

# Statistical Analysis of Hippocampus Shape Using a Modified Mann-Whitney-Wilcoxon Test

Nikhil Ram Mohan

*Johns Hopkins University, Center for Imaging Science  
nikhil@cis.jhu.edu*

Carey Priebe

*Johns Hopkins University, Center for Imaging Science  
cep@jhu.edu*

Youngser Park

*Johns Hopkins University, Center for Imaging Science  
youngser@jhu.edu*

Majnu John

*Weill Cornell Medical College, Dept. of Biostatistics  
maj2023@med.cornell.edu*

## ***Abstract***

*The Mann-Whitney-Wilcoxon (MWW) test statistic, while distribution-free, suffers from a loss of efficacy for certain underlying distributions. In this manuscript, we instead use a data-adaptive weighted generalized Mann-Whitney-Wilcoxon (AWGMWW) test statistic, one that is optimal in the Pitman Asymptotic Efficacy (PAE) sense, to discern differences in hippocampus shape among twin populations with or without Major Depressive Disorder (MDD). Using this more powerful statistic, we verify, based on a previous study using the MWW statistic, that a high-risk group is more similar to the control group than the depressed group, the depressed group is more similar to the control group than the high-risk group, and the control group cannot be distinguished as more similar to one group or the other in terms of hippocampus shape.*

## **1. Introduction**

### **1.1. Nonparametric Statistics**

Wilcoxon and Mann-Whitney [1,2] are principally responsible for the advent of nonparametric statistics. The unbiased Rank-Sum and U-statistic have since been applied in a great deal of applications, and are most effective in situations where information about underlying distributions is not known *a priori*. One such application is in interpoint distance analysis [3]. The major obstacle in preventing even wider use of these statistics is the loss of efficacy owing to their nonparametric and distribution-free properties. To combat this, Xie & Priebe [4] provided weighted generalizations of the Mann-Whitney-Wilcoxon (WGMWW) and of the Wilcoxon-Signed-Rank (WGWSR) statistics, which were shown to be optimal in the Pitman Asymptotic Efficacy (PAE) sense. Unfortunately, these weighted generalizations were not practically viable owing to their containing parameters which were functions of the unknown null distribution. A data-adaptive alternative (AWGMWW), which has efficacy and power as good as WGMWW, and dependent on two parameters  $r$  and  $s$ , was provided by John & Priebe [5]. The efficacy and power of AWGMWW was compared with WGMWW and the classical MWW test statistic for

several underlying densities. The results suggest that AWGMWW is only marginally more powerful and efficacious when the underlying density is normal or mildly left-skewed, but optimal when the underlying distribution is strongly right skewed, asymmetric bimodal or heavily kurtotic [5].

## 1.2. Major Depressive Disorder

Major Depressive Disorder (MDD), or Clinical Depression, is one of the most prevalent mental disorders in the United States, where approximately 16.2% of the population is affected at least once in their lifetime [6]. The disorder is characterized by a “combination of symptoms that interfere with an individual's ability to work, sleep, study, eat, and enjoy once-pleasurable activities” [7]. In order for an individual to be diagnosed with MDD, at least one of the primary symptoms, namely depressed mood or anhedonia, and at least three of the secondary symptoms, among marked weight loss, insomnia, fatigue, and thoughts of suicide, must be present for a period of six months or more [8]. The physiology of MDD has been examined in great detail over the past few decades. It is known that changes in the amounts of the neurotransmitters *serotonin* and *dopamine* are associated with depressive symptoms, and restoring the amounts of these neurotransmitters to normal levels is the function of most modern anti-depressant drugs [9]. Neuroimaging studies have revealed that certain structural abnormalities such as “enlarged ventricles, sulci, or reduced volume of the frontal lobe and basal ganglia” are also significantly correlated with MDD [2,9].

