Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230492

Analysis of the trimming methods in olympics diving competitions

Honglei Gu[1, 5], Yicheng Zhou[2], Junran Jia[3], Yuanji Gao[4]

[1]The experimental high school attached to Beijing Normal University, Beijing, 100032, China
[2]New York University, New York College of Arts and Science, 10003, US
[3]University of Michigan - Ann Arbor, Michigan, College of Literature, Science, and Arts (LSA), 48109, US
[4]The Chinese University of Hong Kong, Shenzhen, School of Science and Engineering, 518172, China

[5]raymondguhonglei@163.com

Abstract. The Olympics is the most well-known sports event in the world. With the aim to ensure fairness, the Olympics has some of the strictest scoring methods in the world. However, unlike track and field where the timing and length can be precisely measured, diving has a more subjective judging procedure. Each judge gives their scores based on their own evaluation of the specific dive. Under such scoring methods, it is reasonable to suspect that the scores are influenced by the judges' personal bias. For example, judges may favor their own countrymen or score the competitors lower. In order to reduce the negative effects of such bias, different scoring systems are used in diving in multiple Olympics. Using Sydney 2000 Summer Olympics' and Tokyo 2020 Summer Olympics' data sets, we investigate the phenomenon of nationalistic judging bias and compare different scoring methods' abilities in handling bias. Using three different tests, we discovered that the Mean method is the worst in resisting judging bias. The other three trimming methods were found to withstand the effect of bias successfully in different circumstances, and none of them stands up against judging bias well at all times. Further research with more abundant data is needed before we could obtain a solid conclusion about the most effective scoring system to be used in diving competitions.

Keywords: judgement in sports, scoring, judging bias, diving, competition.

1. Introduction

Sportsmanship is defined as an aspiration that a sport will be enjoyed for its own sake, with proper consideration for fairness, ethics, respect, and a sense of fellowship [1]. Generally recognized as a concrete reflection of sportsmanship, the Olympics Games strive to create a level playing field for athletes around the world. However, there are often obstacles to its realization.

Due to the varying characteristics of different sports, the performances of athletes are judged in vastly different ways. For example, runners in a 100-meter race are ranked objectively according to the timer, while gymnasts' performances are scored according to a comprehensive evaluation offered by several judges [2,3].

There are inherent risks when it comes to subjective scoring methods since the judges might be biased towards certain athletes, either intentionally or unintentionally [3,4]. This bias, if it exists, undermines fairness, a core component in sportsmanship that we strive to achieve. A scandal about figure skating in the 2002 Winter Olympics made this problem prominent, where French judge Le Gougne was accused of giving the Russian team a gold medal deliberately [5].

Although Le Gougne's immoral judging practice was disclosed to the public, the more common occurrence of nationalistic judging bias is less likely to be detected [3,6]. Nationalistic judging bias happens when "a judge awards higher scores than other judges to his own countrymen but fails to award higher scores to non-countrymen" [7]. Previous research by John Emerson has shown "strong evidence of nationalistic favoritism" in the diving competition of 2000 Summer Olympics [7]. Results of this bias can be manifested in different ways: apparently, it might change the final ranking in a competition; but in some cases, it might also affect the ranking in Preliminary or Semi-Final rounds, such that some athletes are not qualified for the Finals. This makes studying the previous rounds of each competition equally important.

As it is difficult to detect nationalistic judging bias and eliminate biased judges, regulations are made by International Olympic Committee to reduce the effect of such bias by trimming scores. However, different trimming methods have been used in diving for different Olympic Games: the top and bottom scores were disregarded in the 2000 Sydney Summer Olympics, while the top two and bottom two scores were trimmed in the 2020 Tokyo Summer Olympics [8].

This research compares different scoring systems and analyzes their abilities to handle nationalistic judging bias. We choose to study diving competitions due to the availability and transparency of data: judges' names and nationalities are obtainable for each dive, enabling us to further our analysis on nationalistic bias [9].

We acquired the data of 2000 Summer Olympics from 'Assessing Judging Bias: An Example From the 2000 Olympic Games' [7], and the data of 2021 Summer Olympics from https://olympics.com/en/olympic-games/tokyo-2020 [10]. All analyses were performed using R Statistical Software (v4.2.0; R Core Team 2022) [11].

The following parts of the paper will be organized in this way: the second part of this paper presents an overview of the data by showing several descriptive graphs; the third part introduces our research methodology and core analysis. The forth part shows our conclusion and further discussion about our results.

2. Data description

2.1. General data description

The data sets are of CSV form, containing information about the diving competitions in Sydney 2000 Summer Olympics and Tokyo 2020 Summer Olympics [7,10]. We adjusted the format of each CSV file to make them have same columns names. Each observation in the table shows a score given by a judge for the specific dive. Each dive is judged by seven judges, so there are repetitions of the specific dive for seven times. The 2000 Sydney data has 10724 observations, recording 1532 dives, while 2020 Tokyo data has 9051 observations, recording 1293 dives. The first 6 rows of each data set are shown here to help visualize the data:

DOI: 10.54254/2755-2721/6/20230492

First six rows of the dataset

```
##     Event Round    Diver Country Rank DiveNo Difficulty  JScore
## 1 M10mPF Final CAO Yuan     CHN    1      1         3.4    10.0
## 2 M10mPF Final CAO Yuan     CHN    1      1         3.4    10.0
## 3 M10mPF Final CAO Yuan     CHN    1      1         3.4    10.0
## 4 M10mPF Final CAO Yuan     CHN    1      1         3.4    10.0
## 5 M10mPF Final CAO Yuan     CHN    1      1         3.4    10.0
## 6 M10mPF Final CAO Yuan     CHN    1      1         3.4     9.5
##                    Judge JCountry
## 1          AXTELIUS Peter      SWE
## 2          HASSAN Mohamed      EGY
## 3             LEE William      SGP
## 4           PETERSON Gord      CAN
## 5 RUIZ PEDREGUERA Rolando      CUB
## 6          SCHLEPPS Holger      GER

##    X Event Round    Diver Country Rank DiveNo Difficulty JScore
## 1 1 M3mSB Final XIONG Ni     CHN    1      1         3.1    8.0
## 2 2 M3mSB Final XIONG Ni     CHN    1      1         3.1    9.0
## 3 3 M3mSB Final XIONG Ni     CHN    1      1         3.1    8.5
## 4 4 M3mSB Final XIONG Ni     CHN    1      1         3.1    8.5
## 5 5 M3mSB Final XIONG Ni     CHN    1      1         3.1    8.5
## 6 6 M3mSB Final XIONG Ni     CHN    1      1         3.1    8.5
##                    Judge JCountry
## 1 RUIZ-PEDREGUERA Rolando      CUB
## 2             GEAR Dennis      NZL
## 3           BOYS Beverley      CAN
## 4           JOHNSON Bente      NOR
## 5         BOUSSARD Michel      FRA
## 6          CALDERON Felix      PUR
```

Figure 1. First six rows of the dataset.

As we can see, there are 10 variables for each observation. The variables are:

•1. *Event*: abbreviated event names, including four different events,'M10mPF' (men's 10-meter platform), 'M3mSB' (men's 3-meter Spring Board), 'W10mPF' (women's 10-meter platform), and 'W3mSB' (women's 3-meter Spring Board).

•2. *Round*: the round of the competition. Each event has three rounds of competitions: 'Prelim' (Preliminary), 'Semi' (Semi-Final), and 'Final'.

•3. *Diver*: diver's name.

•4. *Country*: diver's nationality.

•5. *Rank*: the rank of the diver in this round. The top 18 divers in the Preliminary round are qualified for the Semi-Final, the top 12 divers in the Semi-Final are qualified for the Final round, and the top three in the Final round are awarded with medals.

•6. *DiveNo*: an index indicating the position in a sequence of dives in a round. It ranges from 1 to 6 for men, and 1 to 5 for women.

•7. *Difficulty*: the difficulty of this diving, indicating the complexity for a diver to perform a diving. Divers can choose the difficulty of diving, ranging from 1.5 to 4.8, keeping one decimal place.

•8. *JScore*: the score given by this judge, ranging from 0 to 10.0, keeping half-point margins. Each dive is graded by 7 judges, the final grade is calculated by multiplying the sum of trimmed judges' scores with difficulty [8].

•9. *Judge*: judge's name.

•10. *JCountry*: judge's nationality.

2.2. Data exploration

2.2.1. Score distribution. To have a general view of the score distribution, two plots of frequency are drew as follows:

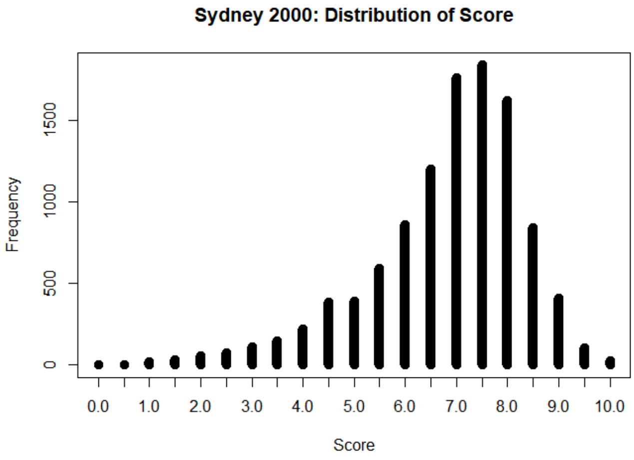

Figure 2. Sydney 2000: Distribution of score.

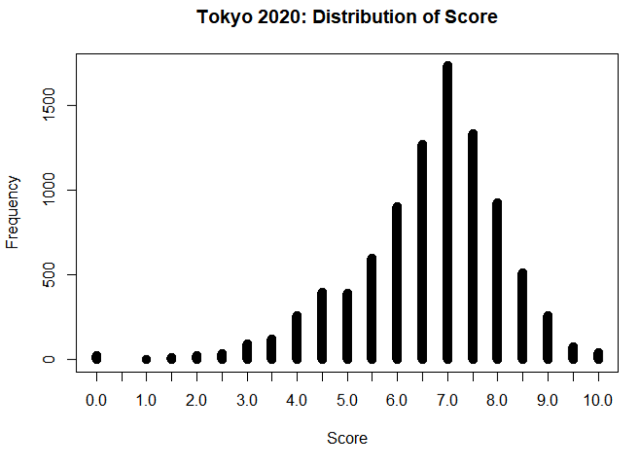

Figure 3. Tokyo 2000: Distribution of score.

