The Best Treatment Modality Selection for Multiple Binary Endpoints

Zhibao Mi[1], Dimitri Novitzky[2] Stephen Bingham[1] and Joseph Collins[1]

[1] Cooperative Studies Program Coordinating Center, VA Maryland Health Care System, Perry Point, MD 21902
[2]Haley VA Medical Center, Tampa, FL 33606

Abstract

The management of brain dead organ donors (BDOD) is complex. The use of inotropic agents and replacement of depleted hormones known as hormonal replacement (HR) is crucial for multiple organ procurements. Yet the optimal HR modality has not been identified, and the statistical adjustment for the best selection is not trivial. Traditional pair-wise comparisons between every pair of treatments, multiple comparisons to all (MCA), are statistically conservative. Hsu's multiple comparison with the best (MCB)—adapted from the Dunnett's multiple comparisons with control (MCC)—has been used for selecting the best treatment with a single continuous outcome. We selected the best HR modality for the success of multiple organ procurements using a two step approach. First, we estimated the predicted margins of the logits by constructing generalized linear models (GLM) or generalized linear mixed models (GLMM), and then we applied the multiple comparison methods to identify the best HR modality given that the testing HR modalities are independent. Among sixteen HR modalities, we found that the combination of thyroid hormone, steroid, vasopressin, and insulin was the best HR modality for multiple organ procurement for transplantation based on the 95% simultaneous confidence intervals or decision limits.

Key Words: Best Treatment Selection, Multiple Binary Endpoints, Hormonal Replacement, Organ Procurement, BDOD

1. Introduction

The quality of organs and lack of supply remain big hurdles for the success of organ transplantation for those organs that are procured from brain dead organ donors (BDOD). Often, hormonal alterations cause hemodynamic instability after brain death. Recently, the use of inotropic agents and replacement of depleted hormones in a process known as hormonal replacement (HR) have helped preserve organ function. Currently, hormonal therapy can include the administration of four hormones: thyroid hormone (triiodothyronine (T3) or levothyroxine (T4)), corticosteroids (cortisol or methylprednisolone), insulin, and anti-diuretic hormone (DDAVP or arginine vasopressin), all of which have been used alone or in various combinations (Cooper et al, 2009; Novitzky et al, 2006; Novitzky et al, 1984; Rosendale et al, 2002; Rosendale et al, 2003). Although there are reports showing that using combination hormonal therapy could significantly improve the success of organ transplantation, there has not been reported an optimal combination hormonal therapy regimen that maximizes multiple organ procurements and improves organ quality from BDOD. In this report, we intend to identify the best HR modality from all possible combinations of the four hormones by using data from the United Network for Organ Sharing (UNOS). However, selecting the best HR modality from multiple HR modalities for multiple organs is statistically challenging.

To select the best HR modality from many available HR regimens is a multiple comparison problem in statistics. Traditional pair-wise comparison (MCA) between every pair of treatments is statistically conservative. Hsu et al proposed a multiple comparison method to select the best treatment, known as the MCB, which is adapted from Dunnett's multiple comparison with control (MCC) method by comparing each treatment with the best of the other treatments (Dunnett, 1955; Dunnett, 1980; Hsu, 1981; Edwards et al, 1983; Hsu, 1992). Since it was developed, the MCB has been widely used to select the best treatment based on single continuous endpoints. However, sometimes one needs to select the best treatment based on multiple endpoints, which usually are not continuous. In this report, we used a two step method to select the best HR modality based on ten binary endpoints.

2. Methods

2.1 Data

The data used for the analysis was provided by the United Network for Organ Sharing (UNOS, Richmond VA). A total of 71,571 organ donors were registered at UNOS from January 1st 2000 to December 31st 2009. A subset of 40,124 subjects, who were not donors of Donation after Cardiac Death (DCD) and had confirmed HR or non-HR therapy, was used for the analysis. There were 10 organs that could be procured from each donor; they included heart, double lungs (transplanted to same recipient), right lung, left lung, double kidneys (transplanted to same recipient), left kidney, right kidney, liver, intestines, and pancreas. Four classes of hormones—thyroid hormone (T3 or T4), corticosteroid (cortisone), vasopressin (arginine vasopressin or DDAVP), and insulin—were used to sustain donor's condition and to improve organ procurement. An individual donor could have received one or more of these four hormones or not received any of them. For convenience of illustration, the hormonal treatment modalities are categorized by thyroid hormone as shown in Table 1.