## 1.3. Computational Anatomy

Computational Anatomy (CA) has only recently emerged as a discipline; it involves “the development of mathematical and software tools...specialized to the study of brain anatomy” [10]. In fact, the shape of the brain as related to its anatomy is quite complex, and thus difficult to quantify. The underlying principle in CA is to “construct a mapping model” [3] that can measure the difference in shape between two brain regions-of-interest (ROIs). Several publications have used CA models on subsections of the brain to further the understanding of several psychological disorders; Posener *et al.* [11] examined the role of the hippocampus, a small structure found in the limbic system of the brain, in MDD, Miller *et al.* [12] studied the role of the cingulate gyrus in Dementia of the Alzheimer's Type (DAT), while Csernansky *et al.* [13] correlated hippocampus shape changes with schizophrenia.

## 1.4. Goals

Recently, significant differences in hippocampus shape were gleaned between twin populations with or without Clinical Depression using the classical MWW test and interpoint comparison analysis [3]. The probability density function (*pdf*) of the underlying distribution, simulated using a kernel estimator, was shown to be mildly right-skewed. The goal of this study is to develop an algorithm for the AWGMWW statistic and compare its efficacy and power with the classical MWW when applied towards interpoint comparison data from Park *et al.* [3].

## 2. Data

Using Missouri birth records, three twin populations, including both monozygotic (MZ) and dizygotic (DZ) twin pairs, were recruited according to varying levels of Clinical Depression: (1) the Control group (CTRL) was unaffected by depression, (2) the High-Risk (HR) group included

one twin in a pair having Depression and the other not, and the (3) Major Depressive Disorder (MDD) group having both twins in a pair being clinically depressed [3]. There were 59 twins (29 pairs, one unpaired) in CTRL, 22 twins in HR, and 33 twins (16 pairs, one unpaired) in MDD. All recruits were female, and a screening procedure excluded individuals with conditions that may influence structural changes of the brain, namely loss of consciousness greater than 5 minutes, pregnancy, and any chronic neurological illnesses [3].

High-resolution Magnetic Resonance Imaging (MRI) scans were obtained for all 114 individuals in the study, a first step in the analysis of hippocampus shape. Three MPRAGE scans (“160 slices at 256x256 FoV, 1 mm<sup>3</sup> isotropic voxels”) were acquired using the Siemens Vision/Sonata 1.5T scanner. Three-dimensional surfaces of the left and right hippocampi were extracted from the scans in accordance with current neuroanatomical guidelines [3].

### 3. Methods

#### 3.1. Image Processing

The following protocol was the same as the one employed by Park *et al.* [3]. For every hippocampus surface, 22 three-dimensional landmarks were anatomically defined. An “interpoint comparison matrix”  $\tilde{D}$  was generated by applying a “non-parametric Landmark Matching (LM) transformation”, which computes as a measure of shape difference the energy of the minimizing diffeomorphism between every possible pair of hippocampus landmarks [3].

For every pair of hippocampi, the following “error criterion” was used [3]:

$$\varphi^* = \arg \inf_{\varphi} \sigma d(\varphi_{identity}, \varphi)^2 + \sum_{i=1}^N \|\varphi(x_i) - y_i\|^2$$

where “ $d$  is a geodesic distance in a group of diffeomorphisms and  $\sigma > 0$  is a regularization parameter which controls the relative contribution of transformation complexity and the landmark mismatch to the optimization objective” [3]. The norm of the optimal mapping i.e.,  $LM(x, y) = \|\varphi^*\|$ , was used as the final measure of shape difference between the pair of hippocampi [3].

Such an analysis was conducted independently on the left hippocampi, the right hippocampi, and finally both hippocampi together. The resulting three interpoint comparison matrices were asymmetric, with dimension 114x114, and zeroes on the diagonal. Each of the matrices were symmetrized from  $\tilde{D}$  to  $D$  using  $d_{ij} = \min(d_{ij}, d_{ji})$ , since the asymmetry was not an accurate reflection of hippocampus shape space. For more details, see Park *et al.* [3].

