MANOVA modelling of a chiropractic longitudinal study using multiple imputation

R. S. Bowden¹, B. Walker² and B. R. Clarke³

Technical Report
December 2012

¹ Mathematics and Statistics, School of Engineering and Information Technology, Murdoch University
² Chiropractic Program, School of Health Professions, Murdoch University
³ Mathematics and Statistics, School of Engineering and Information Technology, Murdoch University
1. Introduction

The purpose of this report is to present the detailed statistical analysis of a randomised, placebo-controlled trial comparing two different treatment modalities to an intervention of no known benefit for people with acute or subacute thoracic spine pain.

The therapy arms consist of Spinal Manipulative Therapy (SMT) and Graston Technique (GT) and the placebo is a non-functional ultrasound. A placebo group was utilised because at present there are no proven treatments for non-specific thoracic pain. This trial is registered with the Australia and New Zealand Clinical Trials Registry. Ethics approval has been granted by Murdoch University Human Research and Ethics Committee, number 2007/274.

The aim of this three arm trial was to test the efficacy of SMT and GT as independent modalities compared to detuned ultrasound for the outcomes of pain and disability. The latter were measured using the Visual Analogue Scale (VAS) and a modified Oswestry Back Pain Disability Index. The study was conducted at the Murdoch University Chiropractic student clinic in Perth, Australia, and the protocol published in Crothers et al (2008).

In this report, Section 2 provides an initial exploratory analysis of the data, Section 3 outlines the statistical models used in the final analysis, Section 4 defines these models in mathematical terms, Section 5 discusses the management of missing values via multiple imputation and Section 6 presents the results of the statistical modelling and hypothesis tests. The clinical study will be published in full elsewhere.

2. Initial Data Analysis

One hundred and forty three participants were recruited for the trial. They were randomly assigned to three treatments being:

- Spinal Manipulative Therapy
- The Graston Technique and
- A placebo employing an un-powered ultra-sound (US) machine

The pain response of participants was measured just before the first treatment and then at intervals of one week, four weeks, three months, six months and twelve months from the two-week treatment. As mentioned previously, the measures were the Visual Analogue Scale where the participant indicates the score level on a continuous physical scale and the Oswestry Disability Index which uses a series of questions. The data was analysed by a statistician blind to the treatment allocation.

The process of selection was:

- Advertise for participants,
- Screen respondents over the phone using a basic selection criteria,
- Further screen respondents for suitability at a face-to-face interview,
- Randomly assign remaining participants to one of the three treatment groups.
One hundred and forty three participants were recruited for the trial. Table 1 provides baseline data for the three groups. There were no important differences noted. However there was an apparent disparity in the number of participants randomised to the three treatment groups (36, 63 and 44 seen in Table 1). The differences in counts were statistically significant (p=0.02) but this p-value may be biased downwards as testing was only undertaken due to the disparate cell counts.

There were two known issues with randomisation:

• One of the persons supervising treatments was found to have returned the allocation envelope to the allocation card box prior to it being opened and chose an alternate envelope, which assigned a participant to a treatment preferred by this clinician. They were able to determine the allocation by holding the “opaque” envelope up to a light. The participant at the time was excluded from the trial results and analysis and the clinician was counselled. As a result all remaining envelopes were opened and the allocation cards were wrapped in aluminium foil. They were then returned to new sealed envelopes. The allocation sequence was maintained.

• On other occasions a volunteer assistant responsible for some allocations was found to be randomly allocating participants in correct sequence but doing so before all instruments of measurement were completed, some of which resulted in exclusion. This was the likely reason for the variation in the group participant numbers.

It seems unlikely that these issues would significantly biased the allocation. And hence it is probable that the disparity in numbers was a random outcome. Further analysis was not undertaken on possible in-balances in the sample following Senn (1994) who argues that “this practice is philosophically unsound, of no practical value and potentially misleading”.