The plots indicate that 7.0, 7.5, 8.0 are the most common grades given to decent dives. Excellent dives might

be awarded full points, while failed dives could be scored below 2.5 and even 0 points. The mean score of dives in the Tokyo data set is 6.83, with mode 7.5 and variance 2.05, which is slightly higher than the scores of dives in the Sydney data set, which has a mean of 6.64, mode of 7.0, and variance of 2.16. Both Olympics have 10 cases of full score (10 judge points). Minimum dive score in Sydney is 0, while minimum dive score in Tokyo is 1.0.

2.2.2. Difficulty and score. There are differences between the two data sets. Two plots are shown below to illustrate the relationship between difficulty of a dive and the score (different rounds are marked by different colors)

DOI: 10.54254/2755-2721/6/20230492

Figure 4. Sydney 2000: Difficulty related with JScore.

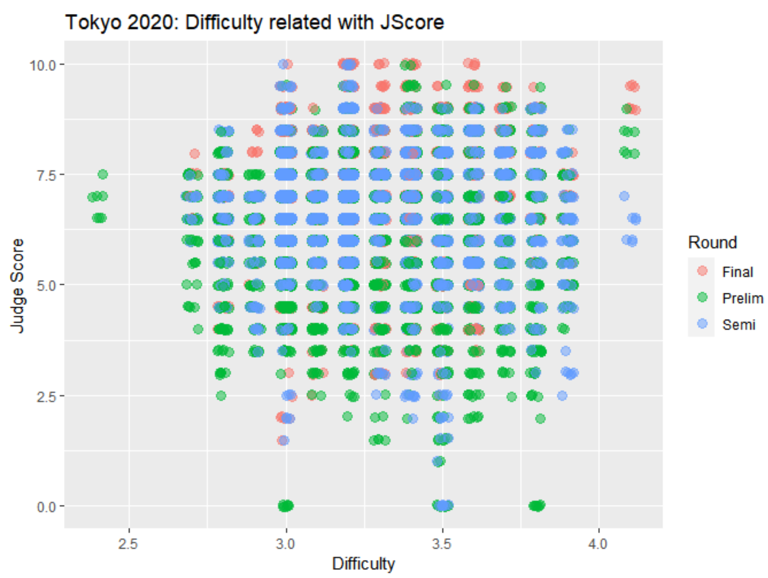

Figure 5. Tokyo 2020: Difficulty related with JScore.

In both Sydney's and Tokyo's data, the range of scores increases when difficulty increases, and both have full points and zero points in some cases. However, Sydney's data have some unique characteristics: there is a blank margin between difficulty 2.1 and 2.4. While final and prelim dives have difficulties equal to or greater than 2.4, all semi-final dives having difficulty less than or equal to 2.1. However, in Tokyo's data, all three rounds have difficulties more than or equal to 2.7 (except for 7 cases in Prelim which have a lower difficulty of 2.4). We suspect that this discrepancy in difficulties is due to some special restriction in the semi-final rounds in Sydney Olympics only. Moreover, most divers in Sydney's final tend to choose difficulties higher than 3.5, while the difficulties in Tokyo's final are more dispersed.

2.2.3. Matching of nationalities between judges and divers. Two more diagrams were made to visualize the matching degree between judges' and athletes' nationality. The x-axis indicates judges' names, and the y-axis indicates the frequency of matching.

DOI: 10.54254/2755-2721/6/20230492

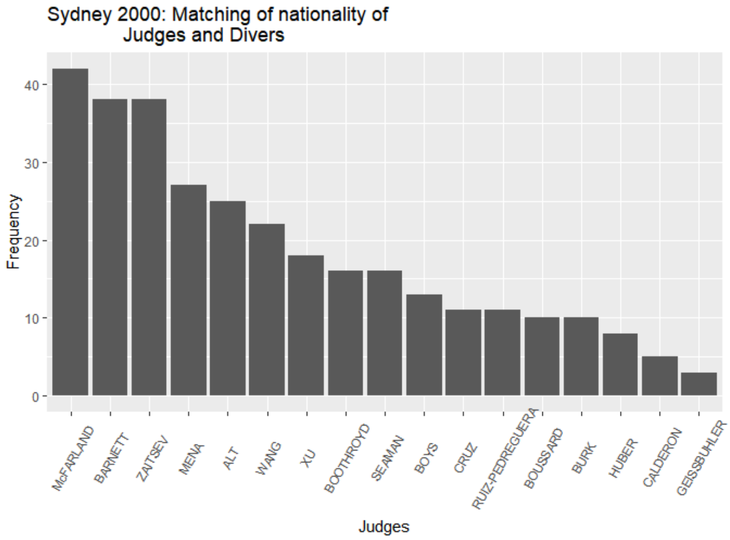

Figure 6. Sydney 2000.

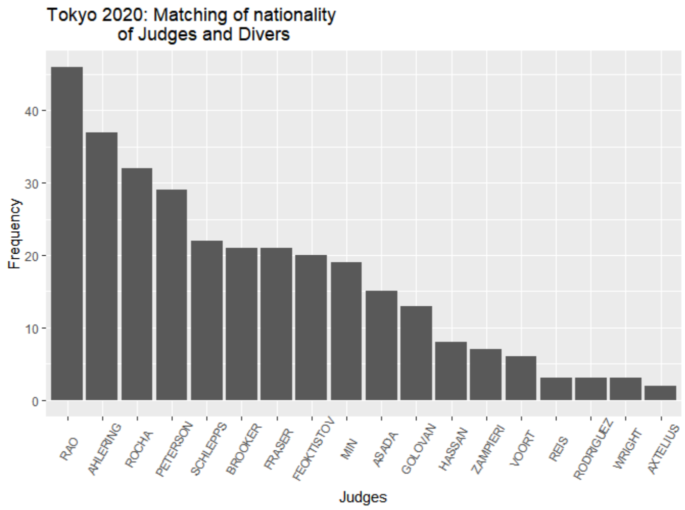

Figure 7. Tokyo 2020.

As seen from above, both Olympics have over three-hundred cases where a judge's and a diver's nationalities are matched: there are 314 cases (20.3%) for Sydney, and 307 cases (23.7%) for Tokyo. In both Olympics, the frequency of matching for some judges is more than 40. With the large number of matching in both cases, the effect of nationalistic judging bias is expected to be seen in the results. It is important to note that as an attempt to reduce the impact of nationalistic judging bias, none of the judges scored divers from the same country in final rounds for both Olympics. However, as mentioned above, some athletes' ranking might still be influenced in the preliminary and semi-final rounds, with some failing to advance to the next round by a tiny score margin. Hence, this study will focus primarily on the preliminary and semi-final rounds of the competition.

The following part of this paper will describe our method and analysis in detail.

3. Analysis

3.1. Test preparation

We aim to compare robustness of different scoring system, and identify one with the best ability to reduce the effect of potential bias. Measurement of the robustness of a scoring system is achieved by evaluating its ability to reduce the impact of nationalistic judging bias in case it exists. The first step is to identify the judges who appears to be nationalistically biased, no matter being intentional or accidental. One prerequisite of the existence of nationalistic judging bias is that there *are* dives where nationality of the diver and that of the judge are the same. Thus, we create a new column in the data frame to indicate the matching of nationality between judges and divers. Attempts such as choosing a judge and simply comparing the scores given to divers with matching and non-matching nationality is problematic, since it does not eliminate the possibility that some divers are indeed better at diving so that they merit higher scores. Comparing a judge's scores given to his own countryman with the panel average is also erroneous, since the judge might award higher than average scores to *all* the divers.

Therefore, we decide to conduct a two-sample t-test on judges' score difference with panel's average when there is matching and that when there is no matching. Here, we introduce two more variables: the average of the seven judges' score in each dive and the difference between the judge's score and the panel average for each dive. The average here would serve as a rough indicator of divers' skill and performance, and the difference implies how deviated the score given by the matched judge is. We choose one judge and collect the scores given to divers with same nationality and divers with different nationality. Then, we conduct two-sample t-test on these two samples to determine whether the difference in means of these two groups of scores is significant. The same procedures are repeated to every judge in the two Olympics We extract p-value from each judge and plot it against the judges' names, obtaining the graphs as follows:

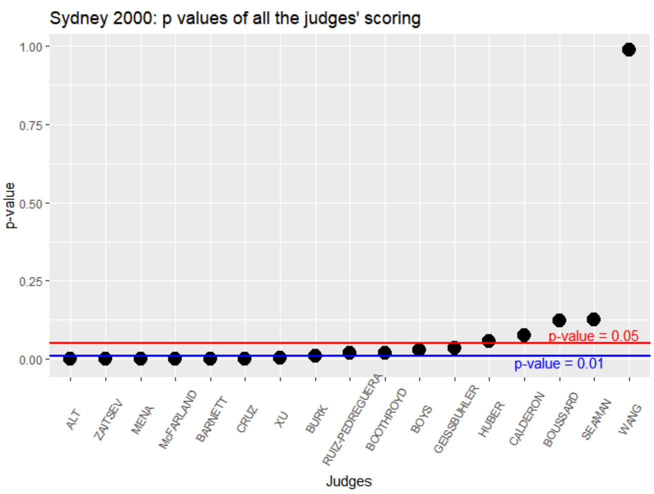

Figure 8. Sydney 2000.

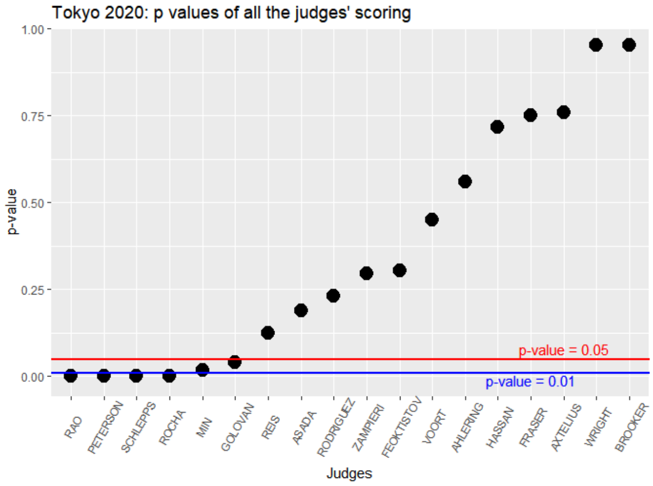

Figure 9. Tokyo 2020.

