

RESEARCH ARTICLE

Robust multivariate nonparametric tests for detection of two-sample location shift in clinical trials

Xuejun Jiang¹, Xu Guo², Ning Zhang^{3,4}, Bo Wang⁵, Bo Zhang^{5*}

1 Department of Mathematics, Southern University of Science and Technology, Shenzhen, Guangzhou, P.R. China, **2** Department of Mathematical Statistics, School of Statistics, Beijing Normal University, Beijing, P.R. China, **3** Department of Health Promotion and Policy, School of Public Health and Health Sciences, University of Massachusetts, Amherst, Massachusetts, United States of America, **4** Meyers Primary Care Institute, A Joint Endeavor of University of Massachusetts Medical School, Reliant Medical Group, and Fallon Health, Worcester, Massachusetts, United States of America, **5** Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, Massachusetts, United States of America

* Bo.Zhang@umassmed.edu



OPEN ACCESS

Citation: Jiang X, Guo X, Zhang N, Wang B, Zhang B (2018) Robust multivariate nonparametric tests for detection of two-sample location shift in clinical trials. PLoS ONE 13(4): e0195894. <https://doi.org/10.1371/journal.pone.0195894>

Editor: Qizhai Li, University of the Chinese Academy of Sciences, CHINA

Received: January 29, 2018

Accepted: April 2, 2018

Published: April 19, 2018

Copyright: © 2018 Jiang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Dr. Xuejun Jiang's research was partially supported by Natural Science Foundation of China (11101432), Natural Science Foundation of Guangdong Province of China (2017A030313012), and Shenzhen Sci-Tech Fund (JCYJ20170307110329106). Dr. Xu Guo's research was partially supported by National Natural Science Foundation of China (11601227 and 11626130) and Natural Science Foundation of

Abstract

This article presents and investigates performance of a series of robust multivariate nonparametric tests for detection of location shift between two multivariate samples in randomized controlled trials. The tests are built upon robust estimators of distribution locations (medians, Hodges-Lehmann estimators, and an extended U statistic) with both unscaled and scaled versions. The nonparametric tests are robust to outliers and do not assume that the two samples are drawn from multivariate normal distributions. Bootstrap and permutation approaches are introduced for determining the p -values of the proposed test statistics. Simulation studies are conducted and numerical results are reported to examine performance of the proposed statistical tests. The numerical results demonstrate that the robust multivariate nonparametric tests constructed from the Hodges-Lehmann estimators are more efficient than those based on medians and the extended U statistic. The permutation approach can provide a more stringent control of Type I error and is generally more powerful than the bootstrap procedure. The proposed robust nonparametric tests are applied to detect multivariate distributional difference between the intervention and control groups in the Thai Healthy Choices study and examine the intervention effect of a four-session motivational interviewing-based intervention developed in the study to reduce risk behaviors among youth living with HIV.

Introduction

In randomized controlled trials, effectiveness (or efficacy) of a treatment effect is constantly characterized by the difference between distributional locations of a treatment group and its control group. Hypothesis testing is the primary statistical inference approach in examining treatment effects in clinical trials, when it is conducted to detect whether there exists any

Jiangsu Province, China (BK20150732). The research project was partially funded by the National Institute of Mental Health (R34MH077523).

Competing interests: The authors have declared that no competing interests exist.

difference between distributional locations of the treatment group and the control group. When the primary endpoint is one-dimensional and normally distributed for both study groups, the two-sample t test is the standard tool. Yet, the two-sample t test may not be valid when normality assumption is violated. The two-sample t test is not robust to outliers and heavy-tail distributions. A number of robust nonparametric tests have been developed in the literature as a complement of the two-sample t test. The classic Wilcoxon-Mann-Whitney test [1] that used the rank sum is a nonparametric counterpart of the two-sample t test. Yuen [2] and Keselman et al. [3] recommended to construct the tests using trimmed means. Recently, Fried and Dehling [4] proposed a series of robust nonparametric tests for detecting univariate two-sample location difference. These tests were constructed based upon unscaled and properly scaled robust location estimators of distributions, including medians and Hodges-Lehmann estimators. The numerical studies reported by Fried and Dehling [4] showed that the test statistics were robust to outliers and non-normality and efficient in detecting univariate two-sample location shift. Mathur [5] proposed a strictly nonparametric bivariate test constructed from an extended U statistic and concluded that the test statistic did not depend on the covariance structure of the underlying population and was more powerful than the existing tests.

In randomized controlled trials, effectiveness of a treatment effect can be defined by not a single or two, but multiple primary endpoints, and significance of the treatment effect is then determined by multivariate location shift between the two multivariate distributions of the treatment and control groups. In these clinical trials, multivariate hypothesis testing procedures are needed to detect a potential location shift between two samples that are drawn with a multivariate primary endpoint. The conventional univariate two-sample t test was extended to the multivariate setting by Hotelling [6] and the proposed test statistic was denominated Hotelling's T^2 statistic. The Hotelling's T^2 test inherits the limitations of univariate two-sample t test, because it is still not robust to multivariate outliers and not valid when the multivariate normality assumption is violated. This motivated the development of multivariate two-sample location tests. Hettmansperger and Oja [7] developed a multivariate sign test for detecting location deviation among multiple multivariate samples. Hettmansperger et al. [8] introduced affine invariant analogues of the two-sample Mann-Whitney-Wilcoxon rank sum test. Neuhaus and Zhu [9] proposed multivariate distribution-free permutation test statistics that were built upon projected univariate versions of multivariate data. Henze et al. [10] introduced a class of consistent tests, in which the test statistic is a weighted integral of the squared modulus of the difference of the empirical characteristic functions of one multivariate sample and another multivariate sample plus a location shift.

In this article, we extend the robust nonparametric test statistics proposed by Fried and Dehling [4] and Mathur [5] to the multivariate setting. A series of robust multivariate nonparametric tests are proposed using the component-wise medians, Hodges-Lehmann location estimators, and an extended U statistic. Univariate test statistics for detecting multivariate two-sample location shift are constructed for the robust multivariate nonparametric tests as (i) unscaled maximum of the component-wise medians or the Hodges-Lehmann estimators, (ii) scaled maximum of the component-wise medians or the Hodges-Lehmann estimators, (iii) maximum of the scaled component-wise medians or the Hodges-Lehmann estimators, or (iv) the extended U statistic. A bootstrap approach and a permutation approach are introduced for determining the p -values of the proposed test statistics. We conduct simulation studies to examine performance of the proposed robust nonparametric test statistics in detecting multivariate two-sample location shift. The numerical results given by the bootstrap procedure demonstrate that the proposed robust multivariate nonparametric tests constructed from the Hodges-Lehmann estimators are more efficient than those based on medians and the extended

U statistic. The permutation approach can provide a more stringent control of Type I error and is more powerful than the bootstrap procedure. To demonstrate the use of these proposed robust multivariate nonparametric tests, the proposed hypothesis tests are applied to detecting the intervention effect of the Thai Healthy Choices study [11], a study that promotes a four-session motivational interviewing-based intervention to reduce risk behaviors among youth living with HIV (the Thai Healthy Choices study was designed jointly by Wayne State University and the Thai Red Cross AIDS Research Center, and implemented in Bangkok, Thailand).

The scientific contribution of this article is multifold. First, a series of new robust nonparametric test statistics are proposed for detecting location shift between two multivariate samples collected from the treatment and control groups, respectively, in clinical trials. Both a bootstrap approach and a permutation approach are introduced to implement the proposed tests for seeking corresponding p -values. These provide the practitioners a variety of choices with two distinct implementation approaches to test on treatment effects when the normality assumption for the samples is violated. Second, comprehensive numerical studies are conducted and the results show explicit benefits from using the proposed tests over the Hotelling's T^2 and the extended U tests in term of controlling Type I error and boosting statistical power. Third, the article presents a representative example, the Thai Healthy Choices study, and shows how the proposed robust nonparametric hypothesis testing procedures can be implemented to test on the treatment or intervention effect in a clinical trial.

Tests on two-sample location shift

Consider two random samples $\{X_1, \dots, X_m\}$ and $\{Y_1, \dots, Y_n\}$ with p -dimensional independent multivariate observations $X_i \in \mathbb{R}^p, i = 1, \dots, m$, and $Y_j \in \mathbb{R}^p, j = 1, \dots, n$, for which

$$X_1, \dots, X_m \stackrel{i.i.d.}{\sim} F(x)$$

and

$$Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} G(x).$$

The null hypothesis of equality of $F(x)$ and $G(x)$ and its alternative hypothesis that there is a location shift in the two multivariate distributions are

$$H_0 : F(x) = G(x) \quad \text{versus} \quad H_1 : F(x) = G(x + \Delta). \tag{1}$$

A natural idea to test the above hypotheses is to compare location estimators of the two distributions. Sample means \bar{X} and \bar{Y} can be used to fulfil this mission, which leads to the prominent Hotelling's T^2 test. However, the Hotelling's T^2 test is constructed under the multivariate normal distributions, and therefore perform poorly when there are outliers or the underlying true distributions of the two samples do not follow multivariate normal distributions.

Tests based on unscaled median difference and Hodges-Lehmann estimators

Here, we propose a series of robust nonparametric test statistics based on robust estimators for distribution locations as competitors of the Hotelling's T^2 test statistics. A general approach to construct such nonparametric tests is to estimate the location difference Δ and then reject the null hypothesis H_0 if Δ is far from zero. As usual, we can replace difference of sample means with difference between two sample medians:

$$\hat{\Delta}_1 = \text{med}\{Y_1, \dots, Y_n\} - \text{med}\{X_1, \dots, X_m\}. \tag{2}$$

In (2), $\text{med}\{X_1, \dots, X_m\}$ and $\text{med}\{Y_1, \dots, Y_n\}$ are the p -dimensional median vectors of the two samples. The median vector of a sample is defined as the vector of component-wise medians. That is, the k th component of $\text{med}\{X_1, \dots, X_m\}$ is the median of X_{1k}, \dots, X_{mk} , where X_{ik} is the k th component of p -dimensional observation X_i for $i = 1, \dots, m$, and the k th component of $\text{med}\{Y_1, \dots, Y_n\}$ is the median of Y_{1k}, \dots, Y_{nk} , where Y_{jk} is the k th component of p -dimensional observation Y_j for $j = 1, \dots, n$. In practice, however, $\hat{\Delta}_1$ cannot be directly used as a test statistic for the pair of hypotheses in (1), since $\hat{\Delta}_1$ is a p -dimensional vector and is not a scalar. Therefore, the following maximum of absolute values of the p medians within $\hat{\Delta}_1$ can be considered:

$$\hat{\Delta}_1^{\max} = \max \{|\hat{\Delta}_{11}|, |\hat{\Delta}_{12}|, \dots, |\hat{\Delta}_{1p}|\},$$

where $\hat{\Delta}_{1k}$ is the k th component of $\hat{\Delta}_1$ for $k = 1, \dots, p$. Under the null hypothesis $\hat{\Delta}_1^{\max}$ should be close to zero, whereas under alternative hypothesis it deviates from zero.

