Comparison of Power from Area under the Curve and Mixed Effects Models Methodologies for Profile Analysis

Robbie A Beyl[1], Jeff Burton[1], and William D Johnson[1]
[1]Pennington Biomedical Research Center, 6400 Perkins Rd Baton Rouge, LA 70808

Abstract

Assume that study subjects are randomly assigned to one of two treatments and assessed for a specific response at time 0 (baseline) and each of three post-treatment times. Further, assume the objective is to compare the treatments with respect to their response profiles across time. This setting is representative of an oral glucose tolerance test used in diabetes research. While the analytical method of choice is to employ mixed effects models for repeated measures, many biomedical researchers prefer comparing the treatments in terms of area under the curve (AUC). Thus, the goal is to determine if one of the analytical methods is more powerful than the other. Data are generated under the null hypothesis and for various configurations of means under the alternative hypothesis. These cases include changes to none, one, two, and all three of the post treatment means. The power curves estimated from these data are adjusted such that estimated power for both tests is equal to the nominal significance level under the null hypothesis. Once adjusted, we compare power between the two testing methodologies across the specified mean configurations.

Key Words: Area under the curve, AUC, mixed model, glucose, profile analysis

1. Introduction

Oral glucose tolerance test (OGTT) is a medical procedure designed to determine how a subject's body processes glucose. The OGTT involves taking 4 measurements of glucose levels though blood draws. The first measurement is taken before the subject ingests the glucose solution and is treated as the subject's baseline glucose level. Each of the sequential measurements is 60 minutes apart for the last. Denote the i^{th} measurement of glucose as g_i for $i=1,2,3,4$. An example two mean glucose profiles with 25 subjects in each group is shown in Figure 1.

The primary goal of the studies that inspired this research is to determine changes in glucose profiles for treatment groups. Several different analyses have been applied to OGTT data to determine if the profiles from each treatment are different (Allison). Area under the curves (AUC) is a metric used in this type of research for more than 75 years (Ross). There appears to be no literature about using a mixed effect linear model to OGTT data. Table 1 shows the p-values based on testing if the two profiles from Figure 1 are equal. Despite testing similar hypotheses, the conclusions based on the p-values would be quite different.

This article details the calculations that go into AUC and the mixed model, and give properties of these two methodologies under different settings.

2. Methods

The two methods for determining differences in glucose profiles considered here are the area under the curve approach and a linear mixed model approach. Since the interest is in how the profiles change and not absolute value of each of the glucose measurements, the baseline value is subtracted off from each of the measurements. The new response is denotes as $y_i = g_i - g_1$.

2.1 Area under the curve

Several different methods can be used to calculate area under the curve with the most common being the trapezoid rule:

$$\frac{1}{2}\sum_{i=1}^{t-1}(x_{i+1} - x_i)(y_{i+1} + y_i),$$

where *x* denotes the time and *t* is the number of time points. This formula greatly simplifies when applied to OGTT data. With $t = 4$, $y_1 = 0$, and $(x_{i+1} - x_i)$ being equal for all *i*'s, the AUC for each subject is proportional to

$$y_2 + y_3 + \frac{1}{2}y_4.$$

Thus, the AUC is a weight mean with the last measure receiving half the weight of the second and third measurements. A two sample t-test is the common statistic used with AUC. The null hypothesis is that the mean AUC for the treatment is equal to the mean AUC for the control group. This statistic is compared to a t-distribution with $N_1 - N_2 - 2$, degrees of freedom with N_k being the total number of subjects from the k^{th} group.

2.2 Mixed effect linear model

The linear mixed effect model is created to model the means over time. The null hypothesis analogous to testing equality between profiles is testing equality of the treatment means at each time point simultaneously. This requires the model to be setup in a non-standard way. Each of the time points will have an overall mean. This effect can be created by treating the time variable as categorical. Next, the model requires the treatment effect to vary at each time point. This effect is created using an interaction of treatment*time. Thus, the treatment*time effect is testing if the means from the treatment group are equal to their corresponding mean from the control group. The code given below shows one way to create this model in SAS. The measurement at the first time point is removed since all values are 0, and an unstructured covariance matrix is used to allow the data to determine how each of three repeated measures are correlated. The treatment*time effect statistics is compared to F-distribution with 3 numerator degrees of freedom and $N_1 - N_2 - 2$ denominator degrees of freedom.

```
proc mixed; where time^=0;
class treatment time;
model diff= time treatment*time;
repeated /subject=subject type=un;
run;
```

2.3 Simulations

The power of the AUC and mixed model are investigated to determine which of these methodologies performs bests. The simulations used to generate power are based on OGTT data taken previous studies performed at Pennington Biomedical Research Center. Data from several control groups was used to estimate means, variances, and co-variances for each time point.