#### 3.2. Statistical Analysis

**3.2.1. Stochastic Ordering:** Each interpoint comparison matrix generated by the LM transformation (see Section 3.1), being nonparametric, has an underlying probability density function (pdf) that is unknown. Park *et al.* [3] had shown, using “kernel probability density estimates” [14], that there exists a stochastic ordering relationship between the interpoint comparisons of HR and CTRL, written  $d(HR, CTRL)$ , and those of HR and MDD, written  $d(HR, MDD)$ , such that  $d(HR, CTRL) <^{st} d(HR, MDD)$  [3]. Figure 1, a plot of the kernel pdfs, demonstrates this relationship.

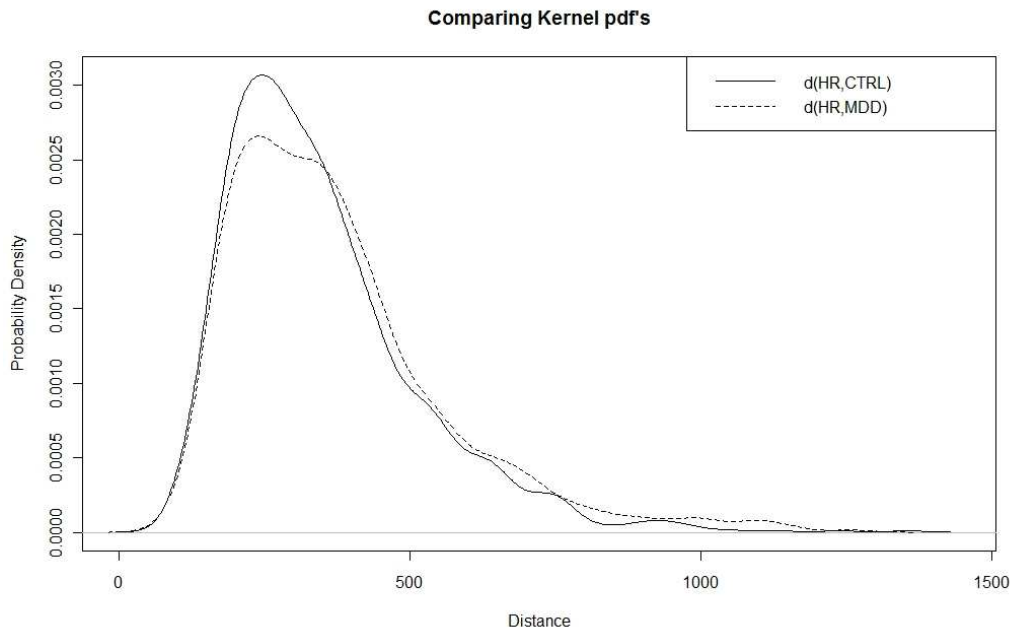


Figure 1: This figure shows two kernel pdfs for  $D_{LM-Left}$ : the solid line shows  $d(HR,CTRL)$  and the dashed line shows  $d(HR,MDD)$  - this representation verifies the relationship obtained by Park *et al.* [3].

Based on the structure of  $D$ , we conclude that every row of the interpoint comparisons corresponding to any one HR subject provides two vectors of comparison: one that compares that HR subject to every CTRL subject, and one that compares that HR subject to every MDD subject [3]. We call these two samples  $\{d(HR_i,CTRL)\}$  and  $\{d(HR_i,MDD)\}$  respectively; the HR subject's own twin is ignored when considering these samples in order to eliminate any bias that may arise from twinnedness not due to the underlying disease condition [3]. Park *et al.* conducted a two-sample hypothesis test using the MWW statistic, where the null hypothesis was equality of the distributions of  $d(HR,CTRL)$  and  $d(HR,MDD)$  against the alternative of stochastic ordering [3]. Here, we instead use the AWGMWW statistic (with  $r = s = 2$ ) to conduct the same hypothesis test and obtain a  $p$ -value.

Figure 2 and Figure 3 compare the quantile-quantile plots of the  $p$ -values obtained by Park *et al.* using the MWW statistic [3], and those obtained in this study using the AWGMWW statistic, respectively.

