.356)	-0.001(0.377)	-0.002(0.475)	0.947(1.553)	0.934(1.459)	0.948(1.541)	0.952(1.945)
	1.02	0(0.383)	-0.001(0.361)	-0.001(0.382)	0.014(0.481)	0.949(1.573)	0.934(1.478)	0.95(1.561)	0.951(1.969)
	1.05	-0.001(0.388)	-0.002(0.366)	-0.002(0.387)	0.033(0.487)	0.949(1.592)	0.937(1.497)	0.95(1.581)	0.949(1.994)
	1.07	0.002(0.392)	0.002(0.369)	0.002(0.391)	0.048(0.492)	0.945(1.606)	0.93(1.51)	0.946(1.595)	0.947(2.014)
	1.1	-0.005(0.398)	-0.006(0.375)	-0.005(0.397)	0.056(0.499)	0.948(1.632)	0.933(1.534)	0.95(1.62)	0.949(2.044)

Table 3: Comparison of Bias, Std, Coverage and CI width for the test of equality of areas under the curve. Data was generated under the ARH(1) assumption with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$, $cv = 0.3$ and $\sigma^2 = 0.1$. Simulations were performed for 10,000 iterations.

p	γ	Bias (std)				Coverage (CI width)			
		UN	CS	ARH(1)	T-Test	UN	CS	ARH(1)	T-Test
$\rho = 0.3$									
0	1	0.001(0.799)	0.001(0.743)	0.001(0.801)	0.001(0.799)	0.953(3.275)	0.936(3.031)	0.953(3.239)	0.953(3.275)
	1.02	-0.003(0.808)	-0.003(0.75)	-0.003(0.81)	-0.003(0.808)	0.954(3.309)	0.933(3.063)	0.953(3.274)	0.954(3.309)
	1.05	0.012(0.821)	0.012(0.763)	0.012(0.824)	0.012(0.821)	0.952(3.365)	0.936(3.115)	0.953(3.33)	0.952(3.365)
	1.07	-0.006(0.829)	-0.006(0.771)	-0.006(0.832)	-0.006(0.829)	0.955(3.398)	0.939(3.147)	0.955(3.362)	0.955(3.398)
	1.1	-0.01(0.841)	-0.01(0.781)	-0.01(0.843)	-0.01(0.841)	0.952(3.444)	0.934(3.188)	0.952(3.408)	0.952(3.444)
0.2	1	-0.004(0.871)	-0.005(0.802)	-0.003(0.86)	-0.004(0.872)	0.953(3.588)	0.939(3.275)	0.954(3.486)	0.954(3.572)
	1.02	-0.014(0.879)	-0.014(0.809)	-0.012(0.867)	0.001(0.881)	0.949(3.622)	0.934(3.304)	0.953(3.515)	0.951(3.609)
	1.05	-0.012(0.89)	-0.013(0.819)	-0.014(0.879)	0.02(0.891)	0.951(3.665)	0.935(3.348)	0.951(3.563)	0.949(3.652)
	1.07	-0.01(0.9)	-0.007(0.829)	-0.008(0.888)	0.042(0.901)	0.945(3.706)	0.929(3.385)	0.948(3.602)	0.948(3.691)
	1.1	-0.01(0.913)	-0.012(0.841)	-0.011(0.902)	0.059(0.915)	0.948(3.767)	0.931(3.437)	0.949(3.658)	0.947(3.75)
$\rho = 0.7$									
0	1	0(1.11)	0(1.041)	0(1.11)	0(1.11)	0.954(4.546)	0.94(4.261)	0.954(4.525)	0.954(4.546)
	1.02	-0.012(1.122)	-0.012(1.052)	-0.012(1.121)	-0.012(1.122)	0.953(4.597)	0.939(4.307)	0.954(4.572)	0.953(4.597)
	1.05	0.009(1.138)	0.009(1.067)	0.009(1.139)	0.009(1.138)	0.953(4.661)	0.938(4.368)	0.953(4.642)	0.953(4.661)
	1.07	0.004(1.15)	0.004(1.079)	0.004(1.15)	0.004(1.15)	0.947(4.71)	0.932(4.415)	0.948(4.688)	0.947(4.71)
	1.1	0.013(1.167)	0.013(1.095)	0.013(1.167)	0.013(1.167)	0.955(4.779)	0.942(4.48)	0.954(4.756)	0.955(4.779)
0.2	1	-0.002(1.136)	-0.001(1.071)	0(1.133)	-0.006(1.119)	0.947(4.661)	0.934(4.382)	0.95(4.626)	0.948(4.586)
	1.02	-0.007(1.15)	-0.009(1.082)	-0.008(1.145)	0.004(1.133)	0.951(4.716)	0.936(4.429)	0.951(4.676)	0.951(4.643)
	1.05	-0.004(1.164)	-0.005(1.096)	-0.007(1.161)	0.026(1.147)	0.953(4.774)	0.94(4.486)	0.955(4.741)	0.952(4.698)
	1.07	-0.006(1.174)	-0.005(1.107)	-0.004(1.171)	0.046(1.156)	0.949(4.818)	0.932(4.529)	0.948(4.781)	0.945(4.736)
	1.1	-0.002(1.194)	-0.004(1.125)	-0.005(1.191)	0.059(1.177)	0.95(4.898)	0.935(4.604)	0.949(4.862)	0.951(4.821)