If we take a look at p-values below 0.05, 6 judges in Tokyo 2020 (Rao Lang, Peterson Gord, Golovan Anatoliy, Rocha Chavez Sergio, Schlepps Holger, and Min Suckhong) and 12 judges in Sydney 2000 (Mena Jesus, Zaitsev Oleg, McFarland Steve, Alt Walter, Barnett Madeleine, Boothroyd Sydney, Ruiz-Pedreguera Rolando, Cruz Julia, Boys Beverley, Burk Hans-Peter, Xu Yiming, and Geissbuhler Michael) are suspicious of being biased.

With specific judges to focus on, we are ready to analyze the robustness of different trimming methods. Four methods will be analyzed and compared. The Mean method scores by multiplying the mean of all seven judges with the difficulty of the dive. Trim 1 method scores by trimming the maximum and the minimum of seven scores and multiplying the sum of the rest five scores with difficulty (adopted in the 2000 Sydney Diving). Trim 2 method scores by trimming two maximum and two minimum scores and multiplying the sum of the rest three scores with difficulty (adopted in the 2020 Tokyo Diving). Result of Trim 3 is literally the product of median of seven scores and the difficulty. To further explore bias involved in judging of Olympics Diving and study robustness of each method, we designed three tests: Random Simulation Test aims to restore the original rank positions of the divers if positive judging bias were absent; Nationalistic Bias Test assumes extreme nationalistic bias to judges' countrymen; Competitive Bias Test assumes extreme competitive bias to competitors of their own countrymen.

In the Random Simulation test, we assumed that *all* the scores given by the potentially biased judges to their countrymen are biased scores. No matter what the trimming method is, those divers are benefited from the biased score directly or indirectly. We replaced all the potentially biased scores with one of the other six scores in the same dive (assumed to be unbiased). This enables us to observe how the ranks should have been, if all the biased scores were absent. We hypothesized that if the change in ranks is obvious, the judges' biases towards their countrymen were strong.

In the second test, we maximized all the potentially biased scores to 10, creating a case of extreme bias. We then compared the robustness of each scoring method: Mean, Trim1, Trim2, and Trim3. The larger the change in rank, the less robust a system is, since positive bias is not removed and manifests itself in the game result.

In the third test, we assumed that all judges exhibit competitive bias towards competitors of their own countrymen and set their scores to a low value. For the purpose of this research, we use the preliminary rankings to locate targets for each judge. If this judge scores the target again in the semi-final round of the same event, the score will be only 5 points. By setting a low score regardless of the diver's performance, we aim to compare the ability of each scoring system in dealing with competitive bias.

DOI: 10.54254/2755-2721/6/20230492

3.2. Random simulation test

At first, all judges with p-values lower than 0.05 are marked as "suspicious judges". For each round, the scores given by the suspicious judges to the divers of the same nationality are replaced by one of the other six judges randomly. Then, the final scores and rankings will be calculated again according to the revised scores. The outcomes before and after randomization are compared and shown in plots. To better illustrate the effects of random simulation, we conduct three simulations, and each of the new lists of scores and ranks is recorded. If a diver has a matched suspicious judge, this diver would be marked with "**" on the left side of the plot, indicating that his/her rank might be affected by randomization. The column "O_rank" represents each diver's original rank, and the columns named "Score x" and "Rank x" represent the scores and rankings after x times of randomization. The simulation result of the preliminary of Men's 10-meter platform in the Tokyo Olympics employing the Mean method is shown below as an example:

Preliminary of Men's 10-meter platform in Tokyo, employing Mean

##	Diver	Match	O_Rank	Score 1	Rank 1	Score 2	Rank 2
## 1	YANG Jian	**	1	180.29	1	181.28	1
## 2	CAO Yuan	**	2	174.82	2	175.08	2
## 3	BONDAR Aleksandr		3	170.94	3	170.94	3
## 4	DALEY Thomas		4	151.64	4	151.64	4
## 5	ZSOMBOR-MURRAY Nathan	**	5	147.54	5	147.31	5
## 6	SEREDA Oleksii	**	6	145.05	6	145.05	6
## 7	WOO Haram		7	142.89	7	142.89	7
## 8	ROUSSEAU Cassiel		8	141.24	8	141.24	8
## 9	VILLARREAL TUDON Andres Isaac	**	9	135.66	9	136.16	9
## 10	VILLA CASTANEDA Sebastian		10	134.92	10	134.92	10
## 11	BARTHEL Timo	**	11	133.84	11	133.89	11
## 12	LOSCHIAVO Brandon		12	133.63	12	133.63	12
## 13	QUINTERO DIAZ Rafael		13	133.25	13	133.25	13
## 14	MINIBAEV Viktor		14	131.25	14	131.25	14
## 15	WINDLE Jordan		15	130.91	15	130.91	15
## 16	FIGUEREDO PEREIRA Kawan		16	124.74	16	124.74	16
## 17	TAMAI Rikuto		17	124.69	17	124.69	17
## 18	KIM Yeongtaek		18	122.64	18	122.64	18
## 19	WIENS Rylan	**	19	122.27	19	122.20	19
## 20	SOUZA FILHO Isaac		20	113.01	20	113.01	20
## 21	EIKERMANN GREGORCHUK Jaden	**	21	110.47	21	110.69	21
## 22	ARIZA Oscar		22	108.46	22	108.46	22
## 23	ISHAK Mohab		23	105.21	23	105.21	23
## 24	GARCIA NAVARRO Ivan	**	24	104.38	25	104.41	25
## 25	NISHIDA Reo		25	104.91	24	104.91	24
## 26	CHAN Jonathan		26	104.25	26	104.25	26
## 27	FRICKER Samuel		27	102.85	27	102.85	27
## 28	WILLIAMS Noah		28	102.15	28	102.15	28
## 29	ROSSET Matthieu		29	92.59	29	92.59	29

##	Score 3	Rank 3	Score 4	Rank 4
## 1	181.27	1	180.79	1
## 2	175.32	2	175.09	2
## 3	170.94	3	170.94	3
## 4	151.64	4	151.64	4
## 5	147.78	5	147.78	5
## 6	145.05	6	145.05	6
## 7	142.89	7	142.89	7
## 8	141.24	8	141.24	8
## 9	136.14	9	136.40	9
## 10	134.92	10	134.92	10
## 11	134.31	11	134.10	11
## 12	133.63	12	133.63	12
## 13	133.25	13	133.25	13
## 14	131.25	14	131.25	14
## 15	130.91	15	130.91	15
## 16	124.74	16	124.74	16
## 17	124.69	17	124.69	17
## 18	122.64	19	122.64	18
## 19	122.73	18	122.46	19
## 20	113.01	20	113.01	20
## 21	110.24	21	110.23	21
## 22	108.46	22	108.46	22
## 23	105.21	23	105.21	23
## 24	104.38	25	104.16	26
## 25	104.91	24	104.91	24
## 26	104.25	26	104.25	25
## 27	102.85	27	102.85	27
## 28	102.15	28	102.15	28
## 29	92.59	29	92.59	29

Figure 10. Prelim of men's 10-meter in Tokyo.

The "Match" column shows that 9 of 29 divers have a corresponding suspicious judge. The scores of these divers may either increase or decrease after simulation, and the results reveal little difference

DOI: 10.54254/2755-2721/6/20230492

between original ranks and simulated ranks, as only the rank of Wiens Rylan (#19) and Garcia Narvado Ivan (#24) have a one-rank change.

For a particular trimming method, if the rank of a marked diver has an evident downward change after random simulation, we may infer that this diver probably benefited from nationalistic bias in this circumstance. On the contrary, if the general ranking does not change obviously, we may infer that nationalistic bias does not have a distinct effect on the final results under this trimming method. Length limited, this paper will only show a few samples of the random simulation instead of all graphs of each round.

To compare the effects of random simulation on different trimming methods, we employ Trim 1 method to the same round, the result is as follows:

Preliminary of Men's 10-meter platform in Tokyo, employing Trim 1

##	Diver	Match	O_Rank	Score 1	Rank 1	Score 2	Rank 2	Score 3	Rank 3	Score 4	Rank 4
## 1	YANG Jian	**	1	906.05	1	907.80	1	904.35	1	907.75	1
## 2	CAO Yuan	**	2	879.45	2	877.45	2	877.65	2	877.45	2
## 3	BONDAR Aleksandr		3	856.15	3	856.15	3	856.15	3	856.15	3
## 4	DALEY Thomas		4	756.80	4	756.80	4	756.80	4	756.80	4
## 5	ZSOMBOR-MURRAY Nathan	**	5	739.05	5	737.65	5	737.45	5	735.80	5
## 6	SEREDA Oleksii	**	6	726.60	6	726.60	6	726.60	6	726.60	6
## 7	WOO Haram		7	713.85	7	713.85	7	713.85	7	713.85	7
## 8	ROUSSEAU Cassiel		8	703.40	8	703.40	8	703.40	8	703.40	8
## 9	VILLARREAL TUDON Andres Isaac	**	9	684.30	9	680.75	9	682.70	9	684.50	9
## 10	VILLA CASTANEDA Sebastian		10	677.65	10	677.65	10	677.65	10	677.65	10
## 11	LOSCHIAVO Brandon		11	671.50	11	671.50	11	671.50	11	671.50	11
## 12	QUINTERO DIAZ Rafael		12	664.70	12	664.70	12	664.70	12	664.70	12
## 13	BARTHEL Timo	**	13	662.60	13	662.60	13	664.20	13	662.60	13
## 14	MINIBAEV Viktor		14	654.35	14	654.35	14	654.35	14	654.35	14
## 15	WINDLE Jordan		15	652.40	15	652.40	15	652.40	15	652.40	15
## 16	TAMAI Rikuto		16	624.45	16	624.45	16	624.45	16	624.45	16
## 17	FIGUEREDO PEREIRA Kawan		17	623.25	17	623.25	17	623.25	17	623.25	17
## 18	WIENS Rylan	**	18	612.65	18	614.75	18	617.95	18	612.90	18
## 19	KIM Yeongtaek		19	611.65	19	611.65	19	611.65	19	611.65	19
## 20	SOUZA FILHO Isaac		20	566.15	20	566.15	20	566.15	20	566.15	20
## 21	EIKERMANN GREGORCHUK Jaden	**	21	547.10	21	550.25	21	550.35	21	551.85	21
## 22	ARIZA Oscar		22	542.30	22	542.30	22	542.30	22	542.30	22
## 23	GARCIA NAVARRO Ivan	**	23	524.00	25	524.10	25	527.40	23	524.10	25
## 24	ISHAK Mohab		24	526.85	23	526.85	23	526.85	24	526.85	23
## 25	NISHIDA Reo		25	525.30	24	525.30	24	525.30	25	525.30	24
## 26	CHAN Jonathan		26	519.65	26	519.65	26	519.65	26	519.65	26
## 27	WILLIAMS Noah		27	514.75	27	514.75	27	514.75	27	514.75	27
## 28	FRICKER Samuel		28	510.80	28	510.80	28	510.80	28	510.80	28
## 29	ROSSET Matthieu		29	457.90	29	457.90	29	457.90	29	457.90	29

Figure 11. Prelim of men's 10-meter in Tokyo employing Trim 1.

Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230492

According to the graph, among the 9 marked divers, only Garcia Narvado Ivan's (#23) ranking is affected, shifting from 23rd to 25th. This result does not differ much from Mean method, so we continue to apply Trim2 and Trim 3 method on the same round.

Preliminary of Men's 10-meter platform in Tokyo, employing Trim 2

##	Diver	Match	O_Rank	Score 1	Rank 1	Score 2	Rank 2
## 1	YANG Jian	**	1	545.20	1	546.90	1
## 2	CAO Yuan	**	2	527.70	2	524.20	2
## 3	BONDAR Aleksandr		3	513.85	3	513.85	3
## 4	DALEY Thomas		4	453.70	4	453.70	4
## 5	ZSOMBOR-MURRAY Nathan	**	5	440.40	5	442.00	5
## 6	SEREDA Oleksii	**	6	435.90	6	434.30	6
## 7	WOO Haram		7	427.25	7	427.25	7
## 8	ROUSSEAU Cassiel		8	423.55	8	423.55	8
## 9	VILLARREAL TUDON Andres Isaac	**	9	410.45	9	408.65	9
## 10	VILLA CASTANEDA Sebastian		10	407.30	10	407.30	10
## 11	LOSCHIAVO Brandon		11	403.85	11	403.85	11
## 12	QUINTERO DIAZ Rafael		12	396.90	12	396.90	12
## 13	BARTHEL Timo	**	13	394.20	13	395.70	13
## 14	MINIBAEV Viktor		14	391.95	14	391.95	14
## 15	WINDLE Jordan		15	390.05	15	390.05	15
## 16	TAMAI Rikuto		16	374.25	16	374.25	16
## 17	FIGUEREDO PEREIRA Kawan		17	371.65	17	371.65	17
## 18	KIM Yeongtaek		18	366.80	19	366.80	18
## 19	WIENS Rylan	**	19	368.55	18	366.70	19
## 20	SOUZA FILHO Isaac		20	339.30	20	339.30	20
## 21	EIKERMANN GREGORCHUK Jaden	**	21	327.65	21	330.75	21
## 22	ARIZA Oscar		22	327.05	22	327.05	22
## 23	ISHAK Mohab		23	318.55	23	318.55	23
## 24	GARCIA NAVARRO Ivan	**	24	313.45	25	313.45	25
## 25	NISHIDA Reo		25	314.30	24	314.30	24
## 26	CHAN Jonathan		26	311.15	26	311.15	26
## 27	WILLIAMS Noah		27	309.55	27	309.55	27
## 28	FRICKER Samuel		28	306.50	28	306.50	28
## 29	ROSSET Matthieu		29	275.70	29	275.70	29

##	Score 3	Rank 3	Score 4	Rank 4
## 1	546.90	1	546.90	1
## 2	527.70	2	527.60	2
## 3	513.85	3	513.85	3
## 4	453.70	4	453.70	4
## 5	442.00	5	442.00	5
## 6	435.90	6	435.90	6
## 7	427.25	7	427.25	7
## 8	423.55	8	423.55	8
## 9	406.85	10	406.70	10
## 10	407.30	9	407.30	9
## 11	403.85	11	403.85	11
## 12	396.90	12	396.90	12
## 13	394.20	13	394.20	13
## 14	391.95	14	391.95	14
## 15	390.05	15	390.05	15
## 16	374.25	16	374.25	16
## 17	371.65	17	371.65	17
## 18	366.80	19	366.80	19
## 19	366.95	18	366.95	18
## 20	339.30	20	339.30	20
## 21	327.60	21	329.15	21
## 22	327.05	22	327.05	22
## 23	318.55	23	318.55	23
## 24	315.15	24	315.15	24
## 25	314.30	25	314.30	25
## 26	311.15	26	311.15	26
## 27	309.55	27	309.55	27
## 28	306.50	28	306.50	28
## 29	275.70	29	275.70	29

Figure 12. Prelim of men's 10-meter in Tokyo employing Trim 2.

DOI: 10.54254/2755-2721/6/20230492

Preliminary of Men's 10-meter platform in Tokyo, employing Trim 3

```
##                              Diver Match O_Rank Score 1 Rank 1 Score 2 Rank 2
## 1                        YANG Jian    **      1  182.40      1  184.15      1
## 2                         CAO Yuan    **      2  176.05      2  175.95      2
## 3                 BONDAR Aleksandr            3  171.30      3  171.30      3
## 4                    DALEY Thomas             4  150.70      4  150.70      4
## 5           ZSOMBOR-MURRAY Nathan    **      5  148.40      5  146.80      5
## 6                   SEREDA Oleksii    **      6  145.30      6  145.30      6
## 7                        WOO Haram            7  144.25      7  144.25      7
## 8                 ROUSSEAU Cassiel            8  141.80      8  141.80      8
## 9  VILLARREAL TUDON Andres Isaac    **      9  136.25      9  136.25      9
## 10      VILLA CASTANEDA Sebastian           10  135.20     10  135.20     10
## 11               LOSCHIAVO Brandon           11  134.55     11  134.55     11
## 12            QUINTERO DIAZ Rafael           12  133.95     12  133.95     12
## 13                 MINIBAEV Viktor           13  131.25     13  131.25     13
## 14                    BARTHEL Timo    **     14  130.80     14  130.80     14
## 15                  WINDLE Jordan           15  129.35     15  129.35     15
## 16                    TAMAI Rikuto           16  124.15     16  124.15     16
## 17                  KIM Yeongtaek           17  123.45     17  123.45     18
## 18                     WIENS Rylan    **     18  122.15     18  124.00     17
## 19        FIGUEREDO PEREIRA Kawan           19  121.60     19  121.60     19
## 20               SOUZA FILHO Isaac           20  110.75     20  110.75     20
## 21    EIKERMANN GREGORCHUK Jaden    **     21  109.70     21  108.05     22
## 22                     ARIZA Oscar           22  108.40     22  108.40     21
## 23                     ISHAK Mohab           23  107.10     23  107.10     23
## 24                    NISHIDA Reo           24  105.30     24  105.30     24
## 25                  WILLIAMS Noah           25  104.35     25  104.35     25
## 26              GARCIA NAVARRO Ivan    **     26  103.20     26  103.20     26
## 27                  CHAN Jonathan           27  102.75     27  102.75     27
## 28                 FRICKER Samuel           28  102.20     28  102.20     28
## 29               ROSSET Matthieu           29   90.30     29   90.30     29
##    Score 3 Rank 3 Score 4 Rank 4
## 1   184.15      1  184.15      1
## 2   177.55      2  175.85      2
## 3   171.30      3  171.30      3
## 4   150.70      4  150.70      4
## 5   146.80      5  148.40      5
## 6   145.30      6  145.30      6
## 7   144.25      7  144.25      7
## 8   141.80      8  141.80      8
## 9   136.25      9  136.25      9
## 10  135.20     10  135.20     10
## 11  134.55     11  134.55     11
## 12  133.95     12  133.95     12
## 13  131.25     13  131.25     13
## 14  130.80     14  130.80     14
## 15  129.35     15  129.35     15
## 16  124.15     16  124.15     16
## 17  123.45     17  123.45     17
## 18  122.40     18  120.55     19
## 19  121.60     19  121.60     18
## 20  110.75     20  110.75     20
## 21  108.05     22  108.05     22
## 22  108.40     21  108.40     21
## 23  107.10     23  107.10     23
## 24  105.30     24  105.30     24
## 25  104.35     25  104.35     25
## 26  103.20     26  103.20     26
## 27  102.75     27  102.75     27
## 28  102.20     28  102.20     28
## 29   90.30     29   90.30     29
```

Figure 13. Prelim of men's 10-meter in Tokyo employing Trim 3.

These two graphs does not reveal obvious differences between trimming methods. For Trim 2, three divers' ranks shift(#9, #19, and #24), and for Trim 3, two diver's ranks shift(#18 and #21). All trimming methods only have a few changed cases, and the number of changed cases does not have a regular pattern. Hence, we speculate that it is hard to make comparison between different trimming methods.

DOI: 10.54254/2755-2721/6/20230492

Considering that choosing one particular round for comparison may have uncertainty, we use another sample from Sydney Olympics.