2.2 Analytical method

Two metrics were calculated to evaluate HR efficacy on BDOD organ procurement and transplantation: 1) average number of organs transplanted per individual donor given a hormonal treatment modality calculated as total number of organs transplanted divided by total number of donors within that treatment modality group and 2) transplantation rate for a specific organ given a hormonal treatment modality calculated as number of the organ transplanted divided by all the organ dispositions among the donors in that modality group. The data management and analysis were performed using SAS 9.2.1 (SAS Inc., Cary NC). Missing values were excluded from the analysis.

2.3 Estimation of predicted margins

To select the best HR modality for organ procurement, we used a logistic regression model to estimate the predicted margins or the logits for each particular organ

Table 1: Combinations of HR modalities

Group	Hormone Combinations	N
A1	Thyroid+Steroid+Insulin+Vasopressin	10669
A2	Thyroid+Insulin+Vasopressin	2935
A3	Thyroid+Steroid+Insulin	1363
A4	Thyroid+Steroid+Vasopressin	1118
A5	Thyroid+Vasopressin	4003
A6	Thyroid+Insulin	1545
A7	Thyroid+Steroid	580
A8	Thyroid	809
B1	Steroid+Insulin+Vasopressin	3553
B2	Insulin+Vasopressin	3655
B3	Steroid+Insulin	985
B4	Steroid+Vasopressin	1328
B5	Vasopressin	2158
B6	Insulin	3028
B7	Steroid	791
B8	No Hormone	1604

type, including heart, double lungs, single lung, liver, double kidneys, single kidney, pancreas, and intestine. The logits for each type of organ was estimated based on the following model, which adjusted for age, gender, and body mass index (BMI):

$$\log it(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \cdot HR + \beta_2 \cdot AGE + \beta_3 \cdot GENDER + \beta_4 \cdot BMI$$

where p is the probability of successfully procuring an organ given certain HR treatment.

Furthermore, we selected the best HR modality for multiple organ procurement by constructing a generalized linear mixed model to estimate the logits, defined as the following:

$$\log it(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \cdot HR + \gamma \cdot Type$$

where p is the probability of successfully procuring an organ given certain HR treatment; β_s is the fixed effect; and γ is the random effect. Four variance-covariance structures were selected to estimate the random effect, i.e. variance component (VC), compound symmetry (CS), autoregressive (AR), and Toepliz (TOEP).

2.4 Multiple comparisons procedures

To select the best HR modality for single and multiple organ procurement for transplantation, multiple comparison procedures (MCA, MCC, and MCB) were applied based on the logits from either the logistic regressions or generalized linear mixed models. Let $\theta_1, \theta_2, \ldots, \theta_k$ denote the treatment effects of logits estimated from the models described above. In the MCA setup, the parameters of interest were $\theta_i - \theta_j$ for all $i \neq j$, and there were $k(k-1)/2$ or 120 pair-wise comparisons. In the MCC setup, the HR modality with the highest value was used as the control and denoted as treatment c. Then the parameters of interest were $\delta_i = \theta_i - \theta_c$ for $i = 1,2,\ldots,k-1$, which in this study involved 15 comparisons. Similar to the MCC, the parameters of interest for the MCB were the difference between each HR modality and the true best of the others. For an HR modality i with an unknown real value θ_i, $\theta_{[k]} = \max_{1\le i\le k} \theta_i$. If modality i is preferred to modality j when $\theta_i > \theta_j$ the parameters $\delta_i = \theta_{[k]} - \theta_i$ reflect inversely the goodness of each treatment relative to the best treatment, for $i = 1,2,\ldots,k-1$, which in this study involved 15 comparisons. The simultaneous confidence intervals on the parameters of interests for the multiple comparison procedures were constructed to assess efficacy for each HR modality for individual and multiple organ procurement. The Tukey MCA procedure (Tukey, 1993) was used for all pair-wise comparisons, and the Dunnett-Hsu MCC and the Hsu MCB procedures were used to select the best HR modality based on the largest mean logits or LS-means.