Table 4: Comparison of Bias, Std, Coverage and CI width for the test of equality of areas under the curve. Data was generated under the ARH(1) assumption with $\boldsymbol{\mu} = (2, 2.5, 3, 3, 2.5, 2)'$, $cv = 0.5$ and $\sigma^2 = 0.1$. Simulations were performed for 10,000 iterations.

p	γ	Bias (std)				Coverage (CI width)			
		UN	CS	ARH(1)	T-Test	UN	CS	ARH(1)	T-Test
$\rho = 0.3$									
0	1	0.012(1.285)	0.012(1.194)	0.012(1.289)	0.012(1.285)	0.951(5.265)	0.934(4.874)	0.951(5.209)	0.951(5.265)
	1.02	0.008(1.297)	0.008(1.204)	0.008(1.299)	0.008(1.297)	0.951(5.312)	0.933(4.917)	0.951(5.252)	0.951(5.312)
	1.05	-0.008(1.317)	-0.008(1.223)	-0.008(1.319)	-0.008(1.317)	0.954(5.394)	0.934(4.994)	0.954(5.332)	0.954(5.394)
	1.07	-0.021(1.331)	-0.021(1.236)	-0.021(1.334)	-0.021(1.331)	0.955(5.451)	0.936(5.047)	0.953(5.394)	0.955(5.451)
	1.1	-0.025(1.348)	-0.025(1.252)	-0.025(1.353)	-0.025(1.348)	0.95(5.522)	0.932(5.112)	0.95(5.471)	0.95(5.522)
0.2	1	0.008(1.394)	0.011(1.286)	0.007(1.38)	0.004(1.344)	0.95(5.746)	0.932(5.254)	0.952(5.595)	0.948(5.508)
	1.02	-0.018(1.412)	-0.006(1.3)	-0.008(1.395)	0.004(1.359)	0.949(5.815)	0.934(5.312)	0.95(5.656)	0.95(5.566)
	1.05	0.015(1.432)	0.018(1.321)	0.018(1.416)	0.043(1.38)	0.951(5.9)	0.932(5.395)	0.949(5.742)	0.947(5.653)
	1.07	0.021(1.443)	0.023(1.331)	0.023(1.428)	0.066(1.39)	0.947(5.942)	0.929(5.437)	0.946(5.789)	0.944(5.693)
	1.1	-0.033(1.47)	-0.033(1.355)	-0.034(1.454)	0.035(1.416)	0.946(6.056)	0.929(5.537)	0.947(5.895)	0.948(5.799)
$\rho = 0.7$									
0	1	-0.02(1.782)	-0.02(1.672)	-0.02(1.78)	-0.02(1.782)	0.945(7.301)	0.929(6.841)	0.946(7.257)	0.945(7.301)
	1.02	-0.02(1.8)	-0.02(1.688)	-0.02(1.798)	-0.02(1.8)	0.946(7.373)	0.932(6.908)	0.947(7.328)	0.946(7.373)
	1.05	-0.009(1.827)	-0.009(1.714)	-0.009(1.826)	-0.009(1.827)	0.948(7.486)	0.931(7.015)	0.947(7.444)	0.948(7.486)
	1.07	0.02(1.847)	0.02(1.733)	0.02(1.847)	0.02(1.847)	0.949(7.567)	0.934(7.091)	0.95(7.529)	0.949(7.567)
	1.1	-0.019(1.873)	-0.019(1.757)	-0.019(1.873)	-0.019(1.873)	0.946(7.672)	0.934(7.188)	0.948(7.634)	0.946(7.672)
0.2	1	-0.023(1.824)	-0.014(1.715)	-0.014(1.816)	-0.014(1.751)	0.951(7.484)	0.936(7.018)	0.95(7.411)	0.949(7.172)
	1.02	0.004(1.844)	0.009(1.736)	0.008(1.836)	0.023(1.772)	0.952(7.567)	0.936(7.103)	0.951(7.496)	0.951(7.258)
	1.05	0.001(1.869)	0.005(1.761)	0.003(1.862)	0.034(1.795)	0.95(7.667)	0.935(7.206)	0.948(7.6)	0.947(7.353)
	1.07	-0.04(1.89)	-0.034(1.779)	-0.035(1.881)	0.019(1.818)	0.95(7.752)	0.938(7.281)	0.95(7.68)	0.951(7.446)
	1.1	-0.011(1.919)	-0.011(1.805)	-0.012(1.909)	0.053(1.843)	0.948(7.871)	0.933(7.387)	0.95(7.794)	0.95(7.548)

Table 5: Vector of coefficients ($\mathbf{c}$) to test for equality of mean areas

c_1	c_2	c_3	c_4
1.5	6	10.5	6

Table 6: Summary of results from the Acupuncture Study

Method	*Estimate* (95% CI)	*SE*	*p − value*
T-Test	-164.50 (-305.2, -23.6803)	72.6143	0.0223
AR(1)	-156.21 (-285.76, -26.6676)	65.8468	0.0183
UN	-149.36 (-285.58, -13.1343)	69.1339	0.0318
CS	-150.17 (-275.22, -25.1166)	63.5423	0.0188

Table 7: Summary of results from the Acupuncture Study using only completers

Method	*Estimate* (95% CI)	*SE*	*p − value*
T-Test	-150.8(-296.7, -4.9712)	75.5975	0.0428
AR(1)	-150.8(-280.81, -20.8812)	65.9938	0.0231
UN	-150.8(-300.07, -1.6203)	75.5975	0.0476
CS	-150.8(-284.29, -17.4038)	67.6681	0.0269