If the null hypothesis were true, the  $p$ -values for both these plots should lie on the diagonal i.e. be distributed  $Uniform(0,1)$ . Both plots, although with different  $p$ -values, clearly suggest the alternative of stochastic ordering between the distributions. This seems to show that the modified statistic does not alter the overall result, but affects the strength of the underlying conclusion.

**3.2.2. Classification:** Recall, we have two samples  $\{d(HR_i,CTRL)\}$  and  $\{d(HR_i,MDD)\}$  for each HR subject, and that a  $p$ -value of 0.5 indicates that the distributions of these samples are equal. Thus, we can use the one-sided ( $d(HR,CTRL) <^{st} d(HR,MDD)$ ) AWGMWW  $p$ -value (instead of the MWW  $p$ -value as implemented in Park *et al.* [3]) to

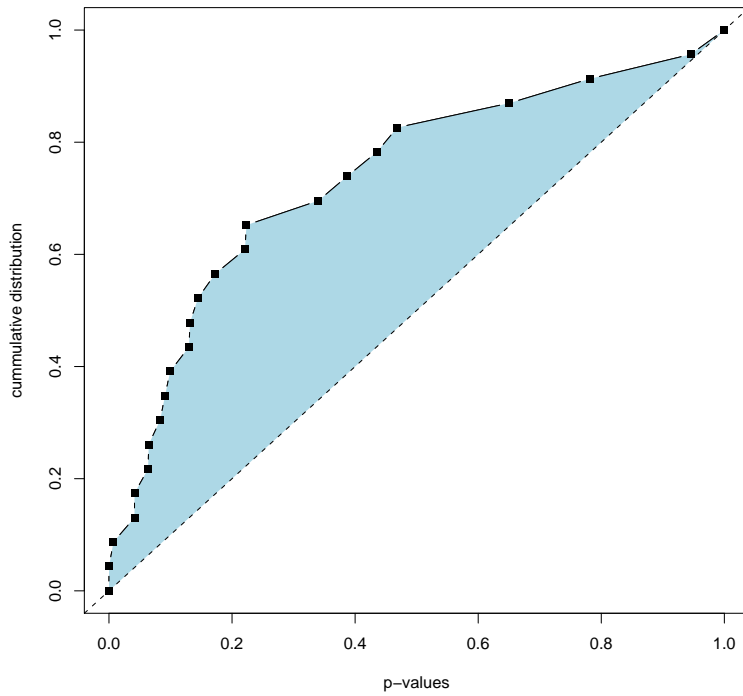


Figure 2: QQ-plot for the two samples  $d(HR,CTRL)$  and  $d(HR,MDD)$  from  $D_{LM-Left}$  using the MWW statistic - this verifies the distribution of p-values obtained by Park *et al.* [3].

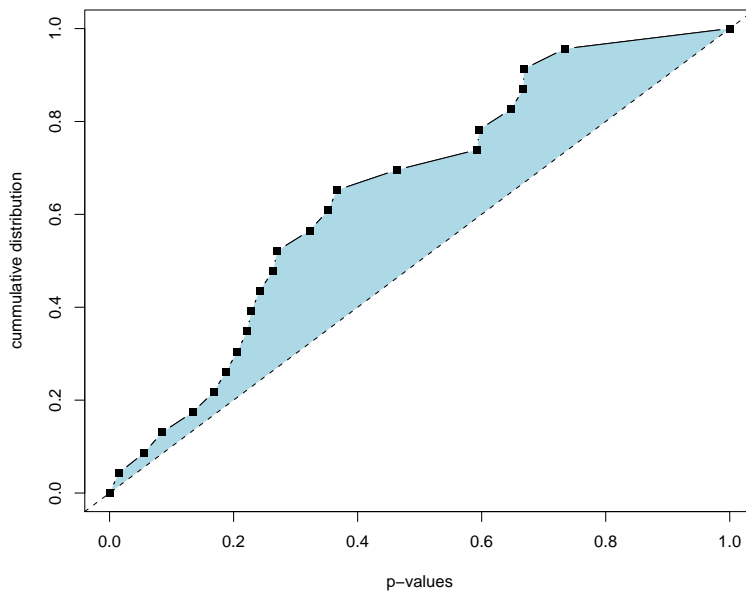


Figure 3: QQ-plot for the two samples  $d(HR,CTRL)$  and  $d(HR,MDD)$  from  $D_{LM-Left}$  using the AWGMWW statistic.