The overall null hypothesis is that the two profiles are equal. Power calculations are based on data that departs from the null hypothesis. While there are many ways to make profiles different, some may not be fair when comparing the two methodologies. For example, crossing profiles can have similar AUC but different means. This is not a fair comparison since the null hypothesis for AUC is still occurring. Therefore, the data is generated using profiles that are equal: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, where $\boldsymbol{\mu}_k$ is mean vector of the k^{th} group. Three cases where the profiles are no longer equal are investigated:

I. $\mu_{12} > \mu_{22}$
II. $\mu_{12} > \mu_{22}$ & $\mu_{13} > \mu_{23}$
III. $\mu_{12} > \mu_{22}$ & $\mu_{13} > \mu_{23}$ & $\mu_{14} > \mu_{24}$,

where μ_{ki} is the mean of the k^{th} group at the i^{th} time.

Define the delta as the sum of the differences between the two mean vectors.
At each value of delta, 1,000 simulations are computed. For each simulation, data for 25 subjects are generated for each treatment group. The result of each simulation is a p-value from the type III test of fixed effects for the treatment*time variable. Power is defined as the number of simulation with a p-value less than or equal to 0.05 over 1,000.

3. Results

The type I error rate can be seen when delta equals 0. Across all three cases, both methods have estimated type I error rate close to the nominal value of 0.05. Since both methods have similar type I error rate which are quite close to 0.05, the power curves are not adjusted.

Figure 2 shows the power curves for case I. The mixed effect model is uniformly more powerful than AUC. Once the delta reaches 5, mixed model gain power faster than AUC.

The power curves for case II in shown in Figure 3. As with case I, the mixed model has uniformly greater power than AUC. However, the rate at which the power increases for AUC appears to be only slightly less than the rate for the mixed model.

The last figure shows the power for case III. Neither method has uniformly greater power; both methods produce similar power for delta less than 5. After a delta of around 5, AUC is more powerful than the mixed model. The rate at which the power increases for the mixed model is only slightly less than the rate for AUC.

4. Conclusion

There appears to be no clear best test to determine differences of profiles based on data from an oral glucose tolerance test. The mixed effect linear model appears to perform the best in cases I and II, while the area under the curve method appears to perform better in case III.

The two methodologies are based on different statistics, so it not surprising to see that they preform differently. However, both methods appear to be testing similar hypotheses which suggest that can be used to test equality of profiles. The difference in power, in part, can be contributed to the number of different means. The power of the mixed model does not appear to be greatly affected by the number of means that are different, whereas this number does affect AUC. This result suggests that delta, sum of the different of mean vectors, may not be the best metric to judge power.

Only three cases were investigated for OGTT data. Future research will expand the number of case to include different combination of means with both increases and decreases.

Acknowledgements

Supported in part by 1 U54 GM104940 from the National Institute of General Medical Sciences of the National Institutes of Health, which funds the Louisiana Clinical and Translational Science Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Ross CW and Tonks EL. The determination of glucose tolerance. Arch Dis Child. 1938: 13(76). 289-309.

Allison DB, Paultre F, Maggio C, Mezzitis N, and Pi-Sunyer FX. The Use of Areas Under Curves in Diabetes Research. Diabetes Care. 1995 18:2 245-250.