### Table 1. Baseline data for the entire sample and the three treatment groups.

<table>
<thead>
<tr>
<th>Variable</th>
<th>All (N = 143)</th>
<th>SMT (N = 36)</th>
<th>Graston (N = 63)</th>
<th>US (placebo)(N = 44)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Age (years)(SD)</td>
<td>45.8 (13.7)</td>
<td>44.4 (13.0)</td>
<td>44.8 (14.3)</td>
<td>48.5 (13.0)</td>
</tr>
<tr>
<td>Range</td>
<td>18-74</td>
<td>18-74</td>
<td>19-71</td>
<td>21-67</td>
</tr>
<tr>
<td>Sex % Male</td>
<td>53.2</td>
<td>55.6</td>
<td>50.2</td>
<td>54.6</td>
</tr>
<tr>
<td>VAS Mean (SD)</td>
<td>5.6 (2.0)</td>
<td>5.5 (2.0)</td>
<td>5.7 (2.1)</td>
<td>5.5 (2.0)</td>
</tr>
<tr>
<td>ODI Mean (SD)</td>
<td>28.5 (10.4)</td>
<td>27.2 (10.2)</td>
<td>29.6 (11.1)</td>
<td>28.1 (9.9)</td>
</tr>
<tr>
<td>Mean length of pain time in years (SD)</td>
<td>9.2 (12.0)</td>
<td>9.0 (16.0)</td>
<td>8.2 (11.0)</td>
<td>10.9 (9.4)</td>
</tr>
</tbody>
</table>

Of total initial participants in the current study, 92% completed week 1 and 62% month 12 (There were no drop-outs for the baseline values). This is a reflection of the standard “dropout” missing value pattern in longitudinal studies. There was no statistically significant greater incidence of dropout for any treatment group at any stage of the study. However older participants were more likely to persist in the study (p=0.0004). Figure 1 shows the retention and dropout of participants at each stage of the study. Missing values are a common problem in longitudinal studies (Acock 2005, Dragset 2009, Myers 2000, Laird 1988 and Nakai and Ke 2011) and the treatment of missing values is addressed in Section 5.
Figure 1. Flow chart of participants (VAS Scores)

Responded to advertisement N = 361
Assessed for eligibility via screening questionnaire

Potentially eligible attend for clinic visit: N = 203 History/Physical exam/Informed consent

Baseline Visit: N = 143
VAS, Oswestry, SF-36 and socioeconomic data
Random group assignment N=144, minus 1 due to randomisation issue.

Excluded: N = 59
Pain referral from neck, thoracic fracture, DISH & ODI too low.

Group SMT : N = 36
Week 1, N=34
Lost to follow up, N=2
Week 4, N=33
Lost to follow up, N=3
3-months, N=32
Lost to follow up, N=4
6-months, N=25
Lost to follow up, N=11
12-months, N=26
Lost to follow up, N=10

Group Graston : N = 63
Week 1, N=58
Lost to follow up, N=5
Week 4, N=54
Lost to follow up, N=9
3-months, N=50
Lost to follow up, N=13
6-months, N=50
Lost to follow up, N=13
12-months, N=36
Lost to follow up, N=27

Group Placebo : N = 44
Week 1, N=40
Lost to follow up, N=4
Week 4, N=37
Lost to follow up, N=7
3-months, N=31
Lost to follow up, N=13
6-months, N=31
Lost to follow up, N=13
12 months, N=26
Lost to follow up, N=18
The mean pain scores at each time point by treatment group are presented in Table 2 along with standard deviations (unbiased) and 95% confidence intervals.