Preliminary of Men's 10-meter platform in Sydney, employing mean

##	Diver	Match	O_Rank	Score 1	Rank 1	Score 2	Rank 2	Score 3
## 1	TIAN Liang	**	1	167.44	1	167.44	1	167.44
## 2	HU Jia	**	2	162.12	2	162.12	2	162.12
## 3	SAOUTINE Dmitri	**	3	156.93	3	156.93	3	157.16
## 4	TERAUCHI Ken		4	152.54	4	152.54	4	152.54
## 5	HELM Mathew	**	5	151.81	5	151.81	5	151.81
## 6	CHOE Hyong-Gil		6	149.55	6	149.55	6	149.55
## 7	PICHLER David	**	7	146.59	7	147.02	7	146.82
## 8	DESPATIE Alexandre	**	8	144.66	8	144.66	8	144.66
## 9	NEWBERY Robert	**	9	144.23	9	144.23	9	144.23
## 10	LOUKACHINE Igor	**	10	143.92	10	144.15	10	143.01
## 11	RUIZ Mark	**	11	142.94	11	142.72	11	142.91
## 12	MEYER Heiko	**	12	137.33	12	137.33	12	137.33
## 13	GUERRA Jose-Antonio		13	134.49	13	134.49	13	134.49
## 14	HEMPEL Jan	**	14	133.50	14	133.50	14	133.50
## 15	PAK Yong-Ryong		15	133.06	15	133.06	15	133.06
## 16	TAYLOR Leon	**	16	132.50	16	132.50	16	132.50
## 17	KALEC Christopher	**	17	129.99	17	129.99	17	129.99
## 18	MAZZUCCHI Massimiliano		18	128.94	18	128.94	18	128.94
## 19	PEREZ Francisco	**	19	128.16	19	127.49	19	127.72
## 20	VOLODKOV Roman		20	126.09	20	126.09	20	126.09
## 21	EMPTOZ-LACOTE Gilles		21	125.31	21	125.31	21	125.31
## 22	ABALLI Jesus-Iory		22	122.62	22	122.62	22	122.62
## 23	SKRYPNIK Oleksandr		23	121.45	23	121.45	23	121.45
## 24	YEOH Ken-Nee		24	120.86	24	120.86	24	120.86
## 25	BAHARI Mohd-Azheem		25	114.09	25	114.09	25	114.09
## 26	JABRAYILOV Emil		26	111.46	26	111.46	26	111.46
## 27	URAN Juan-Guillermo		27	111.19	27	111.19	27	111.19
## 28	DURAN Cassius		28	110.60	28	110.60	28	110.60
## 29	YOO Chang-Joon		29	110.37	29	110.37	29	110.37
## 30	CHO Dae-Don		30	106.73	30	106.73	30	106.73
## 31	GURMAN Alexey		31	106.34	31	106.34	31	106.34
## 32	CHERECHES Gabriel		32	106.21	32	106.21	32	106.21
## 33	WATERFIELD Peter	**	33	105.09	33	105.09	33	105.09
## 34	HAJNAL Andras		34	104.76	34	104.76	34	104.76
## 35	VILLARROEL Luis		35	102.86	35	102.86	35	102.86
## 36	SANTOS Ruben	**	36	100.47	36	100.47	36	100.47
## 37	SANCHEZ Abel		37	97.69	37	97.69	37	97.69
## 38	AVTANDILYAN Hovhannes		38	95.14	38	95.14	38	95.14
## 39	PICHI Suchart		39	89.00	39	89.00	39	89.00
## 40	NASRULLAH Muhammad		40	85.70	40	85.70	40	85.70

## 1	1	167.44	1
## 2	2	162.12	2
## 3	3	156.92	3
## 4	4	152.54	4
## 5	5	151.81	5
## 6	6	149.55	6
## 7	7	146.79	7
## 8	8	144.66	8
## 9	9	144.23	9
## 10	10	143.92	10
## 11	11	142.69	11
## 12	12	137.33	12
## 13	13	134.49	13
## 14	14	133.50	14
## 15	15	133.06	15
## 16	16	132.50	16
## 17	17	129.99	17
## 18	18	128.94	18
## 19	19	127.72	19
## 20	20	126.09	20
## 21	21	125.31	21
## 22	22	122.62	22
## 23	23	121.45	23
## 24	24	120.86	24
## 25	25	114.09	25
## 26	26	111.46	26
## 27	27	111.19	27
## 28	28	110.60	28
## 29	29	110.37	29
## 30	30	106.73	30
## 31	31	106.34	31
## 32	32	106.21	32
## 33	33	105.09	33
## 34	34	104.76	34
## 35	35	102.86	35
## 36	36	100.47	36
## 37	37	97.69	37
## 38	38	95.14	38
## 39	39	89.00	39

Figure 14. Prelim of men's 10-meter in Sydney employing mean.

DOI: 10.54254/2755-2721/6/20230492

The “Match” column shows that 16 of 40 divers have a corresponding suspicious judge. However, none of their ranks change after simulation. Length limited, other graphs are attached in the appendix. We speculate that the results of random simulation are not obvious because the change in the scores is smaller than the original gap, such that no change in ranks has resulted. In the examples shown above, We notice that the effect of random simulation on scores are usually less than 0.5 points. However, in most cases, the score gap between two divers obtaining consecutive ranks are more than 1 points, much more than the change caused by simulation.

According to the samples, the random simulation does not have an immense effect on the final ranking, and the outcomes of different trimming methods do not differ much from each other. In fact, the rank does not change after simulation in most circumstances. Together with the fact that randomization itself has some uncertainty, we cannot conclude that the judges’ potential nationalistic bias benefited their countrymen. Since the potential nationalistic bias of the “suspicious judges” does not make an evident effect on the final results of the diving competitions, according to our hypothesis, the data could reflect the athletes’ real performance in most cases.

3.3. Nationalistic bias test

As mentioned in part two, we picked a list of “biased judges” whose p-values from our two-sample t-tests and assumed that all of them exhibit extreme favoritism towards their countrymen *at the same time*. Since only the matching cases (same nationality between diver and judge) are of our interest, we selected only the subset of dives where 1. the judges include one or more of the biased judges and 2. The divers include one or more who are from the matching country. In these selected cases, the matching divers’ scores were assigned to a maximum of 10, while the non-matching divers’scores remained unchanged. We then performed the four scoring methods on each of the rounds selected, calculating new total scores and generating new ranks.We take the preliminary round of men’s ten-meter platform in the 2020 Tokyo Olympics as an example, and use the Mean method:

Preliminary of Men’s 10-meter platform in Tokyo, employing Mean method

##	Diver	Match	O_Score	O_Rank	Score_New	Rank_New
## 1	YANG Jian	**	181.52143	1	183.52	1
## 2	CAO Yuan	**	175.56429	2	177.75	2
## 3	BONDAR Aleksandr		170.94286	3	170.94	3
## 4	DALEY Thomas		151.63571	4	151.64	4
## 5	ZSOMBOR-MURRAY Nathan	**	148.07143	5	150.43	5
## 6	SEREDA Oleksii	**	145.54286	6	147.96	6
## 7	WOO Haram	**	142.89286	7	142.89	7
## 8	ROUSSEAU Cassiel		141.23571	8	141.24	9
## 9	VILLARREAL TUDON Andres Isaac	**	136.65714	9	141.87	8
## 10	VILLA CASTANEDA Sebastian		134.92143	10	134.92	11
## 11	BARTHEL Timo	**	134.35714	11	138.37	10
## 12	LOSCHIAVO Brandon		133.62857	12	133.63	12
## 13	QUINTERO DIAZ Rafael		133.25000	13	133.25	13
## 14	MINIBAEV Viktor		131.25000	14	131.25	14
## 15	WINDLE Jordan		130.91429	15	130.91	15
## 16	FIGUEREDO PEREIRA Kawan		124.74286	16	124.74	17
## 17	TAMAI Rikuto		124.68571	17	124.69	18
## 18	KIM Yeongtaek	**	122.63571	18	122.64	19
## 19	WIENS Rylan	**	122.46429	19	129.14	16
## 20	SOUZA FILHO Isaac		113.00714	20	113.01	21
## 21	EIKERMANN GREGORCHUK Jaden	**	110.67857	21	115.71	20
## 22	ARIZA Oscar		108.45714	22	108.46	23
## 23	ISHAK Mohab		105.20714	23	105.21	24
## 24	GARCIA NAVARRO Ivan	**	105.13571	24	110.02	22
## 25	NISHIDA Reo		104.91429	25	104.91	25
## 26	CHAN Jonathan		104.25000	26	104.25	26
## 27	FRICKER Samuel		102.85000	27	102.85	27
## 28	WILLIAMS Noah		102.15000	28	102.15	28
## 29	ROSSET Matthieu		92.59286	29	92.59	29

Figure 15. Prelim of Men’s 10-meter in Tokyo employing mean.

‘O_Score’ and ‘O_Rank’ stands for original rank, while ‘Score_New’ and ‘Rank_New’ stands for new score and new rank. The divers are arranged in descending order of their original rank position. In the second column (‘Match’), divers labelled with “**” have had at least one judge who is from the same country in the round. For these divers, there is most likely an improvement in the new score because their matched judges are assumed to be extremely biased towards them and awarded them the full points. When we compare original rank with new rank, we can see that positions 8th to 11th, and 16th to 24th has shifted considerably, with eleven divers shifted by one position, three divers shifted by two position, and three divers shifted by three positions. This drastic change might be explained by the small differences between the original scores of these divers. For instance, Villa Castaneda Sebastian (10th) and Barthel Timo (11th)'s scores were differed by only 0.56. By proportion, nationalistic biases would be manifested more among divers with smaller score differences. Noticeably, since some rank positions before the eighteenth are shifted in this preliminary game, the advancement of divers to the semi-final round of the competition will be affected.For example, Wien Rylan, originally at position 19th, became position 16th and was able to advance to the semi-final round with the case of extreme nationalistic bias. Kim Yeongtaek at position 18th, on the other hand, became position 19th and failed to advance. Apparently, by keeping all the scores, the mean method does a poor job in eliminating supportive bias. Then, what about the other three trimming methods?

To zoom in into each individual case, for each scoring method, we created an arrow graph which directly compares the original rank with the new rank in each round of event, showing an arrow if there is a position change. Take Tokyo's M10mPF Semi as an example:

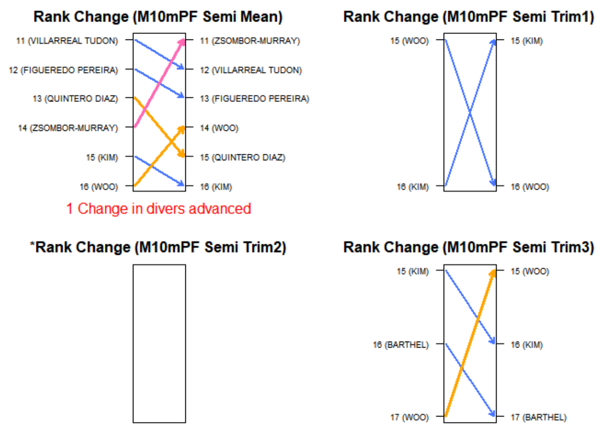

Figure 16. Rank change of each trimming method of M10PF in Tokyo.

The graph at top left corner shows changes in ranking in M10mPF Semi's competition if we were to use the mean method. A blue arrow indicates shift of one position, an orange arrow indicates a shift of two positions, and a pink arrow indicates a shift of three positions. "One change in divers advanced" in red text means that there is one change in the top 12 positions (12 divers in each semi-final round are advanced to final). The graph showing result of trim2 has a '*' in its title, which indicates that this method was adopted in that year of Olympics. We are delighted to see that the optimal scoring method in this event, trim2, was used, reducing the effect of nationalistic bias to the most extent. However, is this always the case?