Table 1: P-values testing equality of the two glucose profiles

	AUC	Mixed
P-value	0.2830	0.0390

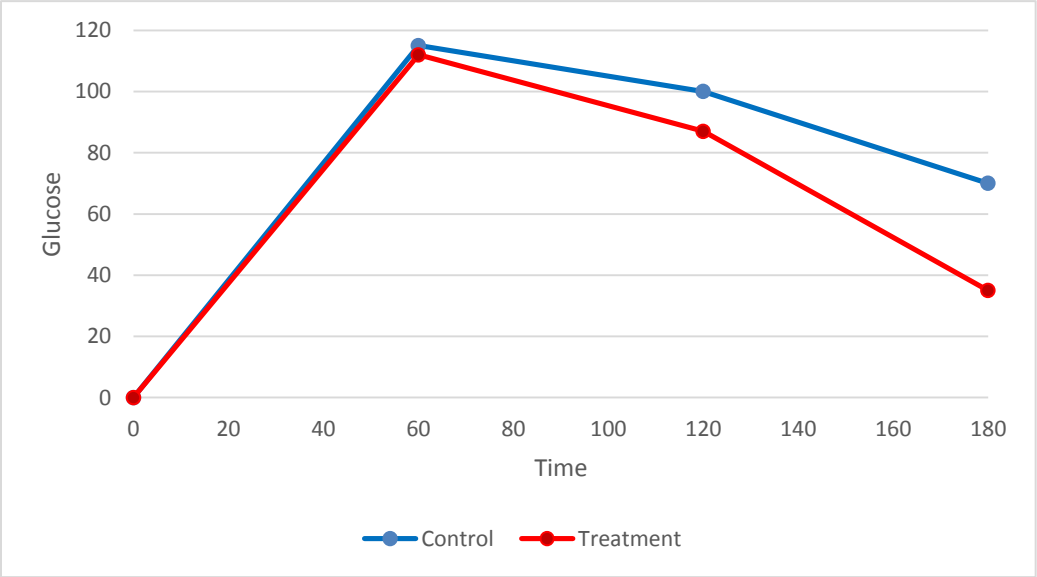

Figure 1: Example of two mean glucose profiles

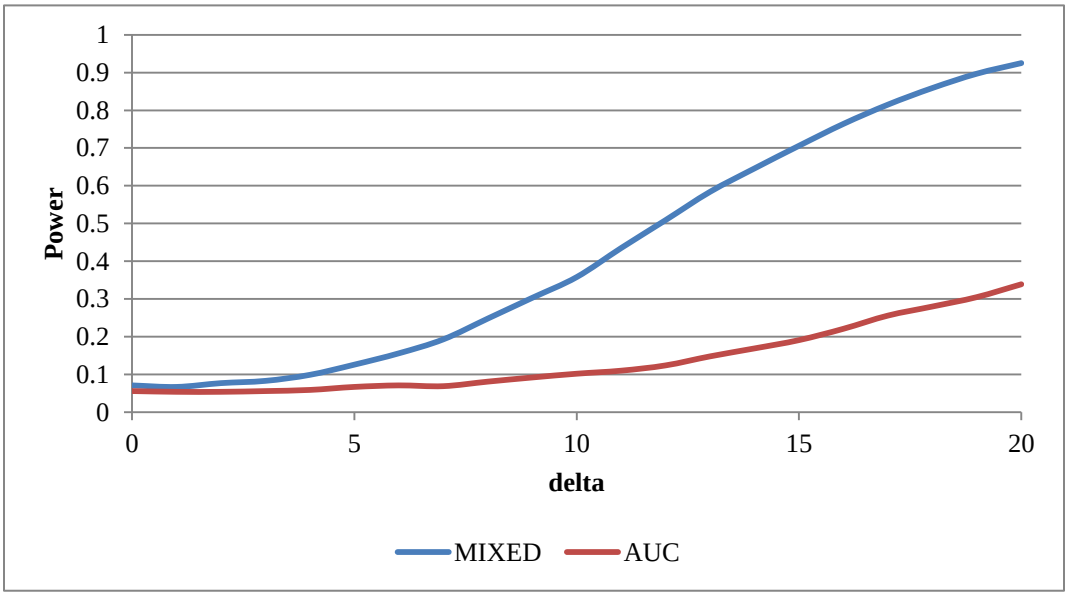