Noted that, although the direct sample medians $\text{med}\{X_1, \dots, X_m\}$ and $\text{med}\{Y_1, \dots, Y_n\}$ in (2) are robust estimators for the locations of two samples, these medians are not very efficient as each of them exploits little information in the sample data. To balance efficiency against robustness, two types of Hodges-Lehmann estimators were developed [12, 13]. Multivariate analogs of the univariate Hodges-Lehmann estimators are

$$\hat{\Delta}_2 = \text{med}\{Y_j - X_i; i = 1, \dots, m, j = 1, \dots, n\}$$

and

$$\hat{\Delta}_3 = \frac{\text{med}\{Y_i + Y_j; 1 \leq i < j \leq n\}}{2} - \frac{\text{med}\{X_i + X_j; 1 \leq i < j \leq m\}}{2},$$

where the p -dimensional multivariate median vectors are likewise defined as in (2). The test statistics using the multivariate Hodges-Lehmann estimators to detect location shift between two multivariate samples are then proposed as the absolute component-wise maximum of $\hat{\Delta}_2$ and $\hat{\Delta}_3$:

$$\hat{\Delta}_2^{\max} = \max \{|\hat{\Delta}_{21}|, |\hat{\Delta}_{22}|, \dots, |\hat{\Delta}_{2p}|\}$$

and

$$\hat{\Delta}_3^{\max} = \max \{|\hat{\Delta}_{31}|, |\hat{\Delta}_{32}|, \dots, |\hat{\Delta}_{3p}|\}.$$

where $\hat{\Delta}_{2k}$ and $\hat{\Delta}_{3k}$ are the k th component of $\hat{\Delta}_2$ and $\hat{\Delta}_3$, respectively, for $k = 1, \dots, p$.

Tests based on scaled median difference and Hodges-Lehmann estimators

Because $\hat{\Delta}_1^{\max}$, $\hat{\Delta}_2^{\max}$ and $\hat{\Delta}_3^{\max}$ only measure the component-wise maximum variability between the two samples, a scaled version of each is required to construct a more robust nonparametric test statistic. To this end, a related measure of the variability within the two samples are needed for the procedure of standardization. For $\hat{\Delta}_1^{\max}$, the following p -dimensional median vector is the measure of component-wise differences between the two samples:

$$S_1 = 2 \text{med}\{Z_i; i = 1, \dots, m + n\},$$

where $(Z_1, \dots, Z_{m+n})' = (X_1 - \tilde{X}, \dots, X_m - \tilde{X}, Y_1 - \tilde{Y}, \dots, Y_n - \tilde{Y})'$ is the joint median-corrected sample and $\tilde{X} = \text{med}\{X_1, \dots, X_m\}$ and $\tilde{Y} = \text{med}\{Y_1, \dots, Y_n\}$. Then, the absolute

component-wise maximum of S_1 is defined as

$$S_1^{\max} = 2 \max \{ \text{med}\{|Z_i|; i = 1, \dots, m + n\} \},$$

and the standardized version of $\hat{\Delta}_1^{\max}$ can be formulated as

$$T_1 = \hat{\Delta}_1^{\max} / S_1^{\max}.$$

For $\hat{\Delta}_2^{\max}$ and $\hat{\Delta}_3^{\max}$, the following p -dimensional median vectors of the absolute set of differences in the samples and within the joint median-corrected sample can be taken as the measure of component-wise differences within the two samples:

$$S_2 = \text{med}\{|X_i - X_j|, 1 \leq i < j \leq m; |Y_i - Y_j|, 1 \leq i < j \leq n\}$$

and

$$S_3 = \text{med}\{|Z_i - Z_j|; 1 \leq i < j \leq m + n\}.$$

Then, the absolute component-wise maximum of them are defined as

$$S_2^{\max} = \max \{ \text{med}\{|X_i - X_j|, 1 \leq i < j \leq m; |Y_i - Y_j|, 1 \leq i < j \leq n\} \}$$

and

$$S_3^{\max} = \max \{ \text{med}\{|Z_i - Z_j|, 1 \leq i < j \leq m + n\} \},$$

respectively. The scaled versions of $\hat{\Delta}_2^{\max}$ and $\hat{\Delta}_3^{\max}$ are consequently constructed as

$$T_2 = \hat{\Delta}_2^{\max} / S_2^{\max} \quad \text{and} \quad T_3 = \hat{\Delta}_3^{\max} / S_3^{\max},$$

and

$$T_4 = \hat{\Delta}_3^{\max} / S_2^{\max} \quad \text{and} \quad T_5 = \hat{\Delta}_3^{\max} / S_3^{\max}.$$

Alternative standardization procedure of $\hat{\Delta}_1^{\max}$, $\hat{\Delta}_2^{\max}$, and $\hat{\Delta}_3^{\max}$ can also be applied, which is to standardize each component of them and then take the maximum of all standardized components. This alternative standardization procedure leads to the following test statistics for detecting a multivariate two-sample location shift:

$$T_1^* = \max \left\{ \frac{|\hat{\Delta}_{11}|}{S_{11}}, \frac{|\hat{\Delta}_{12}|}{S_{12}}, \dots, \frac{|\hat{\Delta}_{1p}|}{S_{1p}} \right\},$$

$$T_2^* = \max \left\{ \frac{|\hat{\Delta}_{21}|}{S_{21}}, \frac{|\hat{\Delta}_{22}|}{S_{22}}, \dots, \frac{|\hat{\Delta}_{2p}|}{S_{2p}} \right\},$$

$$T_3^* = \max \left\{ \frac{|\hat{\Delta}_{31}|}{S_{31}}, \frac{|\hat{\Delta}_{32}|}{S_{32}}, \dots, \frac{|\hat{\Delta}_{3p}|}{S_{3p}} \right\},$$

$$T_4^* = \max \left\{ \frac{|\hat{\Delta}_{31}|}{S_{21}}, \frac{|\hat{\Delta}_{32}|}{S_{22}}, \dots, \frac{|\hat{\Delta}_{3p}|}{S_{2p}} \right\},$$

and

$$T_5^* = \max \left\{ \frac{|\hat{\Delta}_{31}|}{S_{31}}, \frac{|\hat{\Delta}_{32}|}{S_{32}}, \dots, \frac{|\hat{\Delta}_{3p}|}{S_{3p}} \right\},$$

for $l = 1, 2, \dots, 5$, in which S_{lk} denotes the k th element of S_l , $l = 1, 2, 3$. Under the null hypothesis, the proposed test statistics T_l and T_l^* , $l = 1, 2, \dots, 5$, should be close to zero, whereas they should be far from zero under the alternative hypothesis. When the dimension p is equal to 1, these test statistics degenerate to the test statistics introduced by Fried and Dehling [4] and $T_l = T_l^*$ for $l = 1, 2, \dots, 5$.

Tests based on U statistics

An U -statistic recently proposed by Mathur [5] was designated to test a bivariate two-sample location shift. Here, we extend it to serve for detecting the multivariate two-sample location shift. Specifically, the extended U test statistic for multivariate sample location detection is defined as

$$U = \sum_{i=1}^n \sum_{j=1}^m I(D_{1i}^2 > D_{2j}^2) \tag{3}$$

where $D_{1i}^2 = \|X_i\|^2$ is the Euclidean distance from $\{X_1, \dots, X_m\}$ to origin and $D_{2j}^2 = \|Y_j\|^2$ is the Euclidean distance from $\{Y_1, \dots, Y_n\}$ to origin. The null hypothesis is rejected if the observed value of the extended U statistic exceeds a critical value of U obtained by permutation.

Implementation: A bootstrap procedure

Here, a bootstrap procedure is introduced to numerically approximate the p -values of the proposed robust nonparametric tests. Suppose two random samples with p -dimensional independent multivariate observations $X_1, \dots, X_m \stackrel{i.i.d}{\sim} F(x)$ and p -dimensional independent multivariate observations $Y_1, \dots, Y_n \stackrel{i.i.d}{\sim} G(x)$ are collected. The null hypothesis $H_0: F(x) = G(x)$ and its alternative hypothesis for a location shift in the two multivariate distributions $H_1: F(x) = G(x + \Delta)$ are considered. To conduct hypothesis testing on such a pair of hypotheses, distributions of the above proposed test statistics are generally unknown in finite samples. Therefore, a bootstrap method can be adopted to approximate the underlying distribution of a test statistic and subsequently determine the corresponding p -value. In the bootstrap procedure, a pseudo sample $\{X_i^*, i = 1, \dots, m\}$ is drawn from the pooled sample $\{X_i, Y_j; i = 1, \dots, m, j = 1, \dots, n\}$ with replacement, and another pseudo sample $\{Y_j^*, j = 1, \dots, n\}$ is drawn from the same pooled sample also with replacement. Let V denote any one of the investigated test statistics, and let V^* be the bootstrap version of V that is calculated from the paired bootstrap pseudo samples $\{X_i^*, i = 1, \dots, m\}$ and $\{Y_j^*, j = 1, \dots, n\}$. Then, the null hypothesis is rejected if V is larger than the $(100\% \cdot \alpha)$ quantile of the bootstrap distribution of V^* , where α is the level of significance of hypothesis testing. It has been confirmed in the literature that the above bootstrap procedure can produce a valid approximation to the test statistic V [14–16].