Let's examine the preliminary round of the same event:

DOI: 10.54254/2755-2721/6/20230492

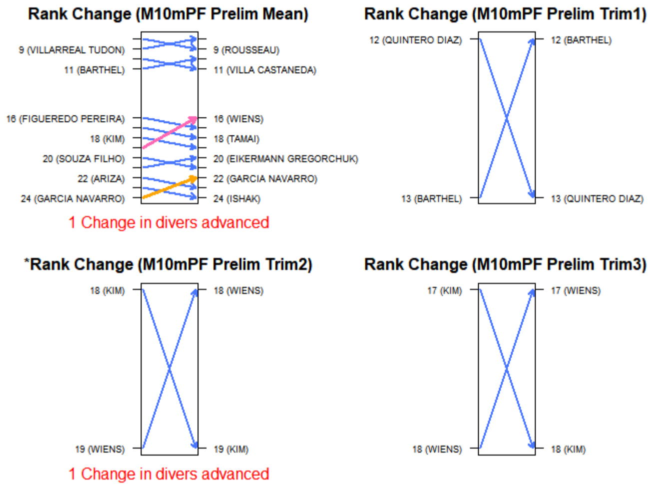

Figure 17. Rank change of each trimming method.

We can see that although all the trimming methods are much better than the mean method and result in only two changes in positions, trim2 is the most interesting case: there is a change in divers who would advance to the semi-finals. Kim, originally at position 18th, became 19th in the case of extreme nationalistic bias and failed to advance. Wiens, with the help of a biased judge, improved his score and became position 18th, advancing to the semi-final round. Note that trim2 was the scoring method chosen in that competition! This critical change in rank affects two divers, not one. This possible change in advancement matters tremendously to these two divers, because one of them could have advanced instead of the other, depending *solely* on the scoring system used, with all other factors kept constant. Hence, we need to be careful to conclude that trim2 is the best method – two divers' career could be at stake.

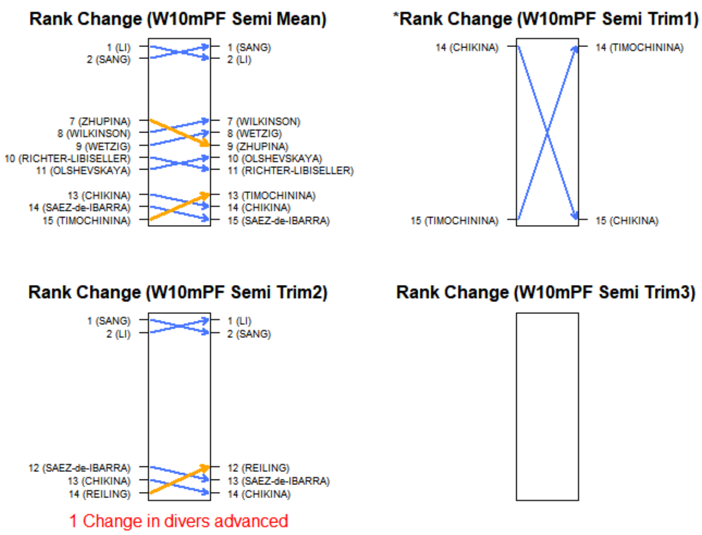

Figure 18. Rank change of each trimming method.

Let's zoom into some specific cases in Sydney. Is trim1, the actual method used, the best? Let's look at the results from W10mPF's preliminary round: Among trim1, trim2, and trim3, trim 3 is the optimal method here, it handles nationalistic bias and results in no change in ranking. Trim2 is the worst method because it not only causes four divers to change by one position and one diver to change by two positions, but it also causes one change in divers who advanced to the final round. This, again, affects advancement of two divers, not one. Trim1, the actual scoring system used, is in the middle in terms of its effectiveness in handling nationalistic bias.

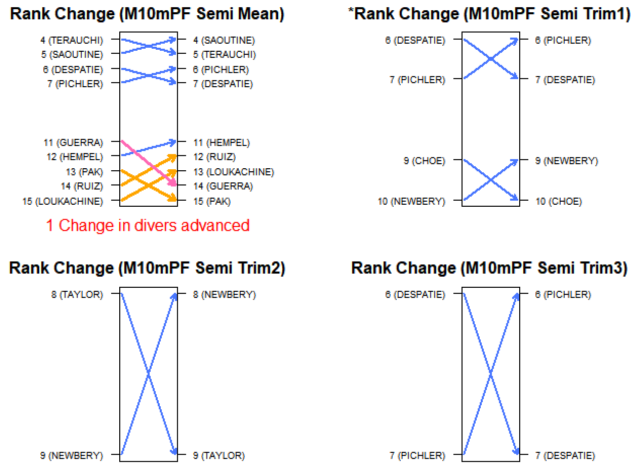

Figure 19. Rank change of each trimming method.

In M10mPF's Semi, trimi2 and trim3 appeared to be better than trim1, since only two divers swapped their rank positions in trim2 and trim3, and four divers' positions were affected in trim1.

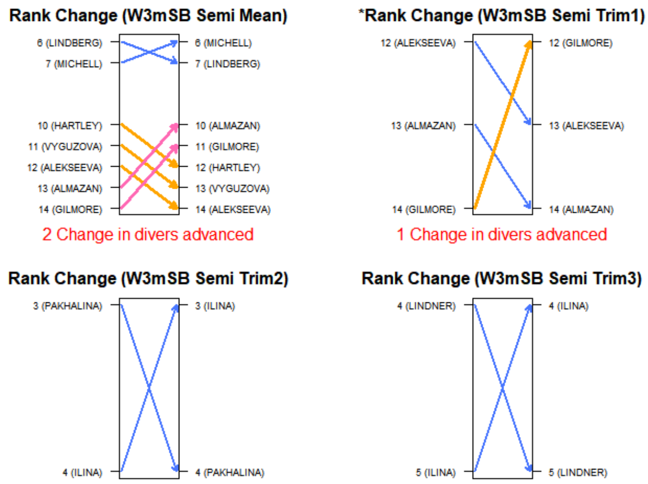

Figure 20. Rank change of each trimming method.

In W3mSB' Semi, trim2 and trim3 were also superior as compared to trim1, with smaller changes in ranks resulting from extreme nationalistic bias. The other arrow graphs are attached in the appendix. To obtain a better overview of the results, we plotted 8 histograms of a number of changes in positions (y-axis) against frequency (x-axis) for each event and repeated each graph with different trimming methods (mean, trim1, trim2, trim3) to get 32 (8*4) graphs. To obtain a general view of the entire picture, we pasted all the graphs together. For each table of histograms, the rows represent the eight events, and the columns represent the four trimming methods.

Let's first observe the 8*4 graph of Sydney:

DOI: 10.54254/2755-2721/6/20230492

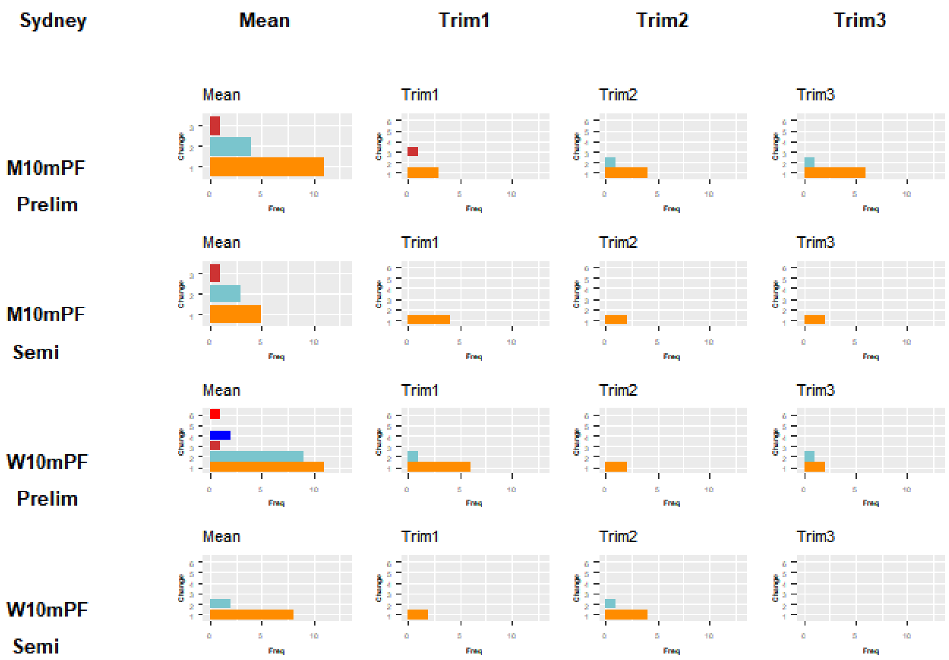

Figure 21. Result of Nationalistic Bias at Sydney 2000 for M10mPF and W10mPF.

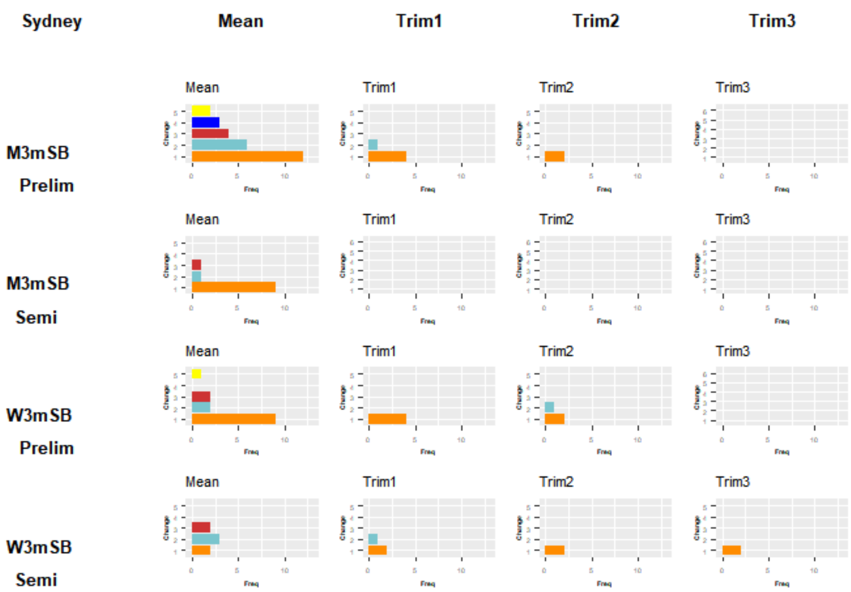

Figure 22. Result of Nationalistic Bias at Sydney 2000 for M3mSB and W3mSB.

