Biostatistics (2019) **20**, 2, *pp.* 363–365
doi:10.1093/biostatistics/kxy061
Advance Access publication on December 26, 2018

Response to Guo *et al*.'s Letter to the Editor

YOUYI FONG*, YING HUANG, MARIA P. LEMOS, M. JULIANA MCELRATH
Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
yfong@fhcrc.org

We thank Drs Guo, Gao, Niu, and Zhang for their comments on the article by Fong *and others* (2018). They pointed out that the more classical methods, MW-MW$_2$ and SR-MW$_2$, which only make comparisons between $\mathcal{X}$ and $\mathcal{Y}$ (paired observations) and between $\mathcal{X}'$ and $\mathcal{Y}'$ (unpaired observations) were useful alternatives to the proposed tests, MW-MW$_0^l$ and SR-MW$_0^l$, which made comparisons between all x's and all y's. Dr Guo *et al*.'s recommendation was "*to use MW-MW$_0^l$ and SR-MW$_0^l$ for* $\{\mathcal{X}, \mathcal{Y}, \mathcal{Y}'\}$*, while use MW-MW$_2$ and SR-MW$_2$ for* $\{\mathcal{X}, \mathcal{Y}, \mathcal{X}', \mathcal{Y}'\}$*, especially when the correlation between the samples is high.*" We agree that MW-MW$_2$ and SR-MW$_2$ are important to study as alternative approaches, and aim to refine the recommendations in this response so that practitioners may find it easier to choose the appropriate methods.

Before discussing power comparison, we would like to propose a variant of the MW-MW$_2$ test. Since MW-MW$_2$ only makes comparisons within the paired subset and the unpaired subset, it is possible to perform permutation tests to obtain p-values to avoid inflated Type 1 error rates under small sample sizes (Tables A.1–A.4 of the supplementary material available at *Biostatistics* online). We will refer to this test as MW-MW$_2^{perm}$.

We study power comparison under four different distributional assumptions: normal (Table 1), logistic (Table B.1 of the supplementary material available at *Biostatistics* online), gamma (Table B.2 of the supplementary material available at *Biostatistics* online), and lognormal (Table B.3 of the supplementary material available at *Biostatistics* online). We also plot the results in Figure 1 and Figures B.1, B.2, and B.3 of the supplementary material available at *Biostatistics* online to help visualize these results. All estimates are based on 10^4 Monte Carlo replicates. m, l, n refer to the number of pairs, the number of independent x's and the number of independent y's, respectively. Three levels of correlation between the two samples are examined: 0, 0.5, and 0.8.

First, focusing on lines 2 and 3 in the figures, we see that SR-MW$_2$ and MW-MW$_2^{perm}$ either outperform or closely match the performance of SR at all times. These empirical results are worth noting, because theoretically a test that combines two independent test statistics using weights proportional to the inverse of their variances is not always more powerful than each component test. Based on these results, we can narrow the choice down to be between SR-MW$_0^l$/MW-MW$_0^l$ and SR-MW$_2$/MW-MW$_2^{perm}$ when there are unpaired observations from both samples.

Now, focusing on lines 1 and 2 in the figures, we see that there is a clear trade-off between SR-MW$_0^l$/MW-MW$_0^l$ and SR-MW$_2$/MW-MW$_2^{perm}$ depending on ρ and sample sizes. This is true for normal,

*To whom correspondence should be addressed.

Table 1. *Estimated power, normal distribution, $m = 20$*

(l, n)	MW-MW$_0^l$			MW-MW$_2^{perm}$			SR			SR-MW$_2$			SR-MW$_0^l$		
	0	0.5	0.8	0	0.5	0.8	0	0.5	0.8	0	0.5	0.8	0	0.5	0.8
(10,5)	19	26	46	17	26	49	14	23	51	17	27	52	19	26	44
(10,10)	20	28	47	18	27	51	14	23	51	19	29	53	20	28	46
(40,5)	23	31	52	19	28	51	14	23	51	19	29	53	23	32	51

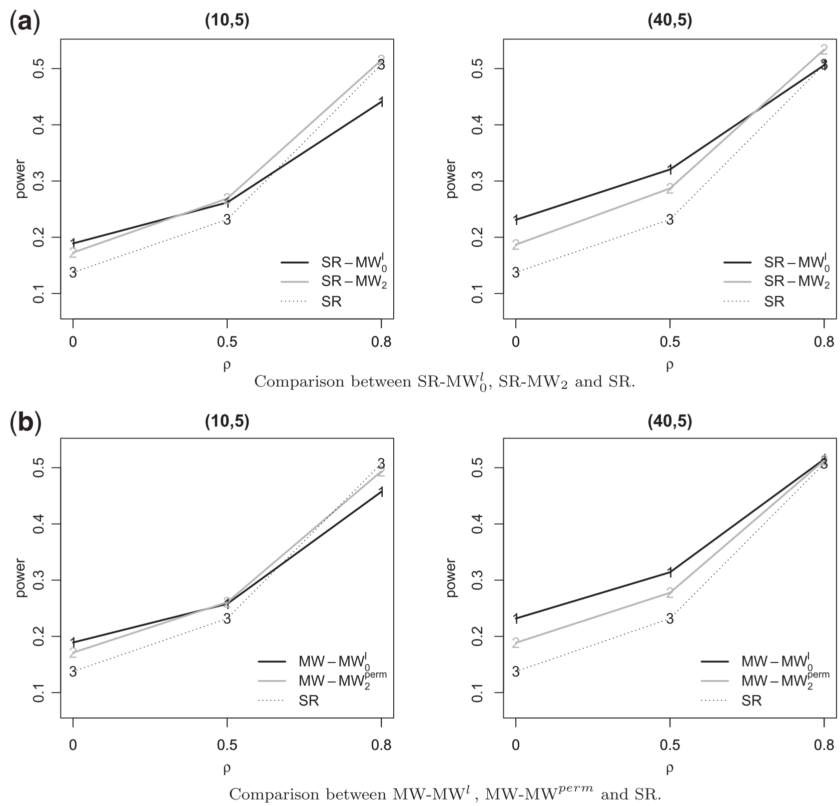

Fig. 1. Power comparison when the marginal distribution is normal. Sample sizes: $m = 20$ and (l, n) are given in the titles.

test. The practical implication of this observation is that we should preprocess the data by applying proper transformation if the distributions appear highly skewed.

Our recommendation for the case when there are unpaired observations from both samples has two parts. If a simple rule of thumb is desirable, our recommendation is to choose SR-MW$_0^l$/MW-MW$_0^l$ when $\rho < 0.5$ and SR-MW$_2$/MW-MW$_2^{perm}$ when $\rho > 0.5$. On the other hand, if an optimal choice is important, we recommend doing a simulation study to find the most powerful approach. To make this a feasible option for practitioners, we provide an easy-to-use function, *choose.test*, in the R package *chngpt*. The only information the function needs is the sample sizes and the estimated first and second moments from the data, and it is fast, for example, it takes only 2 s to run on an Intel i7 processor clocked at 2.6GHz when $m = 20, l = 40, n = 5$.

For the case when there are only unpaired observations from one sample (thus SR-MW$_2$/MW-MW$_2^{perm}$ are not applicable), we recommend choosing between SR and SR-MW$_0^l$/MW-MW$_0^l$ through the *choose.test* function, since there is a trade-off in power between the two tests depending on ρ and sample sizes (Tables D.1–D.3 of the supplementary material available at *Biostatistics* online).

Lastly, given the choice between SR-MW$_0^l$ and MW-MW$_0^l$, we recommend SR-MW$_0^l$ if a monotone transformation can be performed on both samples so that the distributions from both samples are not too skewed. If that is not possible or desirable, for example, when one sample has a highly skewed distribution while the other does not, MW-MW$_0^l$ is preferred because it is a more robust test and invariant to monotone transformations applied to both samples. When using MW-MW$_0^l$, one should proceed with caution as Type 1 error rates may be inflated when sample sizes are small (Tables D.4–D.6 of the supplementary material available at *Biostatistics* online). Similar arguments can be applied to the choice between SR-MW$_2$ and MW-MW$_2^{perm}$, except that there is no concern of inflated Type 1 error rates here.

The chngpt package is available from the Comprehensive R Archive Network, and the Monte Carlo study code can be downloaded at https://github.com/youyifong/response_to_letter_on_rank.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

The authors are grateful to Lindsay N. Carpp for help with editing. *Conflict of Interest*: None declared.

Funding

This work was supported by R01-AI122991, R01-GM106177, UM1-AI068635, UM1-AI068618, and OPP1099507.

Reference

Fong, Y., Huang, Y., Lemos, M. P. and McElrath, M. J. (2018). Rank-based two-sample tests for paired data with missing values. *Biostatistics*, **19**, 281–294.

[*Received September 25, 2018; revised September 25, 2018; accepted for publication September 29, 2018*]