3. Results

The best HR modality selection was based on single organ type procurement and multiple organ procurement and transplantation, including average number of multiple organ transplantations for individual BDOD.

3.1 General information for BDOD organ procurements

For each BDOD, there were ten organs that could be procured and transplanted, for which lungs (left and right) and kidneys (left and right) were considered as two single lung or kidney transplantations when they were transplanted into different recipients and as a double transplantation when two lungs or two kidneys were transplanted into the same recipient. Table 2 lists the overall success rates for each organ type transplanted to recipients without specifying HR information. Liver, single left kidney, single right kidney, and double kidneys transplantations had higher success rates of 81.0%, 74.1%, 73.2%, and 62.8%, respectively, than those of heart and double lungs transplantations (31.1% and 31.55). Pancreas, single right lung, single left lung, and intestine transplantations had the lowest success rates, ranging from 2.5% to 18.9% (Table 2).

Table 2 Success rates for different organ transplantations

Organ	*Transplanted*	*All Donors*	*Rate (%)*
Heart	12461	40124	31.1
Double Lungs	5297	16806	31.5
Left Lungs	1562	23318	6.7
Right Lungs	1432	23318	6.1
Liver	32103	39643	81.0
Double Kidneys	1555	2478	62.8
Left Kidneys	27891	37646	74.1
Right Kidneys	27539	37646	73.2
Pancreas	7595	40121	18.9
Intestine	997	40120	2.5

3.2 The best HR evaluation based on individual organ transplantations

Generally, multiple organs are procured from a BDOD treated with certain HR modality. In order to assess whether a certain HR modality is good for one organ but not good for others, we performed the best HR modality evaluation for each individual organ.

Figure 1 shows predicted margins or LS-means in SAS estimated from logistic regressions based on the success rates for different organ transplantations across sixteen treatment categories, i.e. fifteen HR modalities and no HR treatment. The largest LS-mean of HR modalities was identified for each organ type. The HR effects were reflected by the values of the predicted population margins or LS-means. The higher value had the better effect. The simultaneous confidence intervals for LS-mean differences between each HR modality and the HR modality with the largest LS-mean value are shown in Figure 2. When the Dunnett-Hsu procedure was used, the best HR modalities for heart, double lungs, single lung, liver, double kidneys, single kidney, pancreas, and intestine transplantations were A5, A4, A2, B4, A4, A1, A4, and A4 respectively. However, the best HR A5 for heart transplantation was not significantly better than A1, A2, and A4; similarly, A2 was not significantly better than A1 for single lung transplantation, and A1 was not significantly better than A4 for single kidney transplantation. When the Hsu MCB procedure was used, the best HR modalities for heart, double lungs, single lung, liver, double kidneys, single kidney, pancreas, and intestine transplantations were A5, A4, A2, B4, A4, A1, A4, and A4, respectively. As shown in Figure 3, similar results were found when the Dunnett-Hsu MCC procedure was used; whereas, the best HR A1 was not significantly better than A2 and A5 for single kidney transplantation, and A4 was not significantly better than B4.