classify each HR subject as “closer” to CTRL or MDD. That is, if  $p \leq 0.5$  for some HR subject, we classify that subject as CTRL, and if  $p > 0.5$ , we classify that subject as MDD. Extending this procedure to every HR subject would allow us to assess classifier performance on the HR group as a whole. In this case, we have fixed the HR group as the population of interest; we can equally fix one of the other groups (namely CTRL or MDD) as the population of interest, and follow the same procedure as above. Also note we can apply our classifier on the left hippocampi alone, the right hippocampi alone, or both left and right hippocampi simultaneously. The results are tabulated in Table 1.

	<u>LM-Left</u>	<u>LM-Right</u>	<u>LM-Left&amp;Right</u>
<u>H:CvM</u>	17(19)	13(16)	15(20)
<u>H:MvC</u>	5(3)	8(6)	7(2)
<u>M:HvC</u>	14(5)	12(9)	11(6)
<u>M:CvH</u>	19(28)	21(24)	22(27)
<u>C:HvM</u>	40(48)	30(22)	33(31)
<u>C:MvH</u>	19(32)	29(33)	26(25)

Table 1: Classification of individual subjects based on AWGMWW  $p$ -values. The numbers in parentheses correspond to classification using the MWW statistic from Park *et al.* [3]. The column-label “LM-Left” corresponds to the LM procedure being applied on the left hippocampi alone, and the row label “H:CvM” corresponds to the number of HR subjects classified as CTRL as opposed to MDD, and similarly, “H:MvC” corresponds to the number of HR subjects classified as MDD as opposed to CTRL. The rest of the column- and row-labels are defined analogously.

## 4. Results

Table 1 shows some interesting results, where we observe that the general classification trends for the AWGMWW statistic are consistent with the MWW statistic - more HR subjects are classified as CTRL as opposed to MDD ( $p < 0.00005$  for MWW and  $p < 0.01$  for AWGMWW respectively), more MDD subjects are classified as CTRL than HR ( $p < 0.00005$  for MWW and  $p < 0.05$  for AWGMWW respectively), and CTRL subjects are more or less distributed evenly between the two groups ( $p \sim 0.2$  for alternate hypothesis that CTRL is more similar to HR).

## 5. Conclusions

The overarching results obtained using the AWGMWW statistic are consistent with those obtained using the MWW statistic; in terms of hippocampus shape, the high-risk group is more similar to the control group than the depressed group, the depressed group is more similar to the control group than the high-risk group, and the control group cannot be distinguished as more similar to one group or the other. Considering the hippocampus shape of the three populations in the one-dimensional Euclidean space  $\mathbb{R}^1$  *a la* Park *et al.* [3], we again show that the control group is approximately in the middle between the high-risk and depressed groups (see Figure 4). However, we have demonstrated that AWGMWW, a PAE-optimal statistic, has more statistical power and eliminates the loss of efficacy for this analysis, which increases the validity and veracity of the results obtained. The AWGMWW statistic shows promise for future applications involving nonparametric statistical testing.

# HR

# CTRL

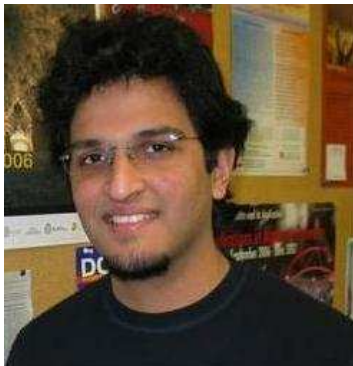
# MDD

Figure 4: Hippocampus shape-space if viewed in one dimension, similar to the representation depicted by Park *et al.* [3].