In Sydney's M10mPF Prelim (row 1), if we use the mean (column 1), there are eleven divers whose rank positions have shifted by one (either up or down), four divers whose positions have shifted by two, and one diver whose position has shifted by three. If instead, we use 'trim1' (column 2), three divers' positions shift by one, and one diver's position shift by three. The same interpretation applies for 'trim2' and 'trim3' (the 3rd and 4th column). If we take a glimpse at the entire table, it is not clear-cut that

DOI: 10.54254/2755-2721/6/20230492

which method has resulted in the least change in ranks across *all* the events. However, this table does provide an overall perspective on which method is the best in *each* event. Undoubtedly, the mean method is the worst in removing the biased scores because it resulted in the greatest change in ranks. Intriguingly, when we compare amongst the other three methods, the table shows a mixed result. In Sydney's dataset, trim3 was the most desirable method in three cases (W10mPF Semi, M3mSB Prelim and W3mSB Prelim), trim2 was the best method in the two cases (M10mPF Prelim and W10mPF Prelim), trim2 and trim3 were equally desirable in handling bias in three cases (M10mPF Semi and W3mSB Semi). In M3mSB Semi, all trimming methods were equally effective in removing bias. It appears that trim3 is the best, trim2 is slightly worse than trim3, and trim1 is the worst. Unfortunately, trim1 was the actual method used in the competition.

Then compare Sydney's 8*4 graph with Tokyo's:

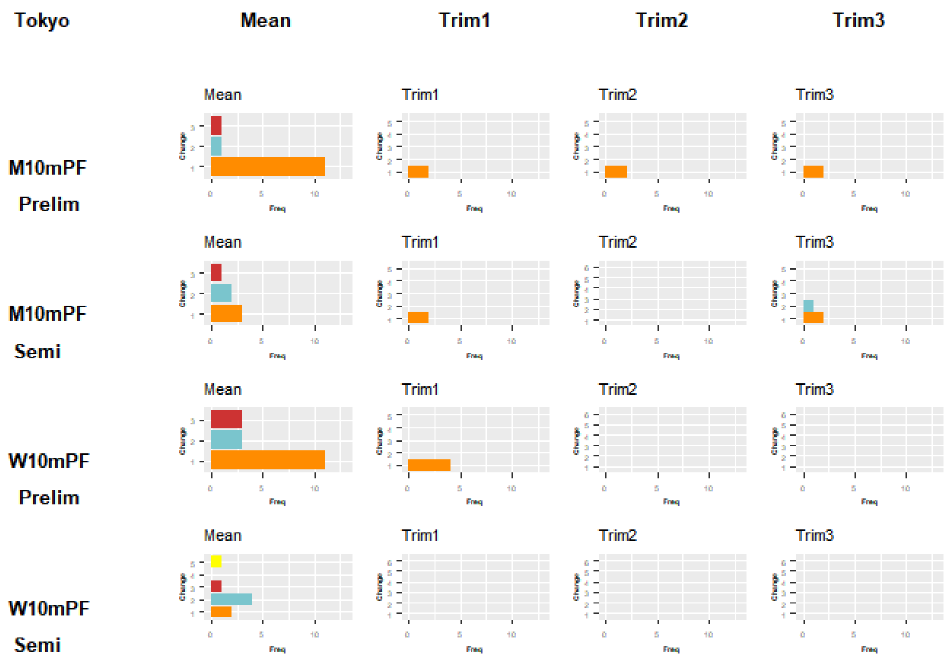

Figure 23. Result of Nationalistic Bias at Tokyo 2020 for M10mPF and W10mPF.

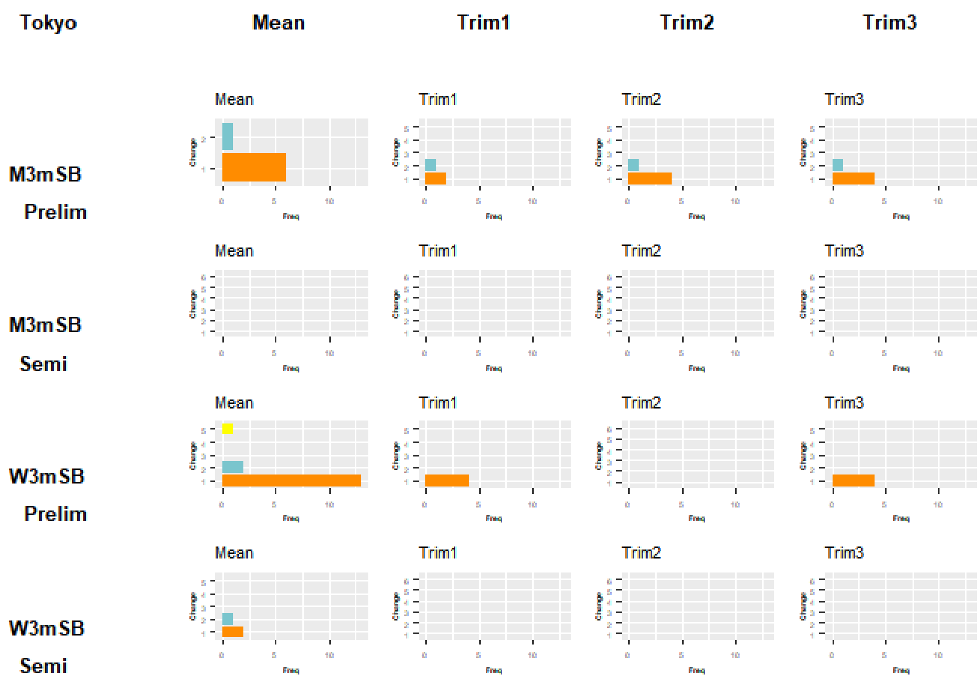

Figure 24. Result of Nationalistic Bias at Tokyo 2020 for M3mSB and W3mSB

Tokyo's dataset presents a slightly different story. Trim2 reduces the bias effect to a greater extent in three cases in Tokyo's dataset (M10mPF Semi, W10mPF Prelim, W3mSB Prelim), trim1 is the most ideal method in two cases (M10mPF Prelim, M3mSB Prelim), in the other three cases all trimming methods were equally effective in dealing with bias and resulted in no change in ranks. Trim2, the method used in this Olympics, appears to be the best in this case, followed by trim1, and then trim3. To sum up, the Nationalistic Bias Test presents us with a mixed result with which scoring system is the most effective in reducing nationalistic bias. What about another potential bias that some judges might have, the competitive bias (bias that causes them to give lower scores to their countrymen's competitors)?

3.4. Negative bias towards competitors

In the real Olympic Games, except for the nationalistic bias which benefits those who share the same nationality with some judges, it is also reasonable to speculate that some judges may take advantage of their position to influence the scoring of other divers who appear to be the main competitor of the judges' own countrymen. They might indirectly contribute to lifting the rank of their countrymen by scoring their competitors with lower scores. In essence, this 'competitor bias' is the contrary form of the matching bias. To investigate the impact, a precise definition of 'competitor bias' is necessary in order to study it. Simply postulating that all the divers other than the ones sharing the same nationality with certain judge are competitors is unrealistic and unreasonable. A strong contender for the medal is certainly not just aiming to defeat the one struggling to enter the final, and a newcomer of the final round might not set the top of the podium as the goal. More realistically, each diver's competitors are the close competitors. For simplicity, we defined competitors for a diver to be the person immediately before and the person immediately after in the rank position. Then, for a judge to help his or her compatriot, we assume that he or she would score those two divers with lower scores.

However, how do the judges decide the close competitors of their compatriots? To answer that question, we need to first find the past ranking which judges might have used to choose their nations' competitors. There might be outer factors such as the politics or the past rank, but we would like to focus

DOI: 10.54254/2755-2721/6/20230492

just on the sport performance itself in that specific Olympic game. We reasonably assume that every four years, the list of divers and the ability of divers would change, making it impossible to determine the “targets” before the Olympics begins. Thus, we regard the result of the Prelim as the basis for judges to locate the competitors.

With the definition of 'competitor bias' and the way to locate the competitors, we are ready to manipulate the data and test the robustness of each trimming method against such bias. In the test, we first obtain the result of preliminary round for each event in both Sydney and Tokyo Olympics. Then, targets of each judge were determined based on this preliminary result. If these judges happen to judge their targets again in semi-final, their competitive bias would be shown by awarding only 5 marks to this target regardless of his or her performance. The result of each event in the Semi is compiled into a data frame containing the original rank and the new rank obtained by the manipulation. Subtracting the original rank from the new rank and only collecting the absolute value, we acquire the data containing the changes of rank. Plotting 16 graphs for 4 events of the semi-final round and 4 trimming methods in the format of a table (rows represent M10mPF, W10mPF, M3mSB, W3mSB and columns represent mean, trim1, trim2, trim3), we obtained two graphs for the two Olympics:

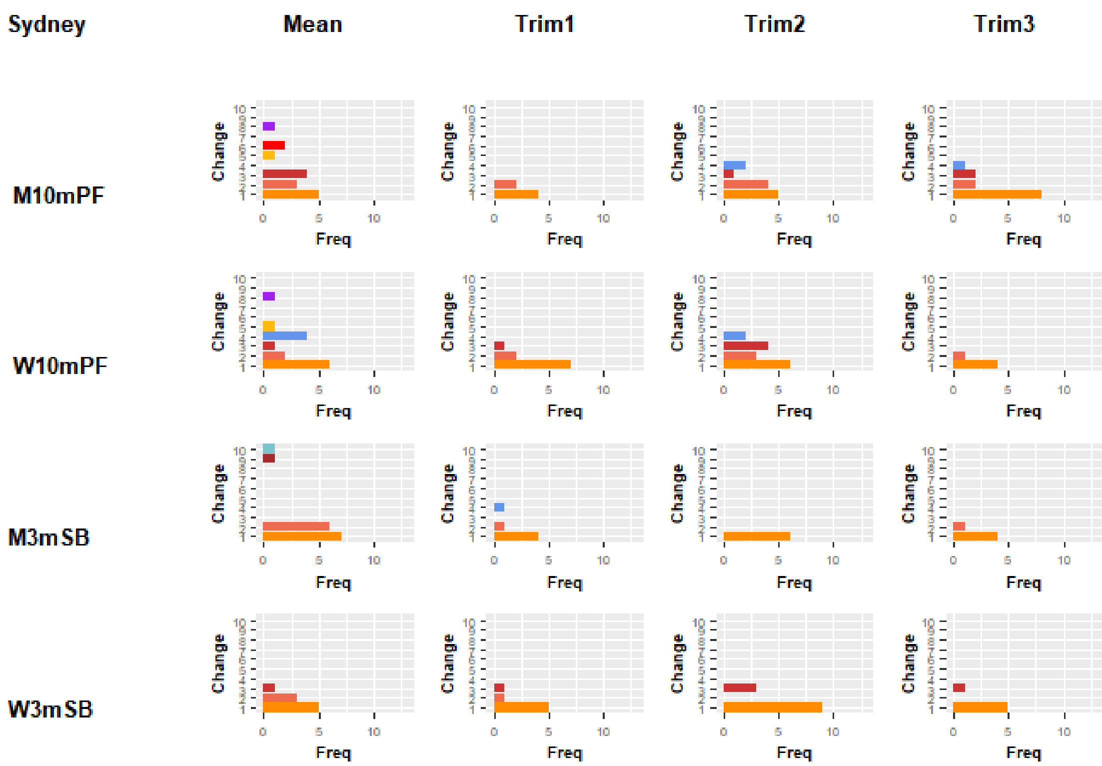

Figure 25. Result of Negative Bias at Sydney 2000.