Implementation: A permutation procedure

A permutation procedure is a competitive alternative to the bootstrap procedure to derive critical values for the proposed robust nonparametric tests. In the permutation procedure, the pooled sample $\{X_i, Y_j; i = 1, \dots, m, j = 1, \dots, n\}$ is repeatedly split to two pseudo samples

$\{X_i^*, i = 1, \dots, m\}$ and $\{Y_j^*, j = 1, \dots, n\}$ without replacement. Let V^* denoted the permuted version of a investigated test statistic V , and V^* is calculated from the paired permuted pseudo samples $\{X_i^*, i = 1, \dots, m\}$ and $\{Y_j^*, j = 1, \dots, n\}$. Then, the null hypothesis is rejected if V is larger than the $(100\% \cdot \alpha)$ quantile of the permutation distribution of V^* , where α is the level of significance of hypothesis testing.

Simulation studies

This section reports numerical results from a simulation study that was conducted to demonstrate merits of the proposed hypothesis tests and compare them with Hotelling's T^2 . In this simulation study, we aim at examining and comparing performance of the proposed hypothesis tests for detecting a location shift among different pairs of two samples. The sample $\{X_1, \dots, X_m\}$ was generated from $F(x)$ and the sample $\{Y_1, \dots, Y_n\}$ was generated from $G(x)$. Four different pairs of distributions of $F(x)$ and $G(x)$ were considered: (i) $F(x)$ was a p -dimensional multivariate normal distribution $N_p(1_p, \Sigma_p)$, where 1_p is a p -dimensional vector with each component equal to one and Σ_p is the variance-covariance matrix, and $G(x)$ was the location shift distribution $N_p(1_p + \Delta, \Sigma_p)$, (ii) $F(x)$ was a p -dimensional multivariate t distribution with 1 degrees of freedom $t_1(1_p, \Sigma_p)$, and $G(x)$ was the location shift distribution $t_1(1_p + \Delta, \Sigma_p)$ (iii) $F(x)$ was a p -dimensional multivariate t distribution with 3 degrees of freedom $t_3(1_p, \Sigma_p)$, and $G(x)$ was the location shift distribution $t_3(1_p + \Delta, \Sigma_p)$, and (iv) $F(x)$ was the p -dimensional joint distribution of the diagonal elements of a Wishart random matrix that followed the Wishart distribution $W_p(3, \Sigma_p)$, where 3 is the degree of freedom and Σ_p is the scale matrix, and $G(x) = F(x + \Delta)$ was the location shift distribution. In this simulation study, two variance-covariance matrices were taken to generated the simulation data: one is an independent variance-covariance matrix $I_{p \times p}$, which is a $p \times p$ identify matrix, and another one is a non-independent variance-covariance matrix with diagonal elements equal to 1 and non-diagonal elements equal to 0.5. The dimension p of the two samples was set to be 4, and the sample sizes were set as $n = m = 10, 25, \text{ and } 50$. In this simulation study, the location vector Δ in each of four distributions was specified as $\Delta = (0.5\delta, \delta, \delta, 2\delta)'$, in which δ varied to take a value of 0, 0.5, 1, 1.5, or 2. The proposed test statistics $\hat{\Delta}_1^{\max}, \hat{\Delta}_2^{\max}, \hat{\Delta}_3^{\max}, T_l$ and $T_l^*, l = 1, 2, \dots, 5$, as well as Hotelling's T^2 and the extended U statistic, were applied in two-sample multivariate hypothesis testing to detect the location shift. A total number of 1000 simulation data sets were generated from each pair of specified distributions of the two samples, and then the proposed hypothesis testing was implemented using these investigated test statistics. The rejection rate was subsequently calculated as the frequency that the null hypothesis $H_0: F(x) = G(x)$ was rejected among the 1000 simulation data sets by each of the investigated hypothesis test statistics. When $\delta = 0$, the pair of true distributions of the two samples have the identical location, and thus the rejection rate is corresponding to simulated Type I error of the hypothesis tests. When $\delta \neq 0$, the pair of true distributions of the two samples reside in different locations, and the rejection rate is corresponding to simulated power of the hypothesis tests. In this simulation study, the number of bootstrap samples was set to be 500 and the significance level was set to be 0.05.

The simulation results of the test statistics based on the bootstrap procedure are presented in S1–S6 Tables. S1–S3 Tables report the Type I errors and power obtained from the simulated paired samples that were generated using the independent variance-covariance matrix with different sample sizes. S4–S6 Tables report the Type I errors and power obtained using the non-independent variance-covariance matrix. It is observed that, when the samples were generated from two multivariate normal distributions with a location shift, Hotelling's T^2 , extended U statistic, and T_l and $T_l^*, l = 2, 3$, performed the best among all the investigated test

statistics in term of Type I errors and power as δ varied. There was not sufficient numerical evidence that the Hotelling's T^2 statistic outperformed other five statistics. The tests based on the Hodges-Lehmann estimators $\hat{\Delta}_2^{\max}$, $\hat{\Delta}_3^{\max}$, T_l and T_l^* , $l = 2, 3, 4, 5$, were more powerful than those based on medians $\hat{\Delta}_1^{\max}$, T_1 and T_1^* . The choice of the measure of variability within the two samples (i.e., the choice of either S_2 or S_3 and the choice of either S_2^{\max} or S_3^{\max}) had very little impact on the performance of test statistics.

When one sample was generated from a multivariate t distribution or a Wishart distribution and another sample was generated from its location shift counterpart, the performance of the proposed robust nonparametric test statistics outperformed the Hotelling's T^2 and extended U statistics in detecting the location shift between the two samples. The power of these robust test statistics was mostly larger than the power given by the Hotelling's T^2 and extended U statistics. Among the nonparametric test statistics, as in the case of multivariate normal distributions, the tests based on the Hodges-Lehmann estimators were more powerful than those based on the medians. The scaled nonparametric tests generally outperform their unscaled counterparts. The nonparametric tests based on T_l and T_l^* , $l = 2, 3$, are most powerful among the investigated test statistics, and the Type I errors given by these four test statistics are mostly close to 0.05. The powers given by the investigated test statistics consistently increased as the location difference between two samples and sample sizes were enlarged.

The simulation results of the test statistics based on the permutation approach are presented in Tables 1–6. Tables 1–3 report the Type I errors and power obtained from the simulated paired samples that were generated using the independent variance-covariance matrix with different sample sizes. Tables 4–6 report the Type I errors and power obtained using the non-

Table 1. Type I errors ($\delta = 0$) and power ($\delta \neq 0$) given by the investigated test statistics based on permutation approach in detecting location shift between two samples generated from the four pairs of $F(x)$ and $G(x)$ with variance-covariance matrix $I_{4 \times 4}$ and sample sizes $n = m = 10$.

$G(x)$	δ value	Type I errors ($\delta = 0$) and power ($\delta \neq 0$)														
		T^2	$\hat{\Delta}_1^{\max}$	$\hat{\Delta}_2^{\max}$	$\hat{\Delta}_3^{\max}$	T_1	T_2	T_3	T_4	T_5	T_1^*	T_2^*	T_3^*	T_4^*	T_5^*	U
(i)	0	0.040	0.054	0.061	0.058	0.053	0.059	0.057	0.062	0.067	0.056	0.059	0.057	0.058	0.050	0.063
	0.50	0.411	0.353	0.437	0.443	0.334	0.418	0.421	0.433	0.430	0.250	0.358	0.361	0.370	0.371	0.576
	1.00	0.970	0.932	0.977	0.979	0.895	0.977	0.977	0.976	0.977	0.797	0.940	0.941	0.940	0.939	0.993
	1.50	1.000	0.998	1.000	1.000	0.995	1.000	1.000	1.000	1.000	0.994	1.000	1.000	1.000	1.000	1.000
	2.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(ii)	0	0.018	0.038	0.048	0.048	0.045	0.048	0.050	0.045	0.052	0.038	0.053	0.050	0.049	0.053	0.054
	0.50	0.044	0.118	0.098	0.092	0.136	0.105	0.110	0.094	0.101	0.163	0.143	0.152	0.116	0.125	0.053
	1.00	0.115	0.367	0.286	0.230	0.464	0.325	0.330	0.271	0.273	0.523	0.478	0.501	0.391	0.422	0.087
	1.50	0.233	0.643	0.561	0.438	0.768	0.623	0.617	0.533	0.528	0.813	0.809	0.804	0.717	0.731	0.165
	2.00	0.398	0.802	0.764	0.653	0.922	0.844	0.834	0.770	0.758	0.930	0.938	0.942	0.860	0.881	0.258
(iii)	0	0.026	0.049	0.052	0.054	0.049	0.059	0.056	0.061	0.059	0.048	0.053	0.056	0.060	0.058	0.053
	0.50	0.174	0.241	0.262	0.249	0.240	0.269	0.266	0.264	0.265	0.247	0.273	0.286	0.283	0.279	0.083
	1.00	0.720	0.747	0.810	0.798	0.751	0.818	0.814	0.827	0.813	0.690	0.811	0.823	0.804	0.811	0.364
	1.50	0.945	0.965	0.985	0.979	0.968	0.988	0.988	0.983	0.985	0.953	0.982	0.983	0.984	0.985	0.810
	2.00	0.993	0.997	1.000	1.000	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.967
(iv)	0	0.023	0.055	0.059	0.047	0.049	0.056	0.063	0.051	0.046	0.047	0.039	0.047	0.041	0.053	0.064
	0.50	0.084	0.101	0.113	0.116	0.087	0.118	0.120	0.118	0.133	0.099	0.114	0.120	0.113	0.117	0.130
	1.00	0.306	0.300	0.334	0.327	0.309	0.351	0.364	0.343	0.360	0.274	0.341	0.379	0.347	0.376	0.345
	1.50	0.652	0.661	0.729	0.667	0.654	0.753	0.759	0.701	0.714	0.589	0.718	0.766	0.689	0.720	0.612
	2.00	0.905	0.904	0.946	0.907	0.897	0.954	0.959	0.933	0.938	0.859	0.966	0.963	0.934	0.939	0.865

<https://doi.org/10.1371/journal.pone.0195894.t001>

Table 2. Type I errors ($\delta = 0$) and power ($\delta \neq 0$) given by the investigated test statistics based on permutation approach in detecting location shift between two samples generated from the four pairs of $F(x)$ and $G(x)$ with variance-covariance matrix $I_{4 \times 4}$ and sample sizes $n = m = 20$.

$G(x)$	δ value	Type I errors ($\delta = 0$) and power ($\delta \neq 0$)														
		T^2	$\hat{\Delta}_1^{\max}$	$\hat{\Delta}_2^{\max}$	$\hat{\Delta}_3^{\max}$	T_1	T_2	T_3	T_4	T_5	T_1^*	T_2^*	T_3^*	T_4^*	T_5^*	U
(i)	0	0.037	0.048	0.047	0.052	0.045	0.049	0.049	0.050	0.052	0.045	0.044	0.040	0.051	0.047	0.049
	0.50	0.847	0.660	0.765	0.770	0.611	0.762	0.768	0.770	0.770	0.557	0.755	0.753	0.744	0.751	0.863
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.983	1.000	1.000	1.000	1.000	1.000
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(ii)	0	0.016	0.053	0.067	0.060	0.052	0.062	0.064	0.065	0.064	0.047	0.052	0.054	0.045	0.046	0.049
	0.50	0.032	0.287	0.211	0.179	0.299	0.227	0.223	0.194	0.190	0.365	0.272	0.293	0.235	0.251	0.061
	1.00	0.141	0.829	0.703	0.628	0.867	0.737	0.728	0.674	0.672	0.902	0.816	0.823	0.788	0.793	0.107
	1.50	0.304	0.987	0.960	0.939	0.994	0.974	0.975	0.958	0.958	0.993	0.987	0.990	0.976	0.982	0.267
	2.00	0.485	1.000	0.993	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	1.000	0.990	0.459
(iii)	0	0.040	0.051	0.051	0.044	0.053	0.052	0.049	0.046	0.041	0.043	0.043	0.046	0.044	0.047	0.048
	0.50	0.432	0.523	0.544	0.545	0.496	0.550	0.548	0.545	0.547	0.494	0.587	0.595	0.579	0.580	0.108
	1.00	0.954	0.989	0.987	0.987	0.976	0.989	0.990	0.990	0.979	0.989	0.991	0.990	0.987	0.991	0.662
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.984
	2.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(iv)	0	0.039	0.050	0.056	0.052	0.053	0.055	0.056	0.050	0.050	0.045	0.059	0.053	0.053	0.056	0.070
	0.50	0.179	0.171	0.214	0.181	0.162	0.226	0.232	0.190	0.199	0.148	0.221	0.240	0.195	0.228	0.201
	1.00	0.659	0.606	0.775	0.651	0.576	0.784	0.791	0.684	0.683	0.544	0.764	0.795	0.658	0.701	0.623
	1.50	0.957	0.929	0.984	0.954	0.930	0.989	0.991	0.965	0.968	0.915	0.989	0.991	0.965	0.975	0.912
	2.00	0.998	0.999	0.999	0.995	0.989	1.000	1.000	0.998	0.999	0.984	0.997	0.999	0.996	0.998	0.992

<https://doi.org/10.1371/journal.pone.0195894.t002>

Table 3. Type I errors ($\delta = 0$) and power ($\delta \neq 0$) given by the investigated test statistics based on permutation approach in detecting location shift between two samples generated from the four pairs of $F(x)$ and $G(x)$ with variance-covariance matrix $I_{4 \times 4}$ and sample sizes $n = m = 50$.

$G(x)$	δ value	Type I errors ($\delta = 0$) and power ($\delta \neq 0$)														
		T^2	$\hat{\Delta}_1^{\max}$	$\hat{\Delta}_2^{\max}$	$\hat{\Delta}_3^{\max}$	T_1	T_2	T_3	T_4	T_5	T_1^*	T_2^*	T_3^*	T_4^*	T_5^*	U
(i)	0	0.047	0.054	0.045	0.046	0.055	0.046	0.046	0.046	0.045	0.044	0.048	0.041	0.042	0.045	0.057
	0.50	1.000	0.974	0.999	0.997	0.970	0.999	0.998	0.998	0.999	0.950	0.999	0.998	0.999	0.999	0.999
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(ii)	0	0.012	0.051	0.055	0.054	0.049	0.054	0.056	0.055	0.054	0.045	0.054	0.057	0.054	0.050	0.055
	0.50	0.043	0.734	0.589	0.565	0.750	0.602	0.601	0.580	0.571	0.732	0.605	0.604	0.578	0.580	0.126
	1.00	0.171	1.000	0.988	0.983	1.000	0.999	0.999	0.997	0.997	0.999	0.998	0.994	0.997	0.997	0.569
	1.50	0.345	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.926
	2.00	0.531	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996
(iii)	0	0.046	0.052	0.048	0.043	0.052	0.055	0.058	0.052	0.053	0.49	0.048	0.048	0.046	0.045	0.052
	0.50	0.854	0.905	0.921	0.926	0.908	0.923	0.924	0.926	0.927	0.887	0.898	0.919	0.916	0.921	0.226
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	2.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
(iv)	0	0.042	0.047	0.053	0.051	0.052	0.049	0.052	0.051	0.054	0.053	0.047	0.047	0.051	0.050	0.055
	0.50	0.471	0.392	0.591	0.440	0.378	0.589	0.600	0.448	0.461	0.364	0.601	0.620	0.429	0.458	0.376
	1.00	0.982	0.954	1.000	0.967	0.926	1.000	1.000	0.971	0.975	0.930	0.999	0.999	0.977	0.980	0.902
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	
	2.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	

<https://doi.org/10.1371/journal.pone.0195894.t003>

Table 4. Type I errors ($\delta = 0$) and power ($\delta \neq 0$) given by the investigated test statistics based on permutation approach in detecting location shift between two samples generated from the four pairs of $F(x)$ and $G(x)$ with the non-independent variance-covariance matrix and sample sizes $n = m = 10$.

		Type I errors ($\delta = 0$) and power ($\delta \neq 0$)														
$G(x)$	δ value	T^2	$\hat{\Delta}_1^{\max}$	$\hat{\Delta}_2^{\max}$	$\hat{\Delta}_3^{\max}$	T_1	T_2	T_3	T_4	T_5	T_1^*	T_2^*	T_3^*	T_4^*	T_5^*	U
(i)	0	0.034	0.056	0.052	0.044	0.052	0.044	0.052	0.038	0.048	0.048	0.046	0.056	0.040	0.050	0.064
	0.50	0.358	0.294	0.364	0.356	0.290	0.332	0.324	0.332	0.328	0.240	0.296	0.298	0.304	0.316	0.178
	1.00	0.974	0.906	0.962	0.964	0.874	0.946	0.944	0.952	0.948	0.752	0.924	0.922	0.922	0.928	0.632
	1.50	1.000	0.995	1.000	0.979	1.000	1.000	1.000	1.000	1.000	0.823	1.000	1.000	1.000	1.000	0.958
	2.00	1.000	0.995	1.000	1.000	0.998	1.000	1.000	1.000	1.000	0.966	1.000	1.000	1.000	1.000	1.000
(ii)	0	0.014	0.044	0.038	0.048	0.044	0.040	0.044	0.044	0.054	0.046	0.044	0.050	0.056	0.040	0.052
	0.5	0.050	0.136	0.104	0.112	0.162	0.112	0.112	0.110	0.114	0.134	0.112	0.134	0.102	0.110	0.072
	1.0	0.252	0.384	0.318	0.276	0.422	0.364	0.384	0.324	0.348	0.452	0.388	0.382	0.346	0.350	0.156
	1.5	0.530	0.742	0.690	0.578	0.824	0.734	0.722	0.670	0.652	0.762	0.744	0.762	0.614	0.652	0.310
	2.0	0.638	0.844	0.814	0.732	0.898	0.862	0.870	0.828	0.804	0.916	0.856	0.870	0.7720	0.788	0.456
(iii)	0	0.024	0.064	0.044	0.040	0.058	0.044	0.044	0.044	0.042	0.046	0.048	0.046	0.042	0.046	0.060
	0.5	0.212	0.252	0.282	0.268	0.238	0.266	0.272	0.264	0.270	0.196	0.218	0.236	0.246	0.254	0.082
	1.0	0.742	0.740	0.784	0.784	0.730	0.780	0.794	0.782	0.782	0.602	0.732	0.750	0.722	0.720	0.312
	1.5	0.958	0.972	0.986	0.984	0.952	0.986	0.986	0.986	0.982	0.900	0.970	0.970	0.966	0.968	0.674
	2.0	0.994	1.000	1.000	1.000	0.992	1.000	1.000	1.000	1.000	0.980	0.998	1.000	0.994	1.000	0.9000
(iv)	0	0.030	0.046	0.047	0.051	0.036	0.051	0.052	0.056	0.053	0.053	0.054	0.048	0.055	0.051	0.066
	0.5	0.069	0.080	0.095	0.096	0.085	0.102	0.102	0.086	0.099	0.085	0.092	0.112	0.096	0.112	0.090
	1.0	0.226	0.262	0.299	0.295	0.251	0.310	0.319	0.311	0.308	0.199	0.293	0.321	0.2780	0.299	0.174
	1.5	0.491	0.583	0.711	0.617	0.582	0.719	0.725	0.651	0.661	0.427	0.625	0.681	0.5470	0.598	0.307
	2.0	0.740	0.851	0.936	0.865	0.835	0.933	0.931	0.885	0.881	0.680	0.870	0.897	0.789	0.850	0.476

<https://doi.org/10.1371/journal.pone.0195894.t004>

Table 5. Type I errors ($\delta = 0$) and power ($\delta \neq 0$) given by the investigated test statistics based on permutation approach in detecting location shift between two samples generated from the four pairs of $F(x)$ and $G(x)$ with the non-independent variance-covariance matrix and sample sizes $n = m = 20$.

		Type I errors ($\delta = 0$) and power ($\delta \neq 0$)														
$G(x)$	δ value	T^2	$\hat{\Delta}_1^{\max}$	$\hat{\Delta}_2^{\max}$	$\hat{\Delta}_3^{\max}$	T_1	T_2	T_3	T_4	T_5	T_1^*	T_2^*	T_3^*	T_4^*	T_5^*	U
(i)	0	0.033	0.043	0.053	0.059	0.047	0.052	0.058	0.064	0.057	0.046	0.057	0.052	0.062	0.059	0.056
	0.50	0.687	0.629	0.732	0.738	0.608	0.736	0.730	0.731	0.734	0.544	0.703	0.715	0.706	0.715	0.587
	1.00	1.000	0.988	1.000	1.000	0.992	1.000	1.000	1.000	1.000	0.974	1.000	1.000	1.000	1.000	0.994
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(ii)	0	0.011	0.054	0.063	0.065	0.055	0.061	0.060	0.062	0.058	0.063	0.059	0.054	0.063	0.059	0.042
	0.50	0.067	0.291	0.212	0.187	0.320	0.226	0.226	0.208	0.212	0.343	0.244	0.248	0.226	0.232	0.076
	1.00	0.243	0.809	0.693	0.641	0.845	0.736	0.731	0.681	0.681	0.853	0.737	0.737	0.694	0.692	0.273
	1.50	0.446	0.982	0.951	0.918	0.995	0.961	0.955	0.947	0.942	0.991	0.966	0.968	0.947	0.953	0.589
	2.00	0.623	0.998	0.998	0.999	0.998	0.995	0.995	0.994	0.997	0.999	0.997	0.998	0.995	0.997	0.813
(iii)	0	0.028	0.047	0.050	0.057	0.052	0.050	0.051	0.049	0.040	0.045	0.049	0.043	0.054	0.055	0.064
	0.50	0.377	0.515	0.554	0.536	0.516	0.550	0.549	0.544	0.536	0.486	0.547	0.556	0.537	0.542	0.101
	1.00	0.915	0.980	0.988	0.981	0.975	0.992	0.992	0.986	0.988	0.985	0.983	0.984	0.980	0.981	0.658
	1.50	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	1.000	1.000	1.000	1.000	0.980
	2.00	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(iv)	0	0.040	0.038	0.051	0.054	0.040	0.050	0.055	0.052	0.039	0.054	0.044	0.051	0.056	0.047	0.054
	0.50	0.155	0.150	0.209	0.199	0.167	0.219	0.221	0.198	0.208	0.125	0.209	0.222	0.185	0.211	0.178
	1.00	0.529	0.612	0.771	0.632	0.606	0.788	0.798	0.656	0.682	0.556	0.761	0.791	0.659	0.699	0.573
	1.50	0.874	0.927	0.976	0.925	0.918	0.981	0.983	0.950	0.949	0.911	0.978	0.980	0.934	0.943	0.869
	2.00	0.985	0.994	1.000	0.995	0.998	1.000	1.000	0.998	0.996	0.995	1.000	1.000	0.996	0.995	0.981

<https://doi.org/10.1371/journal.pone.0195894.t005>

Table 6. Type I errors ($\delta = 0$) and power ($\delta \neq 0$) given by the investigated test statistics based on permutation approach in detecting location shift between two samples generated from the four pairs of $F(x)$ and $G(x)$ with the non-independent variance-covariance matrix and sample sizes $n = m = 50$.

$G(x)$	δ value	Type I errors ($\delta = 0$) and power ($\delta \neq 0$)														
		T^2	$\hat{\Delta}_1^{\max}$	$\hat{\Delta}_2^{\max}$	$\hat{\Delta}_3^{\max}$	T_1	T_2	T_3	T_4	T_5	T_1^*	T_2^*	T_3^*	T_4^*	T_5^*	U
(i)	0	0.036	0.046	0.051	0.053	0.056	0.053	0.053	0.048	0.048	0.052	0.055	0.055	0.047	0.049	0.055
	0.50	0.994	0.955	0.991	0.993	0.993	0.994	0.993	0.992	0.992	0.942	0.991	0.990	0.988	0.987	0.929
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(ii)	0	0.011	0.052	0.049	0.046	0.053	0.050	0.049	0.045	0.049	0.049	0.051	0.052	0.049	0.051	0.052
	0.50	0.059	0.734	0.589	0.565	0.752	0.602	0.601	0.583	0.572	0.732	0.603	0.602	0.578	0.580	0.126
	1.00	0.275	1.000	0.992	1.000	1.000	0.996	0.995	0.989	0.990	1.000	0.993	0.992	0.985	0.985	0.569
	1.50	0.500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.926
	2.00	0.658	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996
(iii)	0	0.037	0.063	0.052	0.053	0.062	0.061	0.050	0.052	0.053	0.057	0.057	0.058	0.057	0.054	0.053
	0.50	0.746	0.905	0.923	0.928	0.908	0.923	0.924	0.926	0.927	0.898	0.919	0.918	0.921	0.920	0.226
	1.00	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.967
	1.50	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(iv)	0	0.038	0.047	0.053	0.054	0.049	0.054	0.051	0.049	0.053	0.052	0.047	0.051	0.047	0.051	0.046
	0.50	0.369	0.392	0.591	0.440	0.378	0.592	0.600	0.448	0.461	0.364	0.601	0.620	0.429	0.458	0.374
	1.00	0.932	0.953	0.992	0.964	0.948	0.996	0.994	0.973	0.977	0.935	0.996	0.995	0.970	0.973	0.902
	1.50	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

<https://doi.org/10.1371/journal.pone.0195894.t006>

independent variance-covariance matrix. It is observed that, when the samples were generated from two multivariate normal distributions with a location shift, the performance of all the test statistics are comparable. Moreover, when one sample was generated from a multivariate t -distribution or a Wishart distribution and another sample was generated from its location shift counterpart, the performance of the proposed robust nonparametric test statistics outperformed the Hotelling's T^2 and extended U statistics in detecting the location shift between the two samples as it was shown by the bootstrap procedure. The tests based on the Hodges-Lehmann estimators $\hat{\Delta}_2^{\max}$, $\hat{\Delta}_3^{\max}$, T_l and T_l^* , $l = 2, 3, 4, 5$, were slightly powerful than those based on medians $\hat{\Delta}_1^{\max}$, T_1 and T_1^* . A cross comparison of the Type I errors and power given by the bootstrap approach and the permutation approach showed that the permutation approach was able to provide a more stringent control of Type I error and was generally more powerful than the bootstrap procedure. The performance of the nonparametric tests T_l and T_l^* , $l = 2, 3, 4, 5$ did not differ when the permutation approach is applied. Although in Tables 1–6 the scaled nonparametric test statistics cannot be distinguished from their unscaled counterparts, these results were not generalizable since Fried and Dehling [4] had explicitly demonstrated the advantages of the scaled nonparametric test statistics over the unscaled ones.

Naturally, the proposed nonparametric test statistics function properly without the multivariate normality assumption that the classical Hotelling's T^2 test requires and therefore are robust to non-normality and outliers. This is the primary reason that we observe in the simulation studies that the proposed tests were comparable to the the Hotelling's T^2 and the extended U tests when the two samples were simulated from multivariate normal distributions and they outperform the two tests when normality does not hold for the simulated samples.

Statistical analysis of the Thai Healthy Choices study

This section introduces the Thai Healthy Choices study and reports the analysis results from hypothesis testing on effect of a four-session motivational interviewing-based intervention developed in the study to reduce risk behaviors among youth living with HIV [11]. The proposed nonparametric robust test statistics T_l and T_l^* , $l = 1, 2, \dots, 5$, as well as Hotelling's T^2 and extended U statistic, were applied in the hypothesis testing procedure to detect multivariate difference between intervention and control groups in the study.

The Thai Healthy Choices study

The Thai Healthy Choices study was conducted at the Thai Red Cross AIDS Research Center in Bangkok [11]. Thai youth living with HIV and attending the Thai Red Cross AIDS Research Centre clinics in Bangkok, who were interested in participation in the study, were referred by their physicians to the study team. The participants eligible to enroll in the study are those between 16 and 25 years old, HIV-positive, and understanding spoken and written Thai enough to participate in study assessments and sessions. Upon completion of consent, participants were randomized in a one-to-one ratio to receive either an designed intervention approach named Healthy Choices (intervention group) or general health education (control group). At the baseline visit, participants completed the assessments. After the baseline visit, participants began to attend either four Healthy Choices sessions in intervention group or four general health education sessions in control group, based on randomization. The sessions in both groups occurred at 1, 2, 6 and 12 weeks after the baseline visit. Each session took approximately 60 min. The assessments similar to the baseline visit were conducted at 1 month after the fourth session and again at 6 months after the fourth session in both groups.

The intervention group received Healthy Choices, a four-session individual-level Motivational Interviewing (MI) counseling that targeted two of three possible risk behaviors, including sexual risks, alcohol use, and antiretroviral adherence. The intervention was delivered in Thai by an MI-trained interventionist. The details of the intervention have been published elsewhere [11]. Session 1 focused on eliciting the participants view of the behavior, exploring barriers as well as sociocultural factors affecting risks and building motivation to initiate the change plan. Session 2 followed a similar format with a focus on the second targeted behaviors. Sessions 3 and 4 were to formalize the personalized behavior change plan, reinforce commitment to change, and identify strategies to maintain healthy behaviors and to prevent relapse. All MI strategies to enhance motivation were used throughout all sessions. The control group received four individualized sessions of general health education unrelated to HIV risk behaviors. Session 1 focused on healthy diet, Session 2 on exercise, and Session 3 on smoking and healthy sleep habits. Session 4 was an overall review of the participants knowledge learned during the prior sessions. The contents of the sessions were adapted from the health education materials published by the Thai Ministry of Public Health. All sessions were delivered didactically by a research assistant who read the contents of the health education manual to the participant. The research assistant received no MI training and was instructed to avoid discussing HIV-related topics, including sexual behavior, HIV disclosure, alcohol and substance use, and medication adherence with the participant.

There were six primary clinical measures for the success of the investigated intervention. (1) HIV sexual risk score. An HIV sexual risk scoring system was empirically created based on eight sexual behavior characteristics: sexual intercourse, condom use, number of partners, HIV status of partners, anal sex, receptive anal sex, receptive vaginal sex, and alcohol use with sex. A score (ranging from 1 to 13) was calculated for each participant at each study visit based on the individuals sexual activities in the previous 30 days. The purpose of the scoring system

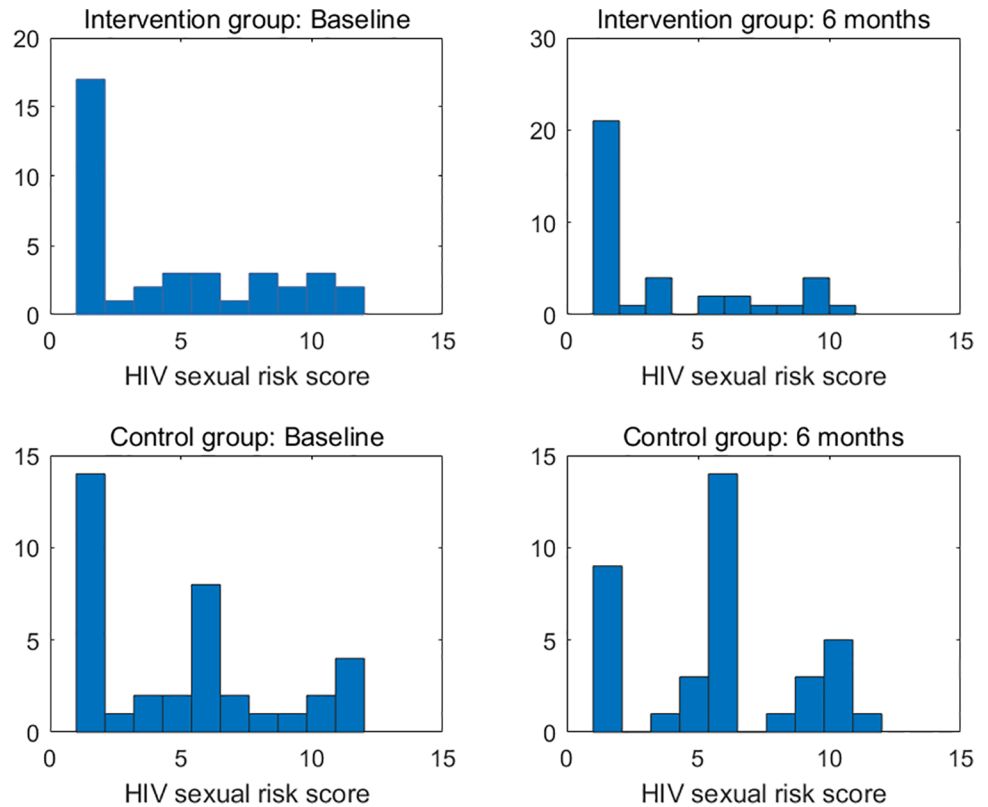


Fig 1. Histograms of HIV sexual risk scores for treatment and control groups at baseline and 6-month visits.

<https://doi.org/10.1371/journal.pone.0195894.g001>

was to provide a broad view of the quantifiable magnitude of an individuals sexual risk. (2) Viral load. Blood samples for plasma HIV viral loads were obtained at baseline, 1 month follow-up, and 6 months follow-up in both study groups and were analyzed by COBAS AmpliPrep/Amplior HIV-1 Monitor Test, version 1.5 (Roche Molecular Systems, Branchburg, NJ), with the lower limit of detection at 50 copies/ml. (3) HIV stigma. Participants completed the 12-item HIV Stigma Scale, which was developed from Berger’s 40-item HIV Stigma Scale [17]. The measure contains four stigma subscales, with three items per each subscale, representing personalized stigma, disclosure concerns, negative self-image, and public attitude stigma. Cronbach’s α was 0.80 in the present study. (4) Mental health. Participants completed the 12-item Thai General Health Questionnaire covering depression, anxiety, social impairment, and somatic complaints. All items were rated on a four-point Likert scale, ranging from 1 (not at all) to 4 (much more than usual). The scores were averaged and a mean score ≥ 2 was considered clinically significant. Cronbach’s α was 0.85 in the present study. (5) Self-efficacy on confidence in avoiding multiple sex partners, and (6) self-efficacy on confidence in using condoms. The Self-Efficacy for Health Promotion and Risk Reduction questionnaire contains 6 items on confidence in using a condom and 3 items on confidence in avoiding sex with multiple partners. Items were rated on a 5-point Likert scale ranging from 1 (very sure I cannot) to 5 (very sure I can). Cronbach’s α was 0.89 in the this study. Figs 1–3 display the histograms of HIV sexual risk scores, self-efficacy on avoiding multiple partners, and self-efficacy on condom use for treatment and control groups at baseline and 6-month visits. Figs 4–6 display the

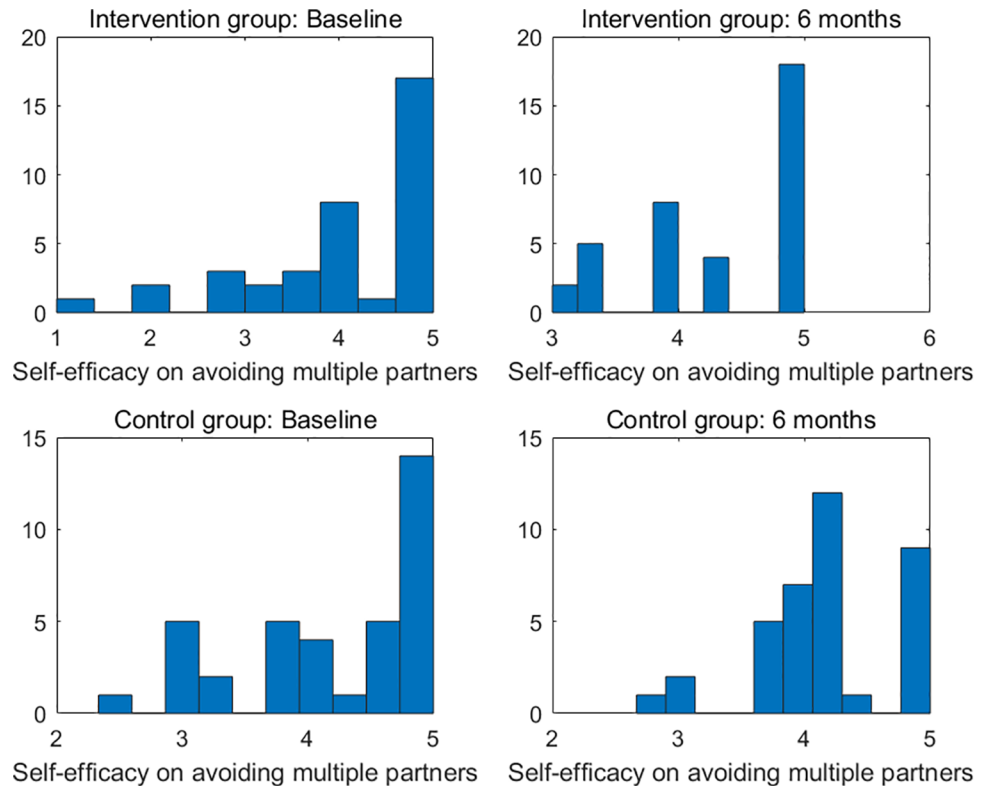


Fig 2. Histograms of self-efficacy on avoiding multiple partners for treatment and control groups at baseline and 6-month visits.

<https://doi.org/10.1371/journal.pone.0195894.g002>

boxplots of visual load, HIV stigma, mental health for treatment and control groups at baseline and 6-month visits.

Hypothesis testing on intervention effect

In the Thai Healthy Choices study, effect of the four-session motivational interviewing-based intervention were simultaneously evaluated by six primary clinical measures: namely HIV sexual risk score, viral load, HIV stigma, mental health, self-efficacy of condom use, and self-efficacy of avoiding multiple sex partners. One approach to determine whether the intervention effect is statistically significant is to conduct a hypothesis test using the nonparametric robust test statistics to determine whether the intervention group and the control group are different in terms of the 6-dimensional multivariate clinical measure at the end of the study (i.e., at 6-month visit). A total number of 74 HIV-positive men who have sex with men were included in this analysis: 37 individuals in intervention group and 37 individuals in control group [18]. Among all participants, 16 of them had missing values and these missing values were replaced with the sample mean of the corresponding variables in each group.

Differences between sample means, medians and two Hodges-Lehmann location estimators of intervention and control groups are reported in S7 Table for each individual clinical measure. These differences demonstrate that the intervention effect may be driven by HIV sexual risk score and HIV stigma. Hypothesis tests were conducted to formally determine whether there was distributional difference between intervention and control groups at baseline and

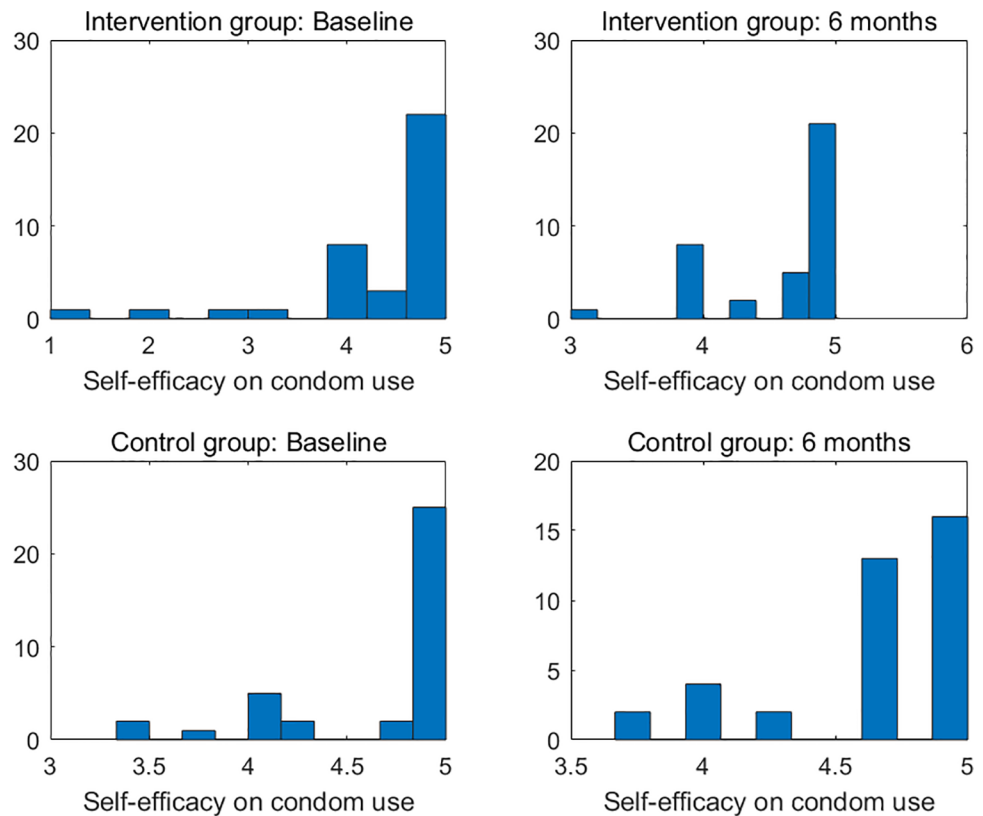


Fig 3. Histograms of self-efficacy on condom use for treatment and control groups at baseline and 6-month visits.

<https://doi.org/10.1371/journal.pone.0195894.g003>

6-month visits. The null hypothesis was that probability distributions of the multivariate clinical measure for intervention and control groups are identical, and the alternative hypothesis was that there was a location shift between the distributions of the multivariate clinical measure for intervention and control groups. The proposed test statistics T_l and T_l^* , $l = 2, 3, 4, 5$, as well as Hotelling's T^2 and extended U statistic, are applied to detect the location shift. The hypothesis testing results, including the values of test statistics and the corresponding p -values, are reported in the upper panel in Table 7. For the baseline samples, all test statistics failed to reject the null hypothesis at the significant level of 0.05, suggesting there was not any statistically significant difference between the probability distributions of the two groups at the baseline visit. For the samples collected at the 6-month follow-up visit, the test statistics T_l and T_l^* , $l = 2, 3$, and Hotelling's T^2 statistic rejected the null hypothesis but others did not. This implied that distributional locations of the samples collected from the two study groups may be statistically different after 6 months of intervention. Furthermore, we compared the two samples collected at the baseline visit and the 6 month visit within each of the two study groups. The null hypothesis was that probability distributions of the multivariate clinical measure are identical at the baseline and the 6 month visits for the intervention group or for the control group, and the alternative hypothesis was that there was a location shift between the distributions of the multivariate clinical measure at the baseline and the 6 month visits within each group. The hypothesis testing results are reported in the lower panel in Table 7. For the intervention group, the test statistics T_2^* and T_3^* rejected the null hypothesis whereas others did not. For the control group, none of the test statistics rejected the null hypothesis.

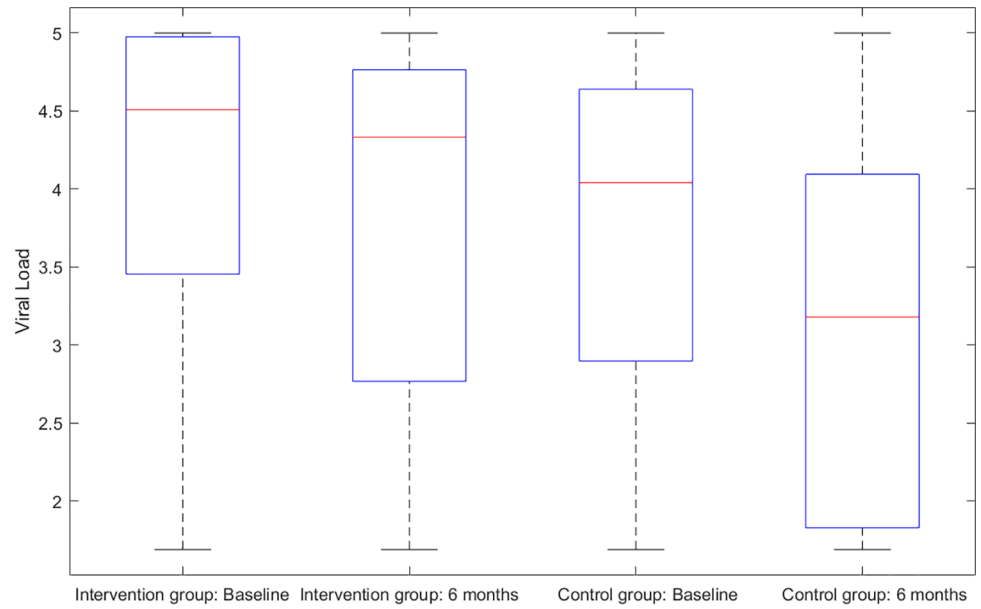


Fig 4. Boxplots of viral load for treatment and control groups at baseline and 6-month visits.

<https://doi.org/10.1371/journal.pone.0195894.g004>

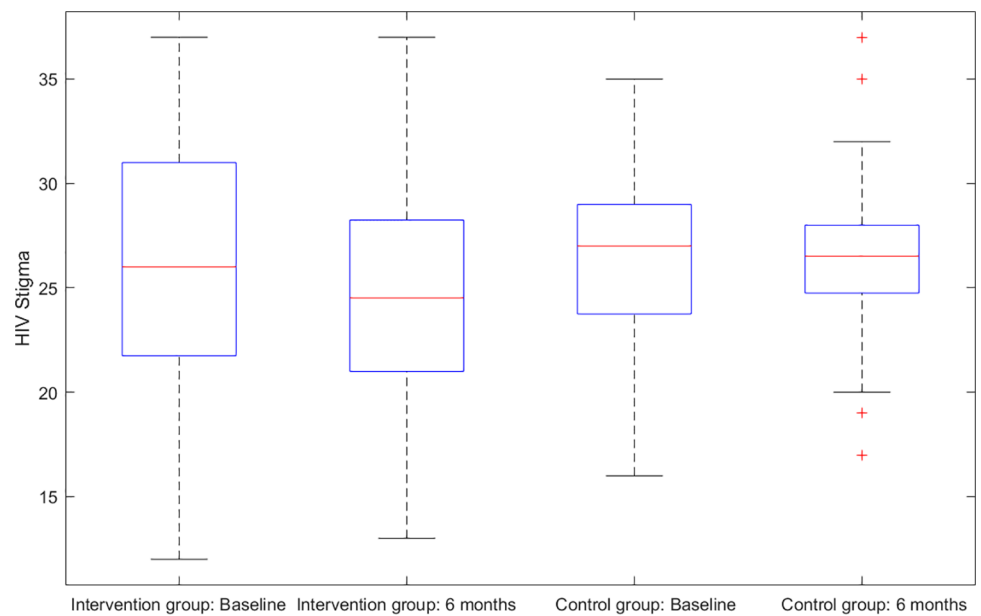


Fig 5. Boxplots of HIV stigma for treatment and control groups at baseline and 6-month visits.

<https://doi.org/10.1371/journal.pone.0195894.g005>

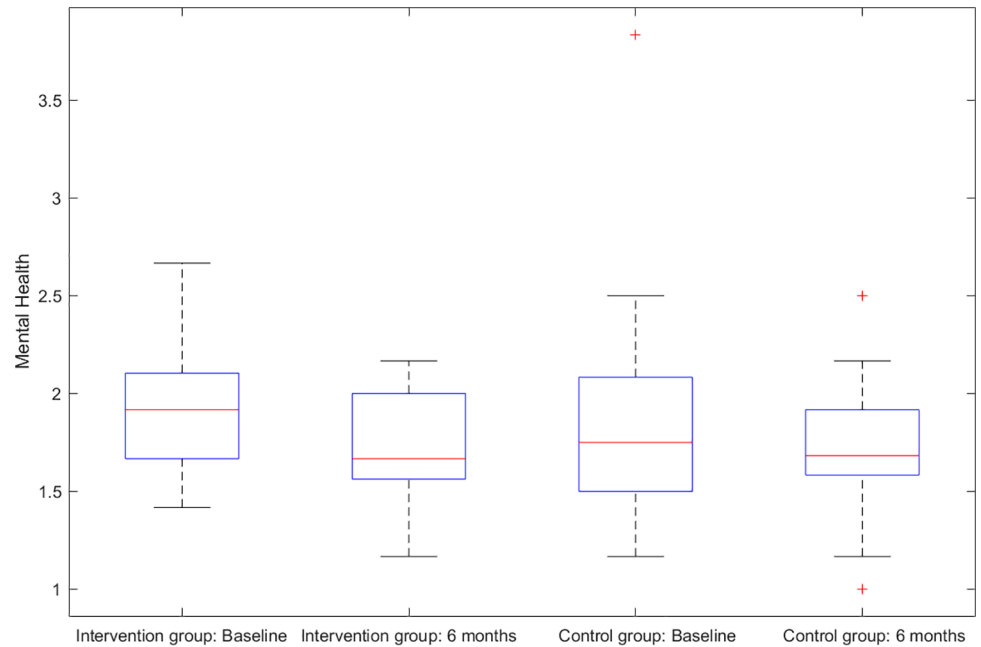


Fig 6. Boxplots of mental health for treatment and control groups at baseline and 6-month visits.

<https://doi.org/10.1371/journal.pone.0195894.g006>

Table 7. Values of test statistics and corresponding *p*-values for comparison of intervention and control groups at baseline and 6-month visits and for comparison of baseline (upper panel) and 6-month visits within each group (lower panel).

		T^2	T_2	T_3	T_4	T_5	T_2^*	T_3^*	T_4^*	T_5^*	U
Intervention vs. control at baseline	Statistic value	1.076	0.033	0.033	0.100	0.100	0.500	0.500	1.000	1.000	652
	<i>p</i> -value (bootstrap)	0.386	0.788	0.788	0.954	0.952	0.274	0.274	0.288	0.286	0.820
	<i>p</i> -value (permutation)	0.386	0.880	0.880	0.982	0.982	0.198	0.246	0.364	0.342	0.768
Intervention vs. control at 6 month	Statistic value	2.747	0.534	0.483	0.443	0.400	0.804	0.689	0.681	0.681	524
	<i>p</i> -value (bootstrap)	0.019	0.024	0.048	0.138	0.164	0.040	0.048	0.184	0.184	0.144
	<i>p</i> -value (permutation)	0.019	0.024	0.032	0.150	0.172	0.026	0.032	0.144	0.164	0.112
Baseline vs. 6 month for intervention	Statistic value	1.719	0.167	0.167	0.207	0.207	0.667	0.667	0.758	0.758	774
	<i>p</i> -value (bootstrap)	0.130	0.578	0.560	0.594	0.590	0.030	0.030	0.386	0.350	0.284
	<i>p</i> -value (permutation)	0.130	0.706	0.704	0.552	0.552	0.034	0.026	0.376	0.360	0.292
Baseline vs. 6 month for control	Statistic value	1.420	0.148	0.148	0.173	0.173	0.481	0.514	0.561	0.599	710
	<i>p</i> -value (bootstrap)	0.220	0.364	0.360	0.620	0.614	0.230	0.142	0.286	0.270	0.576
	<i>p</i> -value (permutation)	0.220	0.418	0.418	0.552	0.552	0.234	0.114	0.220	0.202	0.692

<https://doi.org/10.1371/journal.pone.0195894.t007>

The analysis results of hypothesis testing conclude that there existed statistically significant intervention effect for the four-session motivational interviewing-based intervention developed in the Thai Healthy Choices study to reduce risk behaviors among youth living with HIV. Difference in probability distributions of the multivariate clinical measure for intervention and control groups was detected after 6-month of intervention. Such difference was also confirmed between baseline and 6-month follow-up visits for the intervention group.

Conclusions

This article proposes a series of robust nonparametric test statistics for detecting location shifts between two multivariate samples. The test statistics are constructed based upon the robust estimators of distribution location, including the medians, the two Hodges-Lehmann estimators, and the extended U statistic. Four classes of test statistics are proposed, which include (i) maximum of the component-wise medians or the Hodges-Lehmann estimators, (ii) scaled maximum of the component-wise medians or the Hodges-Lehmann estimators, (iii) maximum of the scaled component-wise medians or the Hodges-Lehmann estimators, and (iv) the extended U statistic. The simulation studies suggest that the proposed robust nonparametric test statistics are effective alternatives to the Hotelling's T^2 . The simulation studies also show that the nonparametric tests built upon the Hodges-Lehmann estimators are generally more powerful than others. Numerous nonparametric hypothesis testing procedures have been proposed for comparing a treatment group and a control group in clinical trials with a multivariate endpoint, in the context of nonparametric Behrens-Fisher hypothesis testing problem [19–22]. Further investigation that compares these hypothesis testing procedures with the procedures included in this article may be relevant.

Supporting information

S1 Table. Type I errors ($\delta = 0$) and power ($\delta \neq 0$) given by the investigated test statistics based on bootstrap approach in detecting location shift between two samples generated from the four pairs of $F(x)$ and $G(x)$ with variance-covariance matrix $I_{4 \times 4}$ and sample sizes $n = m = 10$.

(PDF)

S2 Table. Type I errors ($\delta = 0$) and power ($\delta \neq 0$) given by the investigated test statistics based on bootstrap approach in detecting location shift between two samples generated from the four pairs of $F(x)$ and $G(x)$ with variance-covariance matrix $I_{4 \times 4}$ and sample sizes $n = m = 20$.

(PDF)

S3 Table. Type I errors ($\delta = 0$) and power ($\delta \neq 0$) given by the investigated test statistics based on bootstrap approach in detecting location shift between two samples generated from the four pairs of $F(x)$ and $G(x)$ with variance-covariance matrix $I_{4 \times 4}$ and sample sizes $n = m = 50$.

(PDF)

S4 Table. Type I errors ($\delta = 0$) and power ($\delta \neq 0$) given by the investigated test statistics based on bootstrap approach in detecting location shift between two samples generated from the four pairs of $F(x)$ and $G(x)$ with the non-independent variance-covariance matrix and sample sizes $n = m = 10$.

(PDF)

S5 Table. Type I errors ($\delta = 0$) and power ($\delta \neq 0$) given by the investigated test statistics based on bootstrap approach in detecting location shift between two samples generated from the four pairs of $F(x)$ and $G(x)$ with the non-independent variance-covariance matrix and sample sizes $n = m = 20$.

(PDF)

S6 Table. Type I errors ($\delta = 0$) and power ($\delta \neq 0$) given by the investigated test statistics based on bootstrap approach in detecting location shift between two samples generated from the four pairs of $F(x)$ and $G(x)$ with the non-independent variance-covariance matrix and sample sizes $n = m = 50$.

(PDF)

S7 Table. Differences between sample means, medians and two Hodges-Lehmann location estimators of intervention and control groups in baseline and 6-month visits.

(PDF)

S1 File. Data from the Thai Healthy Choices study.

(XLSX)

Acknowledgments

The authors are thankful to the Academic Editor and the two anonymous referees for their valuable comments and suggestions, which have improved the article substantially. Dr. Xuejun Jiang's research was partially supported by Natural Science Foundation of China (11101432), Natural Science Foundation of Guangdong Province of China (2017A030313012), and Shenzhen Sci-Tech Fund (JCYJ20170307110329106). Dr. Xu Guo's research was partially supported by National Natural Science Foundation of China (11601227 and 11626130 and Natural Science Foundation of Jiangsu Province, China (BK20150732). The research project was partially funded by the National Institute of Mental Health (R34MH077523). The authors thank the principal investigator of the project, Dr. Chokechai Rongkavilit at Valley Children's Hospital for generously providing data for statistical analysis.

Author Contributions

Conceptualization: Xuejun Jiang, Xu Guo, Ning Zhang, Bo Wang, Bo Zhang.

Formal analysis: Xuejun Jiang, Xu Guo, Bo Wang, Bo Zhang.

Investigation: Xuejun Jiang, Bo Zhang.

Methodology: Xuejun Jiang, Xu Guo, Ning Zhang, Bo Wang, Bo Zhang.

Supervision: Xuejun Jiang, Ning Zhang, Bo Zhang.

Writing – original draft: Xuejun Jiang, Xu Guo, Ning Zhang, Bo Wang, Bo Zhang.

Writing – review & editing: Xuejun Jiang, Xu Guo, Ning Zhang, Bo Wang, Bo Zhang.

References

1. Mann HB, Whitney DR. On a test of whether one of two random variable is stochastically larger than the other. *Annals of Mathematical Statistics*. 1947; 18: 50–60. <https://doi.org/10.1214/aoms/1177730491>
2. Yuen KK. The two-sample trimmed t for unequal population variances. *Biometrika*. 1974; 61: 165–170. <https://doi.org/10.2307/2334299>
3. Keselman HJ, Wilcox RR, Kowalchuk RK, Olejnik S. Comparing trimmed or least squares means of two independent skewed populations. *Biometrical Journal*. 2002; 44: 478–489. [https://doi.org/10.1002/1521-4036\(200206\)44:4%3C478::AID-BIMJ478%3E3.0.CO;2-5](https://doi.org/10.1002/1521-4036(200206)44:4%3C478::AID-BIMJ478%3E3.0.CO;2-5)
4. Fried R, Dehling H. Robust nonparametric tests for the two-sample location problem. *Statistical Methods and Application*. 2011; 20: 409–422. <https://doi.org/10.1007/s10260-011-0164-1>
5. Mathur SK. A new nonparametric bivariate test for two sample location problem. *Statistical Methods and Applications*. 2009; 18: 375–388. <https://doi.org/10.1007/s10260-008-0095-7>
6. Hotelling H. The generalization of Student's ratio. *Annals of Mathematical Statistics*. 1931; 2: 360–378. <https://doi.org/10.1214/aoms/1177732979>

7. Hettmansperger TP, Oja H. Affine invariant multivariate multisample sign tests. *Journal of Royal Statistical Society, Series B*. 1994; 56: 235–249.
8. Hettmansperger TP, Möttönen J, Oja H. Affine invariant multivariate rank tests for several samples. *Statistica Sinica*. 1998; 8: 785–800.
9. Neuhaus G, Zhu LX. Permutation tests for multivariate location problem. *Journal of Multivariate Analysis*. 1999; 69: 167–192. <https://doi.org/10.1006/jmva.1998.1808>
10. Henze N, Kar B, Zhu LX. Checking the adequacy of the multivariate semiparametric location shift model. *Journal of Multivariate Analysis*. 2005; 93: 238–256. <https://doi.org/10.1016/j.jmva.2004.04.006>
11. Rongkavilit C, Naar-King S, Wang B, Panthong A, Bunupuradah T, Parsons JT, et al. Motivational interviewing targeting risk behaviors for youth living with HIV in Thailand. *AIDS and Behavior*, 2013; 17: 2063–2074. <https://doi.org/10.1007/s10461-013-0407-2> PMID: 23325376
12. Hodges JL, Lehmann EL. Estimates of location based on rank tests. 1963; *Annals of Mathematical Statistics*. 1963; 34: 598–611. <https://doi.org/10.1214/aoms/1177704172>
13. Chaudhuri P. Multivariate location estimation using extension of R-estimates through U -statistics type approach. *Annals of Statistics*. 1992; 20: 897–916. <https://doi.org/10.1214/aos/1176348662>
14. Ghosh M, Parr WC, Singh K, Babu GJ. A note on bootstrapping the sample median. *The Annals of Statistics*. 1984; 1:1130–1135. <https://doi.org/10.1214/aos/1176346731>
15. Fernández VA, Gamero MJ, García JM. A test for the two-sample problem based on empirical characteristic functions. *Computational statistics & data analysis*. 2008; 52:3730–3748. <https://doi.org/10.1016/j.csda.2007.12.013>
16. Alba-Fernández MV, Batsidis A, Jiménez-Gamero MD, Jodrá P. A class of tests for the two-sample problem for count data. *Journal of Computational and Applied Mathematics*. 2017; 318: 220–229. <https://doi.org/10.1016/j.cam.2016.09.050>
17. Berger BE, Ferrans CE, Lashle FR. Measuring stigma in people with HIV: Psychometric assessment of the HIV stigma scale. *Research in Nursing and Health*, 2001; 24: 518–529. <https://doi.org/10.1002/nur.10011> PMID: 11746080
18. Rongkavilit C, Wang B, Naar-King S, Bunupuradah T, Parsons JT, Panthong A, et al. Motivational interviewing targeting risky sex in HIV-positive young Thai men who have sex with men. *Archives of sexual behavior*. 2015; 44: 329–340. <https://doi.org/10.1007/s10508-014-0274-6> PMID: 24668304
19. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics*. 1984; Dec 1: 1079–1087. <https://doi.org/10.2307/2531158>
20. Huang P, Tilley BC, Woolson RF, Lipsitz S. Adjusting O'Brien's test to control type I error for the generalized nonparametric Behrens-Fisher problem. *Biometrics*. 2005; 61: 532–539. <https://doi.org/10.1111/j.1541-0420.2005.00322.x> PMID: 16011701
21. Liu A, Li Q, Liu C, Yu K, Yu KF. A rank-based test for comparison of multidimensional outcomes. *Journal of the American Statistical Association*. 2010; 105: 578–587. <https://doi.org/10.1198/jasa.2010.ap09114> PMID: 21625372
22. Li Z, Cao F, Zhang J, Li Q. Summation of absolute value test for multiple outcome comparison with moderate effect. *Journal of Systems Science and Complexity*. 2013; 26: 462–469. <https://doi.org/10.1007/s11424-012-0272-5>