## 6. References

- [1] F. Wilcoxon, "Individual Comparisons by Ranking Methods", *Biometrics* 1, 1945, pp. 80-83.
- [2] H.B. Mann and D.R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other", *Annals of Mathematical Statistics* 18, 1947, pp 50-60.
- [3] Y. Park, C. E. Priebe, M.I. Miller, N.R. Mohan, and K.N. Botteron, "Statistical Analysis of Twin Populations using Dissimilarity Measurements in Hippocampus Shape Space", *Journal of Biomedicine and Biotechnology*, 2008, Article ID 694297, 5 pages. doi:10.1155/2008/694297.
- [4] J. Xie and C.E. Priebe, "A weighted generalization of the Mann-Whitney-Wilcoxon statistic", *Journal of Statistical Planning and Inference* 102, 2002, pp 441-466.
- [5] M. John and C.E. Priebe, "A data-adaptive methodology for finding an optimal weighted generalized Mann-Whitney-Wilcoxon statistic", *Computational Statistics & Data Analysis* 51, 2007, pp. 4337-4353.
- [6] R.C. Kessler, P. Bergland, O. Demler, R. Jin, D. Koretz, K. Merikangas, A.J. Rush, E.E. Walters, and P.S. Wang, "The Epidemiology of Major Depressive Disorder: Results from the National Comorbidity Survey Replication (NCS-R)", *Journal of the American Medical Association* 289, 2003, pp. 3095-3105.
- [7] Depression: National Institute of Mental Health, <http://www.nimh.nih.gov/health/publications/depression/complete-index.shtml>
- [8] *Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Association, Fourth Edition, 1994.
- [9] L.B. Alloy, J.H. Riskind, and M.J. Manos, "Abnormal Psychology: Current Perspectives", McGraw-Hill Companies, New York, Ninth Edition, 2005, pp. 278.
- [10] M. Miller, A. Banerjee, G. Christensen, S. Joshi, N. Khaneja, U. Grenander, L. Matejic, "Statistical Methods in Computational Anatomy", *Statistical Methods in Medical Research* 6, 1997, pp. 267-299.
- [11] J. Posener, L. Wang, J. Price, M. Gado, M. Province, M. Miller, C. Babb, and J. Csernansky, "High-dimensional mapping of the hippocampus in depression", *American Journal of Psychiatry* 160, 2003, pp. 83-89.
- [12] M.I. Miller, C. Priebe, A. Qiu, B. Fischl, A. Kolasny, T. Brown, Y. Park, E. Busa, J. Jovicich, P. Yu, B. Dickerson, and R.L. Buckner, "The Morphometry BIRN: Collaborative computational anatomy: The perfect storm for MRI morphometry study of the human brain via diffeomorphic metric mapping", *Human Brain Mapping* 30, 2008, pp. 2132-2141.
- [13] J. Csernansky, S. Joshi, L. Wang, J. Haller, M. Gado, J. Miller, U. Grenander, and M. Miller, "Hippocampal morphometry in schizophrenia by high dimensional brain mapping", *Proceedings of the National Academy of Sciences* 95, 1998, pp.11406-11411.
- [14] B. Silverman, "Density Estimation for Statistics and Data Analysis", Chapman and Hall, New York, 1996.

## Authors



Nikhil Ram Mohan was born in Chennai, India in 1987. He enrolled at The Johns Hopkins University in the fall of 2005. He graduated in May 2008 with a Bachelor of Science degree in Applied Mathematics and Statistics, with a second degree in Psychological and Brain Sciences. He then pursued Master of Science in Engineering in Applied Mathematics and Statistics also at JHU, and graduated in May 2009. Throughout his academic career at JHU, Nikhil has been actively involved in research. He began working at the Center for Imaging Science on the Homewood Campus in March 2006, and since then has contributed to several projects in the area of nonparametric statistics and computational anatomy. These projects include the study of hippocampus shape differences between twin populations with or without Clinical Depression, the use of novel

nonparametric tests for the purpose of interpoint comparison analysis, and the examination of the respective roles of anatomy and physiology of the planum temporale in the onset of Schizophrenia. After graduating the Master's program, Nikhil plans to enroll in a combined MD/PhD program, and continue with medically-related research.