Table 2. VAS, ODI scores and number of treatments for the three treatment groups

<table>
<thead>
<tr>
<th></th>
<th>All (n=143)</th>
<th>Group SMT (n=36)</th>
<th>Group Graston (n=63)</th>
<th>Group Placebo (n=44)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAS Baseline</td>
<td>Mean 5.55</td>
<td>5.46</td>
<td>5.66</td>
<td>5.48</td>
</tr>
<tr>
<td></td>
<td>SD 2.01</td>
<td>2.00</td>
<td>2.11</td>
<td>1.95</td>
</tr>
<tr>
<td></td>
<td>CI Lower 95%</td>
<td>4.80</td>
<td>5.14</td>
<td>4.90</td>
</tr>
<tr>
<td></td>
<td>CI Upper 95%</td>
<td>6.11</td>
<td>6.18</td>
<td>6.06</td>
</tr>
<tr>
<td>VAS 1 Week</td>
<td>Mean 4.81</td>
<td>5.08</td>
<td>4.73</td>
<td>4.70</td>
</tr>
<tr>
<td></td>
<td>SD 1.90</td>
<td>2.07</td>
<td>1.90</td>
<td>1.83</td>
</tr>
<tr>
<td></td>
<td>CI Lower 95%</td>
<td>4.38</td>
<td>4.24</td>
<td>4.13</td>
</tr>
<tr>
<td></td>
<td>CI Upper 95%</td>
<td>5.77</td>
<td>5.22</td>
<td>5.27</td>
</tr>
<tr>
<td>VAS 4 Weeks</td>
<td>Mean 3.91</td>
<td>4.33</td>
<td>3.44</td>
<td>4.24</td>
</tr>
<tr>
<td></td>
<td>SD 2.14</td>
<td>2.07</td>
<td>1.95</td>
<td>2.44</td>
</tr>
<tr>
<td></td>
<td>CI Lower 95%</td>
<td>3.63</td>
<td>2.92</td>
<td>3.45</td>
</tr>
<tr>
<td></td>
<td>CI Upper 95%</td>
<td>5.04</td>
<td>3.96</td>
<td>5.02</td>
</tr>
<tr>
<td>VAS 12 Weeks</td>
<td>Mean 3.51</td>
<td>4.00</td>
<td>3.20</td>
<td>3.50</td>
</tr>
<tr>
<td></td>
<td>SD 2.44</td>
<td>2.19</td>
<td>2.66</td>
<td>2.37</td>
</tr>
<tr>
<td></td>
<td>CI Lower 95%</td>
<td>3.24</td>
<td>2.46</td>
<td>2.66</td>
</tr>
<tr>
<td></td>
<td>CI Upper 95%</td>
<td>4.75</td>
<td>3.94</td>
<td>4.33</td>
</tr>
<tr>
<td>VAS 26 Weeks</td>
<td>Mean 3.55</td>
<td>3.57</td>
<td>3.51</td>
<td>3.61</td>
</tr>
<tr>
<td></td>
<td>SD 2.43</td>
<td>2.22</td>
<td>2.51</td>
<td>2.59</td>
</tr>
<tr>
<td></td>
<td>CI Lower 95%</td>
<td>2.70</td>
<td>2.81</td>
<td>2.69</td>
</tr>
<tr>
<td></td>
<td>CI Upper 95%</td>
<td>4.44</td>
<td>4.21</td>
<td>4.52</td>
</tr>
<tr>
<td>VAS 52 Weeks</td>
<td>Mean 3.40</td>
<td>3.84</td>
<td>3.15</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>SD 2.41</td>
<td>2.47</td>
<td>2.37</td>
<td>2.53</td>
</tr>
<tr>
<td></td>
<td>CI Lower 95%</td>
<td>2.89</td>
<td>2.38</td>
<td>2.32</td>
</tr>
<tr>
<td></td>
<td>CI Upper 95%</td>
<td>4.79</td>
<td>3.93</td>
<td>4.27</td>
</tr>
<tr>
<td>ODI Baseline</td>
<td>Mean 28.55</td>
<td>27.19</td>
<td>29.6</td>
<td>28.14</td>
</tr>
<tr>
<td></td>
<td>SD 10.42</td>
<td>10.20</td>
<td>11.09</td>
<td>9.90</td>
</tr>
<tr>
<td></td>
<td>CI Lower 95%</td>
<td>23.86</td>
<td>26.87</td>
<td>25.21</td>
</tr>
<tr>
<td></td>
<td>CI Upper 95%</td>
<td>30.53</td>
<td>32.34</td>
<td>31.06</td>
</tr>
<tr>
<td>ODI 1 Week</td>
<td>Mean 22.87</td>
<td>22.32</td>
<td>22.61</td>
<td>23.68</td>
</tr>
<tr>
<td></td>
<td>SD 11.42</td>
<td>10.14</td>
<td>11.86</td>
<td>12.27</td>
</tr>
<tr>
<td></td>
<td>CI Lower 95%</td>
<td>18.91</td>
<td>19.58</td>
<td>19.93</td>
</tr>
<tr>
<td></td>
<td>CI Upper 95%</td>
<td>25.73</td>
<td>25.64</td>
<td>27.44</td>
</tr>
<tr>
<td>ODI 4 Weeks</td>
<td>Mean 19.54</td>
<td>19.79</td>
<td>18.05</td>
<td>21.46</td>
</tr>
<tr>
<td></td>
<td>SD 12.04</td>
<td>11.78</td>
<td>12.15</td>
<td>12.45</td>
</tr>
<tr>
<td></td>
<td>CI Lower 95%</td>
<td>15.83</td>
<td>14.87</td>
<td>17.55</td>
</tr>
<tr>
<td></td>
<td>CI Upper 95%</td>
<td>23.76</td>
<td>21.24</td>
<td>25.37</td>
</tr>
<tr>
<td>ODI 12 Weeks</td>
<td>Mean 18.27</td>
<td>21.00</td>
<td>16.24</td>
<td>18.71</td>
</tr>
<tr>
<td></td>
<td>SD 14.07</td>
<td>14.57</td>
<td>13.20</td>
<td>15.29</td>
</tr>
<tr>
<td></td>
<td>CI Lower 95%</td>
<td>15.95</td>
<td>12.58</td>
<td>13.33</td>
</tr>
<tr>
<td></td>
<td>CI Upper 95%</td>
<td>26.05</td>
<td>19.90</td>
<td>24.09</td>
</tr>
<tr>
<td>ODI 26 Weeks</td>
<td>Mean 16.88</td>
<td>18.24</td>
<td>16.15</td>
<td>16.94</td>
</tr>
<tr>
<td></td>
<td>SD 13.61</td>
<td>14.44</td>
<td>13.24</td>
<td>14.33</td>
</tr>
<tr>
<td></td>
<td>CI Lower 95%</td>
<td>12.58</td>
<td>12.44</td>
<td>11.89</td>
</tr>
<tr>
<td></td>
<td>CI Upper 95%</td>
<td>23.90</td>
<td>19.85</td>
<td>21.98</td>
</tr>
<tr>
<td>ODI 52 Weeks</td>
<td>Mean 17.68</td>
<td>21.23</td>
<td>16.28</td>
<td>16.08</td>
</tr>
<tr>
<td></td>
<td>SD 15.18</td>
<td>16.30</td>
<td>13.67</td>
<td>16.58</td>
</tr>
<tr>
<td></td>
<td>CI Lower 95%</td>
<td>14.97</td>
<td>11.81</td>
<td>9.70</td>
</tr>
<tr>
<td></td>
<td>CI Upper 95%</td>
<td>27.50</td>
<td>20.74</td>
<td>22.45</td>
</tr>
</tbody>
</table>
The distribution of VAS and ODI scores conditional on the time point was relatively symmetric with some skewness to the right. No transformation was deemed necessary to satisfy the assumptions for statistical testing.