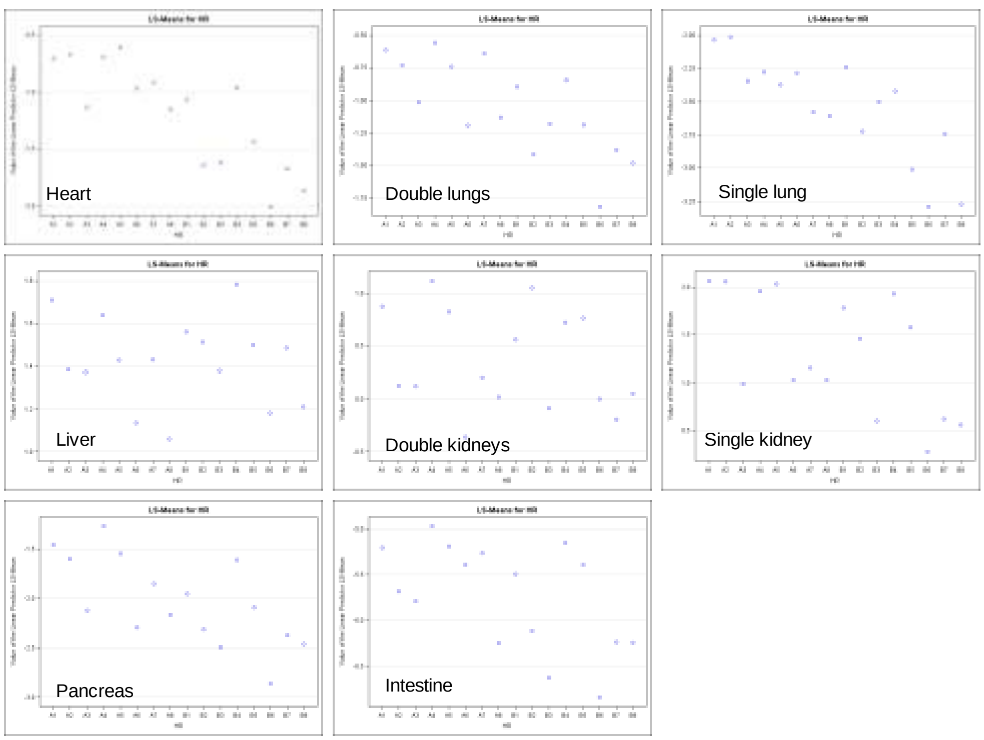

Figure 1 Predicted margins of the logits or LS-means estimated from each logistic regression model based on the successful transplantation rates for each organ (clockwise: heart, double lungs, single lung, liver, double kidneys, single kidney, pancreas, and intestine) across HR groups.

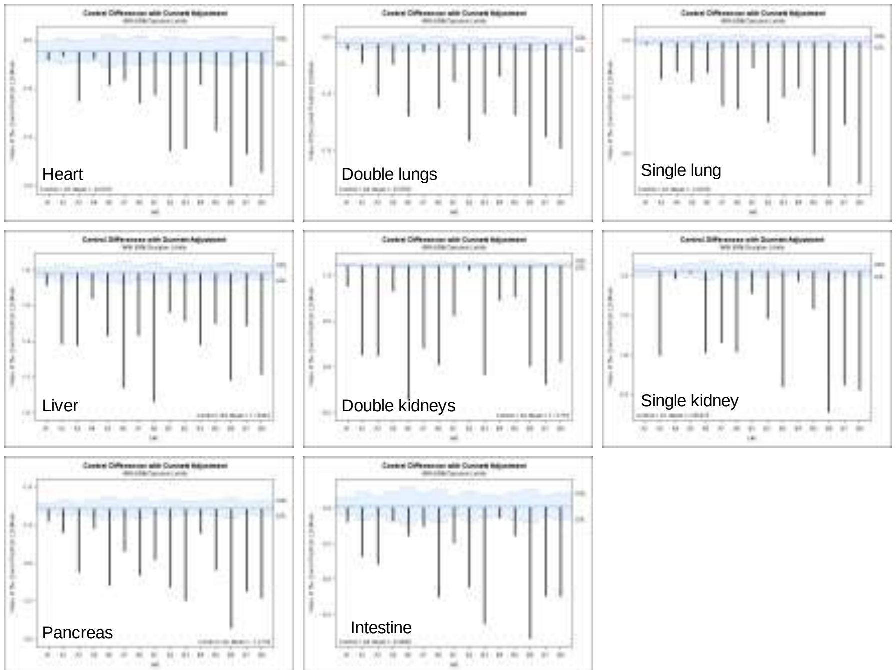

Figure 2 Dunnett-Hsu's MCC simultaneous confidence intervals for the differences of the logits estimated from each logistic regression model using multiple comparison with the highest LS-mean or logit (clockwise: heart, double lungs, single lung, liver, double kidneys, single kidney, pancreas, and intestine) across HR groups.