Figure 2: Power curves for case I

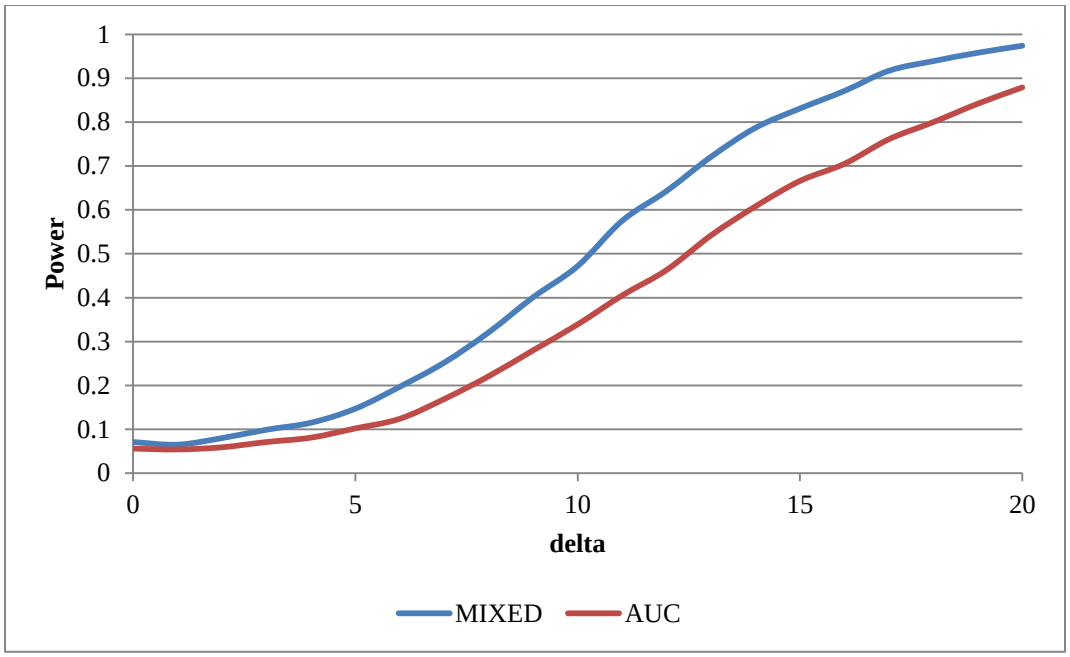

Figure 3: Power curves for case II

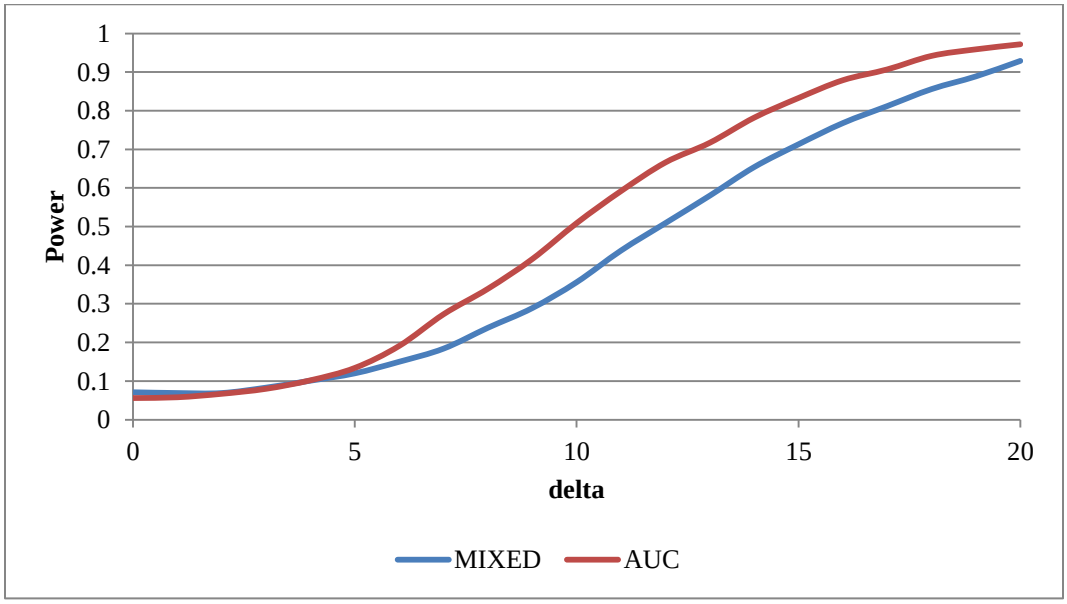

Figure 4: Power curves for case III