Figure 2 shows the mean VAS and ODI pain scores for each time point (without imputation of missing values). Note that the x-axis is not linear in the number of weeks. The confidence bounds (CI’s) have been so constructed that a statistically significant difference between the sample means (two-sided 5% level) will be indicated by the bounds not crossing. This is an approximate test only but provides a visual assessment tool.

From the plots there appears to be a significant reduction in pain scores across all three treatments whilst noting that one of these is a placebo. However there doesn’t appear to be a difference between groups. The preliminary conclusion here is that mean pain scores tend to reduce over time equally for each treatment including a placebo.

Figure 2. Mean VAS and ODI Pain Scores by Treatment Group and Time Point. The confidence intervals are designed so that they do not overlap if the difference is (approximately) significant at 5% level. The group means are for SMT (full line), Graston (large dashed line) and placebo (small dashed line) treatments.
3. Statistical Methods

There are a range of statistical models that can be used to analyse longitudinal data (see Dunn and Pickles 2005, Nakai and Ke 2009 and Everitt 1998) and these include:

- ANOVA (Clarke 2008),
- Multivariate version of ANOVA, i.e. MANOVA (Morrison 2005),
- Seeming unrelated regression (SUR) (Pindyck and Rubinfeld 1997),
- Random effects models (Clarke 2008 - see comments below),
- Time series models (Dunn and Pickles 2005 and Chatfield 2004),
- Functional trend models, and
- Generalised linear models (Dobson 1990) which allow more general specification of the error distribution in regression models.

An initial examination of the pain score data indicated that the readings from time point to time point and between pain measures are correlated for each participant. Furthermore, because only one reading was recorded for each participant for each pain measure and time point, random effects models cannot be used (see further comments below).

The current study will use ANOVA and MANOVA models whilst noting the shortcomings of these specifications. ANOVA models concentrate on univariate aspects of the data and do not encompass the multivariate relationships. However they do allow a more fulsome use of the data (i.e. with minimal missing values).

MANOVA is a truly multivariate technique which accommodates the correlation between pain scores. In the current analysis it was decided to use MANOVA to model the differences between the baseline VAS and ODI scores and the scores at each time point. This allowed the modelling to automatically control for differences in the initial pain scores.

MANOVA runs the risk of over-parameterisation whereby, for the three treatments and ten variables (i.e. differences), eight-five (being 30 means, 10 pooled variances and 45 pooled covariances) parameters must be estimated from 143 data points. In the present circumstance this will likely result in reduced power given the indicated sample size and the moderate drop-out rate. However MANOVA is a standard for longitudinal studies which has the least modelling assumptions. No covariates were included in the modelling.