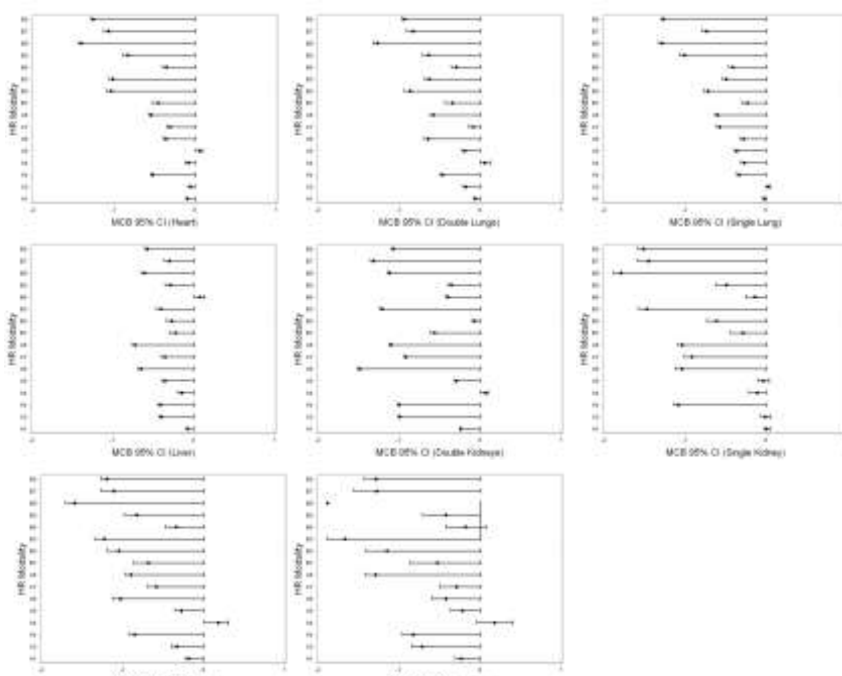

Figure 3 Hsu's MCB simultaneous confidence intervals for the differences of the logits estimated from each logistic regression model using multiple comparison with the highest LS-mean or logit (clockwise: heart, double lungs, single lung, liver, double kidneys, single kidney, pancreas, and intestine) across HR groups.

3.3 The best HR selection for multiple organ transplantations

There were two parameters used for the best HR modality selection for multiple organ procurement and transplantation, i.e. average number of organs transplanted given individual donors and the success rates for multiple organ transplantations.

This approach was based on the LS-means estimated from the generalized linear mixed models using four covariance structures for estimating the random effect. As shown in Table 3, though there were slight LS-mean value differences among these four covariance structure models, there were no differences in terms of the ranks of the HR effect, and the HR group A1 was the top rank of all four models. Pair-wise comparisons were performed using the Tukey procedure. The results showed that group A1 was the best HR modality; however, it was not significantly better than HR groups A4, A2 and A5 after the Tukey adjustment (Figure 4).