DOI: 10.54254/2755-2721/6/20230492

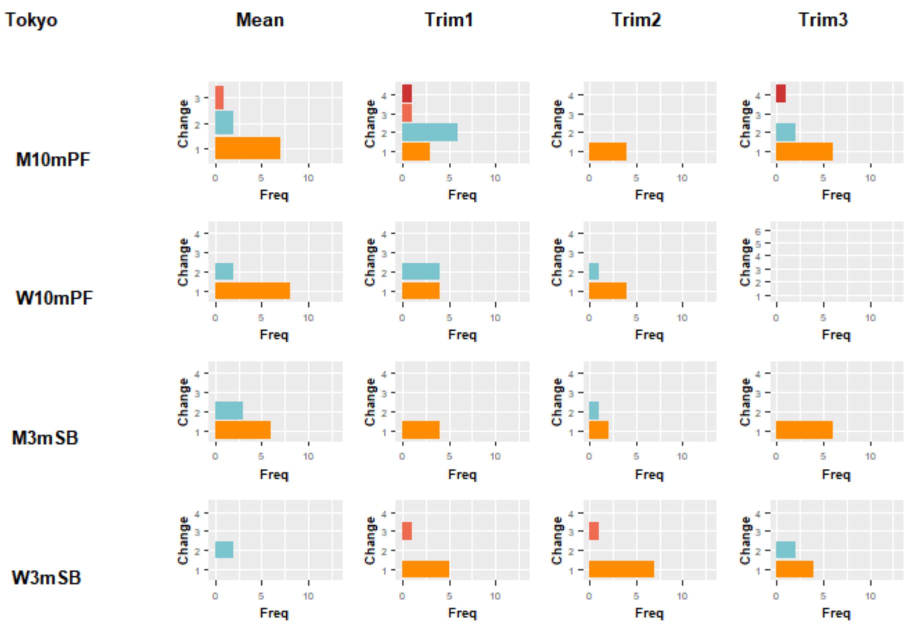

Figure 26. Result of Negative Bias at Tokyo 2020.

For each graph, the subplot represents the changes in rank for the specific event, round, and the trimming method. The y-axis stands for the number of position change, and the x-axis stands for the number of divers affected. Generally, we can observe that the changes of rank in Tokyo are less than those of Sydney under any trimming method. This phenomenon might imply two things about the original scores. First, since we set the score to 5, the distribution of the original scores in the Semi for any event is more condensed around or closer to the 5 than that of Sydney, which causes a smaller degree of change. Such discrepancy could be visualized by the graphs below.

Second, the differences of scores between adjacent divers are smaller in Sydney than those in Tokyo, resulting in Sydney's scores to be more sensitive to the bias. This might happen because the performance of divers in Sydney is relatively closer.

Examining each 4*4 graph, it is hard to conclude anything by comparing vertically, implying that the event does not predict the degree of change. However, in most cases, except for the M3mSB in Sydney, the mean method is always more sensitive than the other three trimming strategies. This is predictable, because without trimming, the change to the score will be directly reflected on the total score. One remarkable observation is made from the Trim1 method in Sydney. Recall that in Sydney's Olympics, Trim1 is the official method for scoring. However, except for the mean method, Trim1 is evidently the second most sensitive trimming method. It might indicate that Trim2 was probably not the best trimming method for Sydney Olympics. Both Trim2 and Trim3 shows their advantage to similar extent.

As for Tokyo, beside the mean method, which has apparent disadvantage, we cannot tell a better trimming method among the other three simply through the comparison of change. As mentioned before, the distribution of scores decides the sensitivity of each trimming method to a degree. Considering both phenomena, it would be reasonable to conclude that under different distributions of scores, different trimming methods will show different adaptive ability to the bias, and there might be no method staying the best.

DOI: 10.54254/2755-2721/6/20230492

4. Conclusion

In the real world, there are various kinds of bias affecting the results of sports. As it is very challenging to eliminate these biases, we shall seek to reduce their effects to render the scoring and ranking process with objectivity out of respect for every participant. In this paper, our main focus is on the nationalistic bias. We explored the degree of nationalistic bias in the Olympic Games and test the ability of four trimming methods to reduce the impact. In the Random test, we replace all the scores given by the matching judges by one ofthe other six scores in the same dive. This process, after repeated on four trimming methods, reveals us little or no change to the rank in almost all events. This conveys the good news that the original ranking are able to reflect the sports performance of divers to a large extent. In the nationalistic bias test, we maximize each judge's bias by setting scores given to their countrymen 10, practicing the pressure test on each trimming method. While the Mean method demonstrates lowest robustness, no one of the other three methods shows a permanent advantage. Similar conclusion is obtained from the third test, the competitor bias test, in which we define the competitors and score 5 to those divers who are targeted as the competitors. However, Trim2's performance is relatively stable across most circumstances. This may imply that the shift of scoring system from Trim 1 in Sydney to Trim 2 in Tokyo is reasonable.

Although there are numerous intriguing implications that can be drawn from this research, it is admitted that there are inherent limitations in our methodology that could have interfered with the accuracy and reliability of the results. To begin with, we were only able to acquire the 2000 and 2020 Summer Olympics' diving data, which are only *two* events separated by *twenty* years of time. With more data, we might be able to conduct our robustness tests of scoring system on other years of Olympic games or even other diving competitions (like those hosted by FINA and the domestic competitions). To further the research, the very first step is to collect more data as comprehensive and detailed as possible.

Furthermore, in the Competitive Bias Test, due to limited availability of other rounds' data, we choose to analyze only the semi-final round's results by using preliminary's scores as the basis for locating competitors. However, in the real world, scores of Olympics' pre-events and other non-Olympics diving competition might also allow the judges to gauge the ability of competing nations' Olympic participants and predict their countrymen's competitors. Moreover, we set the scores given to each judges' targets to 5 for ease of comparison between original rank and final rank (setting the scores to 0 would result in too drastic a change and is unrealistic). However, due to difference in quality of each diver's performance, the score 5 may lie at the bottom, middle or even top of the list of scores given by all the judges. This means that our intension of inserting a competitive bias may not be successful in some cases, where 5 is close to or higher than the mean of other six scores. For example, 5 is a score low enough for a dive with mean 8 but is not very low for a dive with mean 5.5. Also, our assumption that *all* judges exhibit competitive bias is very likely not true in the real world. Accuracy of our results would be improved considerably if we could use statistical tests to generate a list of potentially biased judges for competitive bias and apply the change of scores to only those judges.

Nevertheless, research that compares effectiveness of scoring systems in handling judging bias is still in the very early stage. To our knowledge, this is the very first research that study the robustness of different scoring system in the Olympic Games. Hence, any conclusions should be drawn with caution. Future study in this area should aim to gather abundant data in diving competition and run the tests on a larger sample.

To sum up, it is promising to see that that as compared to twenty years ago, Tokyo's 2020 Olympics diving adopted a more robust scoring system that handles extreme nationalistic bias more adequately. It is also intriguing to observe that there are fewer judges employed in Tokyo's 2020 Olympics diving who were determined to be 'nationalistically biased' by our two-sample t-test (6 judges, as compared to 12 judges in 2000 Sydney's Olympics diving). It is beyond the scope of this research to discover why nationalistic or competitive bias exists in certain judges. The possibility of their co-existence on the same judge might be an intriguing topic for future exploration.

4.1. Acknowledgement

Gu honglei, Zhou yicheng, Jia junran and Gao yuanji contributed equally to this work and should be considered co-first authors

References

[1] En.wikipedia.org. (2022). Sportsmanship - Wikipedia. https://en.wikipedia.org/wiki/Sportsmanship.

[2] Kramer RSS. (2017) Sequentialeffects in Olympic synchronized diving scores. R.Soc.opensci. **4**: 160812.http://dx.doi.org/10.1098/rsos.160812

[3] Campbell, B., & Galbraith, J. W. (1996). Nonparametric Tests of the Unbiasedness of Olympic Figure-Skating Judgments. Journal of the Royal Statistical Society. Series D (The Statistician), 45(4), 521–526. https://doi.org/10.2307/2988550

[4] Luedeker B, McGee M (2022) Relationship between judges' scores and dive attributes from a video recording of a diving competition. PLoS ONE 17(8): e0273374. https://doi.org/10.1371/journal.pone.0273374

[5] Clarey, C. (2002, May 1). 2 French officials suspended 3 years in skating scandal. The New York Times. Retrieved August 13, 2022, from https://www.nytimes.com/2002/05/01/sports/figure-skating-2-french-officials-suspended-3-years-in-skating-scandal.html

[6] Heiniger, S. and Mercier, H. (2021). Judging the judges: evaluating the accuracy and national bias of international gymnastics judges. Journal of Quantitative Analysis in Sports, 17(4), pp.289-305.

[7] John W. Emerson, Miki Seltzer & David Lin (2009) Assessing Judging Bias: An Example From the 2000 Olympic Games, The American Statistician, 63:2, 124-131, DOI: 10.1198/tast.2009.0026

[8] Grannan, C. (2016, August 1). How Is Diving Scored? Encyclopedia Britannica. https://www.britannica.com/story/how-is-diving-scored

[9] John W. Emerson & Silas Meredith (2011) Nationalistic Judging Bias in the 2000 Olympic Diving Competition, Math Horizons, 18:3, 8-11, DOI: 10.4169/194762111X12954578042812

[10] Ioc. (2022, July 7). Tokyo 2020 summer Olympics - athletes, medals & results. Olympics.com. Retrieved August 14, 2022, from https://olympics.com/en/olympic-games/tokyo-2020

[11] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistica Computing, Vienna, Austria. URL https://www.R-project.org/.