4. Mathematical Model

To more clearly define the actual models used, let $x_{i,j,k,m}$ be the pain score for the $i^{th}$ measure ($i=1,2$ i.e. VAS v ODI) at the $j^{th}$ time interval ($j=0,...,5$) on the $k^{th}$ treatment ($k=1,2,3$) for the $m^{th}$ participant ($m=1,...,n_k$). It is assumed that a participant’s VAS or ODI ($i=1,2$) pain score at a given time point is related to their initial score, $x_{i,0,k,m}$, plus a mean effect for the time period ($\delta_{i,j}$), a mean effect for the treatment ($\lambda_{i,k}$) and an interaction effect ($\eta_{i,j,k}$). Hence, for the $i^{th}$ measure, the $j^{th}$ time point and the $k^{th}$ treatment, the response from the $m^{th}$ participant is,

$$
x_{i,j,k,m} = x_{i,0,k,m} + \delta_{i,j} + \lambda_{i,k} + \eta_{i,j,k} + \omega_{i,j,k,m}, \quad m = 1,...,n_k.
$$

(1)
All (2×5) error terms for participant, \( m \), on treatment, \( k \), \( \{ \{ \omega_{i,j,k,m} \}^2 \}_{j=1}^5 \), are potentially cross-correlated. In other words, for the \( i \)th participant on the \( k \)th treatment, the error terms for the pain scores at all time points are cross-correlated. Technically the vector of error terms, \( \Omega'_{k,m} = (\{ \{ \omega_{i,j,k,m} \}^2 \}_{j=1}^5 \) \), is assumed to have a (common) non-diagonal variance matrix.

The above model implies that the pain score at each time point for each treatment is the pain score for the participant at the base value plus a unique constant for each time point plus an error term. This is an additive model, although a multiplicative model could be employed by using the logarithms of the pain scores in (1). Note that either model references the pain score to the initial value for the participant.

Let us transform the above equation (1) by finding the difference of each pain score from its value at the initial time point,

\[
\Delta x_{i,j,k,m} = x_{i,j,k,m} - x_{i,0,k,m} = \delta_{i,j} + \lambda_{i,k} + \eta_{i,j,k} + \omega_{i,j,k,m}.
\]  

(2)

The column vector of these scores is defined as \( \Delta'_{k,m} = (\{ \{ \Delta x_{i,j,k,m} \}^2 \}_{j=1}^5 \) \). Using MANOVA modelling, this formulation can be tested for significant differences between treatments \((k=1,2,3)\) in the mean improvements in pain scores,

\[
\{ \{ \pi_{i,j,k} \}^2 \}_{j=1}^5, \text{ where } \pi_{i,j,k} = \delta_{i,j} + \lambda_{i,k} + \eta_{i,j,k}.
\]

Give the above model, data analysis was undertaken as follows:

1. A univariate ANOVA was undertaken on the VAS improvements between the baseline and 4 week scores, i.e.,

\[
\Delta x_{1,2,k,m}; k = 1,2,3, m = 1,\ldots, n_k,
\]


to examine differences between the short-term changes in pain levels using the inherently continuous variable (VAS) while employing the maximum amount of non-missing data at a clinically effective time interval (4 weeks); and

2. A MANOVA was carried out on the vector of dependent data using the differenced values in (2) and a range of statistical tests (see below).

Before presenting the test results, the imputation of missing values will be discussed in the following section.

5. Missing Values and Multiple Imputation

There are a number of methods that can be used to accommodate missing values but in this study multiple imputation (MI) will be employed (Scafer 1997). MI uses the relationships within the non-missing data and the associated variability to randomly generate a sequence of substitutes for the missing data. New complete datasets are created which have in-filled
missing values but with a degree of variation reflecting the estimation process. Each of the new datasets are then analysed using standard models and the results combined using specific algorithms.

MI is superior to single imputation methods because the range of results from the multiple datasets reflect the variability inherent from in-filling the missing values. Single imputation methods are typically blind to this uncertainty. The MI used here employed the “mice” package in R and utilised linear least-squares regression modelling for imputation. The MI model was restricted to the ten “differenced” pain readings on each individual and their age category.

The parameter estimates and their standard errors are combined using results such as those developed by Rubin (1987). However it was not immediately clear how this could be applied to p-values from MI datasets resulting from (M)ANOVA F-tests.