Table 3 LS-means estimated from different covariance structure models

HR Group	CS	AR	VC	TOEP
A1	-0.4858	-0.4145	-0.4858	-0.2523
A2	-0.5713	-0.5000	-0.5713	-0.3378
A3	-1.0438	-0.9725	-1.0438	-0.8104
A4	-0.4889	-0.4176	-0.4889	-0.2554
A5	-0.5544	-0.4831	-0.5544	-0.3209
A6	-1.0706	-0.9993	-1.0706	-0.8372
A7	-0.8898	-0.8185	-0.8898	-0.6564
A8	-1.1378	-1.0665	-1.1378	-0.9043
B1	-0.7301	-0.6589	-0.7301	-0.4967
B2	-1.0317	-0.9604	-1.0317	-0.7983
B3	-1.3200	-1.2487	-1.3200	-1.0865
B4	-0.6089	-0.5376	-0.6089	-0.3755
B5	-0.9198	-0.8485	-0.9198	-0.6863
B6	-1.5969	-1.5256	-1.5969	-1.3634
B7	-1.3001	-1.2288	-1.3001	-1.0667
B8	-1.4209	-1.3496	-1.4209	-1.1875

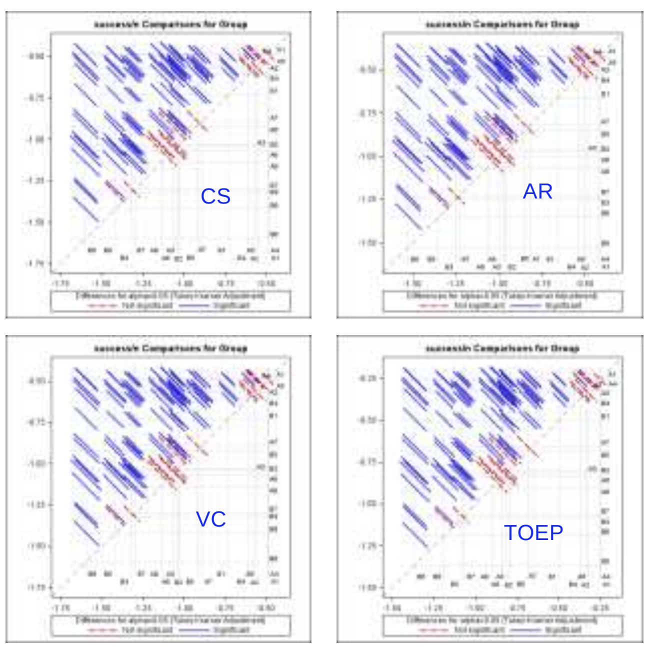

Figure 4 Pair-wise confidence intervals from generalized linear mixed models with four different covariance structures using the Tukey procedure (clockwise: compound symmetry, autoregressive, variance component, and toepliz) across HR groups.

When the Dunnett-Hsu procedure was used to construct the simultaneous confidence intervals or 95% decision limits, the results showed similar results to those when pair-wise comparisons were performed. A1 was the best HR modality for multiple organ procurement and transplantation (Figure 5). However, it was not significantly better than A2 and A5.

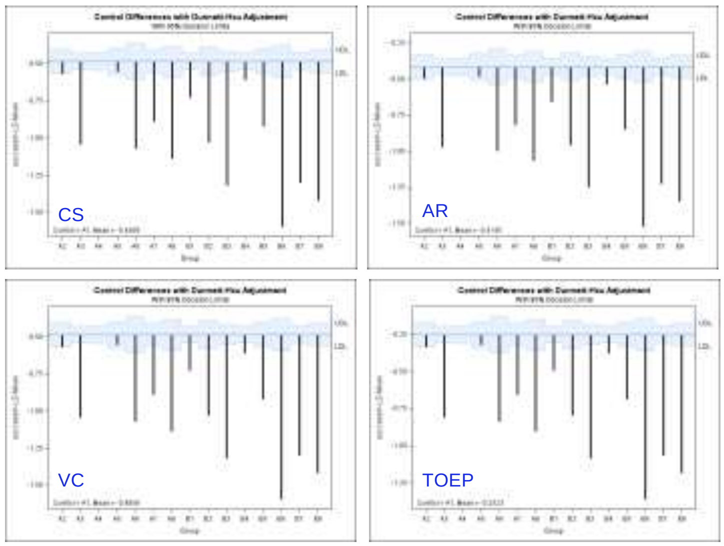

Figure 5 Simultaneous confidence intervals for the differences of the logits estimated from generalized linear mixed models with four different covariance structures using the Dunnett-Hsu procedure (clockwise: compound symmetry, autoregressive, variance component, and toepliz) across HR groups.

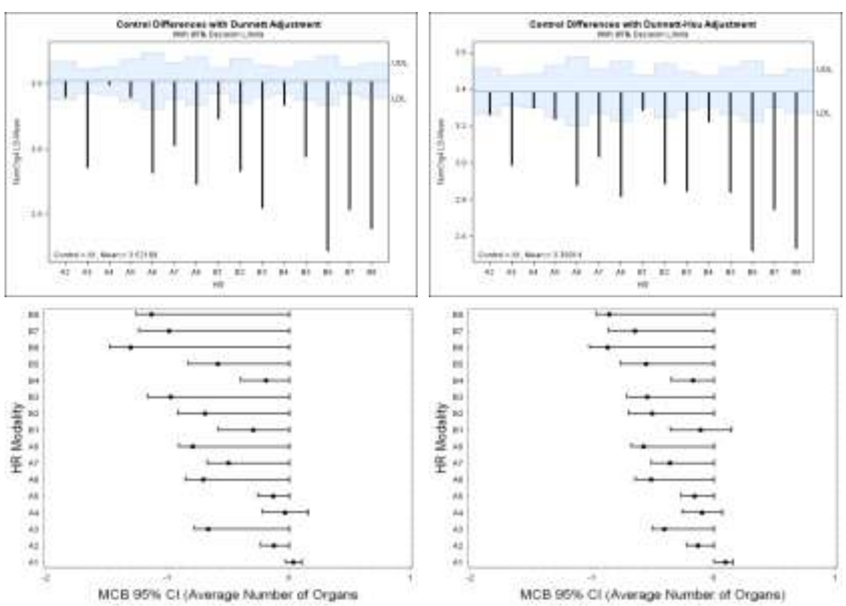

Figure 6 LS-means and the simultaneous confidence intervals for the differences from generalized linear model based on the average number of organs transplanted for individual BDOD using the Dunnett_Hsu MCC procedure (upper panels) and the Hsu MCB procedure (lower panels) across HR groups. The right panels were based on the models adjusted for age, gender and BMI.

The same HR modality was identified using the average number of organs procured and transplanted for individual BDOD based on the general linear models with or without adjusting the covariates, age, gender, and BMI. As shown in Figure 6, group A1 was the best HR modality using both Dunnett-Hsu MCC and Hsu MCB procedures. However, without the covariate adjustment, A1 was not significantly better than A2, A4, and A5 when the MCC was used, and was not significantly better from A4 and B4 when the MCB was used; whereas after the covariate adjustment, A1 was not significantly better than the A4 when the MCB was used.

4. Conclusion

HR has been widely used to increase BDOD organ procurement. So far there is no research showing what the optimal HR combinations are for specific types of organ and multiple organ procurement. In this report, we proposed a two-step approach using generalized linear models (GLM) or generalized linear mixed models (GLMM), and multiple comparison procedures and analyzed 40,124 BDOD with HR treatment information to identify the best HR modality for both individual and multiple organ procurement using Tukey's MCA, Dunnett-Hsu's MCC, and Hsu's MCB. We found that thyroid, steroid, and vasopressin hormones played important roles in hormone management for organ procurement and transplantation. The combination of thyroid hormone, steroids, vasopressin, and insulin was the best hormonal therapy for multiple organ procurement and transplantation. The detailed HR combinations for each organ type and multiple organ procurement are summarized in the following table.

Organ Procurement	*HR group: HR Combinations*
Single Organ Procurement	
Heart	A5: Thyroid+Vasopressin
Double lungs	A4: Thyroid+Steroid+Vasopressin
Single lung	A2: Thyroid+Insulin+Vasopressin
Liver	B4: Steroid+Vasopressin
Double kidneys	A4: Thyroid+Steroid+Vasopressin
Single Kidney	A1: Thyroid+Steroid+Insulin+Vasopressin
Pancreas	A4: Thyroid+Steroid+Vasopressin
Intestine	A4: Thyroid+Steroid+Vasopressin
Multiple Organ Procurements	
Multivariate success rates	A1: Thyroid+Steroid+Insulin+Vasopressin
Average number of organs	A1: Thyroid+Steroid+Insulin+Vasopressin

To select the best HR modality, we applied multiple comparison methods to control type I error rates. Tukey's MCA is the most powerful test when performing all pair-wise comparisons and Dunnett's MCC is the most powerful test when comparing to a 'control'; whereas Hsu's MCB is the most powerful test when not requiring all pair wise comparisons, which compares between each sample mean and the "best" of all the other means, where one specify that "best" means either largest or smallest. It is a modification of Dunnett's MCC method by treating the 'best' as an unknown parameter. The purpose is to select which group(s) is/are the best: not significantly different from each other but significantly better than the others. In this data analysis, we yielded the similar results using the three methods because of the large sample size. We applied the Dunnet-Hsu MCC by treating the highest value by treatment effect as the 'control' group and testing whether the other treatment effects were inferior to the ”control,” which then we considered as the best HR, as a comparison with the results from the Hsu MCB. The best HR selections were based on the simultaneous confidence intervals using both MCC and MCB procedures.

Acknowledgements

The authors gratefully thank United Network for Organ Sharing (UNOS) for sharing their database.

References

D.K. Cooper, D. Novitzky, W.N. Wicomb, M. Basker, J.D. Rosendale, M.H. and Kauffman, “A review of studies relating to thyroid hormone therapy in brain-dead organ donors,” *Front Biosci*, vol.1, no.14, pp.3750-70, 2009.

C. W. Dunnett, "A multiple comparison procedure for comparing several treatments with a control", *Journal of the American Statistical Association*, vol. 50, pp.1096–1121, 1955.

C. W. Dunnett, “Pairwise Multiple Comparisons in the Unequal Variance Case,” *Journal of the American Statistical Association*, vol. 75, pp.796–800, 1980.

D.G. Edwards, and J. C. Hsu, “Multiple comparisons with the best treatment”, *Journal of the American Statistical Association*, vol. 78, no. 384, pp. 965-971, 1983.

J.C. Hsu, "Simultaneous Confidence Intervals for All Distances From the 'Best'," *Annals of Statistics*, vol. 9, pp. 1026-1034, 1981.

J.C. Hsu, “The Factor Analytic Approach to Simultaneous Inference in the General Linear Model,” *Journal of Computational and Graphical Statistics*, vol.1, pp.151–168, 1992.

D. Novitzky, W.N. Wicomb, D.K.C. Cooper, A.G. Rose, R.C. Fraser, and C.N. Barnard, “Electrocardiographic, hemodynamic and endocrine changes occurring during experimental brain death in the Chacma baboon,” *J Heart Transplant,* vol. 4, pp.63-69, 1984.

D. Novitzky, D.K.C. Cooper, J.D. Rosendale and H.M. Kauffman, “Hormonal therapy of the brain-dead organ donor: experimental and clinical studies,” *Transplantation,* vol. 82, pp.1396-1401, 2006.

J.D. Rosendale, F.L. Chabalewski. M.A. McBride, E.R. Garrity, B.R. Rosengard, F.L. Delmonico and H.M. Kaufmann, “Increased transplanted organs from the use of a standardized donor management protocol,” *Am J Transplant,* vol. 2, pp.761-768, 2002.

J.D. Rosendale, H.M. Kauffman, M.A. McBride, F.L. Chabalewski, J.G. Zaroff, E.R. Garrity, F.L. Delmonico, and B.R. Rosengard, “Aggressive pharmacological donor management results in more transplanted organs,” *Transplantation,* vol. 75, pp.482-487, 2003.

J. W. Tukey, “The problem of multiple comparisons”, Unpublished manuscript. *In The Collected Works of John W. Tukey VIII. Multiple Comparisons: 1948-1983 1-300.* Chapman and Hall, New York. 1993.