Licht (2010) discusses combining MI p-values and presents a number of alternative methods, both for univariate and multivariate testing. Possibly the simplest of these assumes the p-values are uniformly distributed under the null hypothesis and are transformed using the inverse standard normal distribution. This produces variates with an assumed standard normal distribution and a mean of zero (if the null hypothesis holds). This is one of the methods used in the current analysis. However it likely that the MI p-values are positively correlated given the common nature of the observed (non-missing) data and do not have a uniform distribution (see Figure 3).

Li et al (1991a) (see also Scafer 1997, Harell and Zhou 2007 and Li et al 1991b), in a similar way to Brand (1999), discuss combining p-values by merging of the Wald statistics associated with the p-values. Each MI p-value from each MANOVA test statistic is assumed to be derived from Wald statistics which are chi-squared distributed with, in this case, 10 degrees of freedom (the dimension of the MANOVA). If only p-values are available, the Wald statistics can be calculated by using the p-values and the inverse chi-squared distribution to (reverse) look-up the associated quantiles of the chi-squared distribution with 10 degrees of freedom. These quantiles are then combined using Li’s algorithm which provides a statistic which is distributed as an F-distribution with degrees of freedom equal to the dimension of the MANOVA model (i.e. 10) and a calculated value.

Li’s method is calibrated for 3 imputation datasets and hence in the current study the Li p-values are calculated for the first three multiple imputations and the full 300. Furthermore Li et al (1991a) recommends that users assume that the true p-value is between a half to double the calculated MI p-value.

An upper bound on the true p-value can be derived by noting that the true p-value will be greater than the mean p-value from the multiple imputations because the latter does not reflect the variability in the MI results. With any variability in the p-values, the true p-value will be a more conservative, i.e. lower, quantity than the mean imputed p-value which doesn’t increase with increased variability in the imputations. Accordingly the mean imputed p-value can be used as an approximate lower bound on the true p-value.
6. Effectiveness of Treatments

Given the dropout rate for all treatment groups after 4 weeks the analysis will include both the differences from baseline for the total dataset (using MANOVA) and the difference to week 4 only (using ANOVA on VAS scores). This leaves 19 out of the total of 143 with missing values for the VAS score at week 4 versus 55 out of 143 with at least one missing value in the dependent variable vectors of differences for the whole dataset.

6.1 ANOVA

Table 3 sets out the results of an MI ANOVA analysis (second column) on the improvement from the baseline to the week four in the imputed VAS scores. Figure 3 shows histograms of the p-values for the ANOVA (and MANOVA (Pillai)) results from 300 imputed datasets and reveals that all of the ANOVA p-values are above 0.05.

<table>
<thead>
<tr>
<th>P-Value Estimate</th>
<th>ANOVA</th>
<th>MANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pillai</td>
</tr>
<tr>
<td>Mean p-value⁴</td>
<td>0.519</td>
<td>0.224</td>
</tr>
<tr>
<td></td>
<td>(0.007)⁵</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Licht (2010)</td>
<td>0.482</td>
<td>0.749</td>
</tr>
<tr>
<td>Li et al (1991a), m=300</td>
<td>0.613</td>
<td>0.376</td>
</tr>
<tr>
<td>Li et al (1991a), m=3</td>
<td>0.776</td>
<td>0.172</td>
</tr>
</tbody>
</table>

It is clear that none of the alternative methods of calculating the combined imputed p-values for the ANOVA test are significant at the 5% level. This is the case even if the Li et al (1991a) results are halved or the mean p-values are used as lower bounds on the true p-values.

⁴ These are the means of all 300 MI p-values, only the first three of which are used in calculating the Li et al (1991a) (m=3) combined MI p-values.

⁵ Standard Error.
6.2 MANOVA

A MANOVA analysis was undertaken using the imputed dataset with a dependent vector composed of all ten improvements (i.e. differences) in pain scores for the five post-treatment time periods and two pain scales. From Figure 3 the vast majority (96%) of the p-values are above 0.05.

The combined p-value results are shown in Table 5 for:

- Pillai’s Trace
- Wilks’ Lambda
- Hotelling-Lawley’s Trace, and
- Roy’s Largest Root.

The MI analysis suggests that there is no significant difference between the treatments at the 5% level for the improvement in pain scores, again, even if the Li et al (1991a) results are halved or the mean p-values are used as a lower bound.
Bibliography:


