

Background:

- Nationalistic bias commands the most attention in newspapers and scholarly articles about judged competitions like diving, ice skating, gymnastics, ski jumping and power lifting.
- Other sources of bias are difficulty bias, gender bias, age bias, reputation bias, round bias, dive order bias, seeding bias and conformity bias
- Estimation of bias requires estimation of the true ability of an athlete, θ
- We propose an estimate for θ and calculate various types of biases using judges' scores as estimates of θ .
- We show that gender and age bias are minimal, while difficulty bias and round bias are present in the data.

Aims:

- Determine a reasonable estimate of diver ability given subjective scores for the same diver in multiple meets
- Examine the discrepancy between overall diver ability and scores for a particular meet given degree of difficulty and meet round.

Previous Work:

Measured diver ability by characteristics of dive, such as the height of the dive, distance of entry from the board, and height of the splash gathered from painstaking analysis of videos of divers⁵. We found ordinal rankings from judges' scores closely matched scores calculated from video evidence.

Methods:

- Scraped data from divemeets.com, a data base of high school, collegiate, and club diving for meets in the US.

11 rounds per meet (r)	261 meets (m)
Divers aged 14 – 19	1911 divers (i)
39105 dives	86 dives (d)

- Calculate diver "competency"^{1,3,4}, an estimate of diver ability, θ , using

$$C_i = \frac{1}{mr} \sum_{\text{all meets and rounds}} NetScore_{i,m,r}$$

where NetScore = sum of the middle 3 scores the judges in a round. There are typically 5 or 7 judges in a meet.

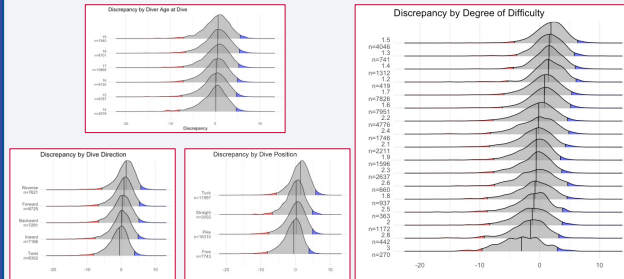
- Calculate "discrepancy", an estimate of judges' bias

$$D_{i,r,m} = NetScore_{i,m,r} - C_i$$

A positive discrepancy indicates that a judge gives the diver a greater score than his or her ability. A negative discrepancy indicates that a judge gives a diver a smaller score than his or her ability would indicate.

Results:

Ridgeline plots show the distribution of the discrepancy based on various factors. For each plot, the discrepancy is on the horizontal axis, the line in the middle of each plot is the median, the red area on the left is the 2.5%-tile and the blue area on the right is the 97.5%-tile. The ridges are aligned in ascending order of the median value.

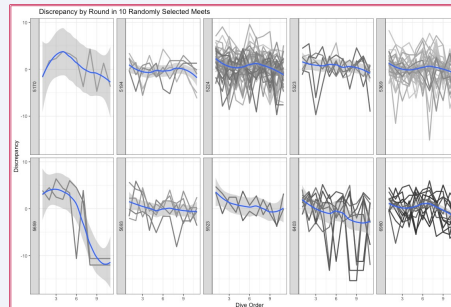


Mixed-Effect Model with Random Intercept for Diver and Discrepancy as the Response

Effect	Est	SE	T
Int	1.8	0.33	5.47
Round	-0.09	0.006	-14.79
Age	0.28	0.019	14.63
DD	-3.26	0.065	-50.04

Effect	Est	SE	T
Int	-0.33	0.37	-0.91
Round	-0.08	0.007	-11.56
Age	0.30	0.021	14.26
DD	-2.20	0.063	-34.84

Girls, n = 1045, ICC = 4%, REML = 107K Boys, n = 866, ICC = 2%, REML = 87K



Discrepancies for all divers within 10 randomly selected meets. The round number is on the horizontal axes and the discrepancy is on the vertical axis. Each black line represents discrepancy scores for a diver for each round. The blue line is the loess regression line.

Scoring Diving Competitions:

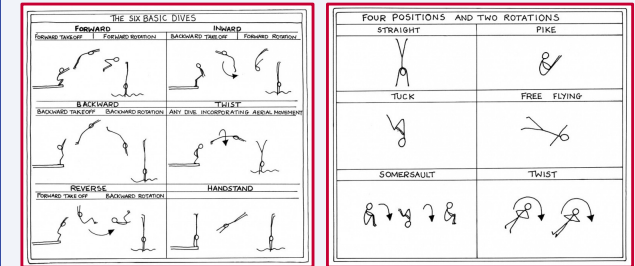
Each judge gives a score in the range of 0 – 10 in half-point increments

Divers must perform 2 of each 5 basic dives (excluding handstand) in any position and one free dive

Each dive is assigned a degree of difficulty (DD) by Fédération Internationale de Natation (FINA)

Award = Net score * Degree of difficulty (DD) per round

Sum of awards is the diver's Total (Meet) Score



Bias Exists:

- Age (experience)
- Degree of difficulty
- Positive bias in early rounds, negative bias in later rounds

Future Work:

- Adjust for dive/order effect: the same dive is done in different order by different divers.
- Use estimated bias to adjust scores for divers within a meet
- Compare new rank order to previous order to see if bias adjustment changes meet outcome.

References:

- Campbell, B., and J. W. Galbraith (1996). "Nonparametric Tests of the Unbiasedness of Olympic Figure-Skating Judgments." *The Statistician* 45: 521-6.
- Emerson, J. W. and S. Meredith (2011). "Nationalistic Judging Bias In The 2000 Olympic Diving Competition." *Math Horizons*, February.
- Emerson, J. W., M. Seltzer, and D. Lin (2009). "Assessing Judging Bias: An Example from the 2000 Olympic Games." *The American Statistician* 63 (2): 124-31.
- Heiniger, S. and H. Mercier (2021). "Judging the judges: evaluating the accuracy and national bias of international gymnastics judges". *Journal of Quantitative Analysis in Sports*, 17(4): 289 – 305.
- Luedeker, B. and M. McGee (2022). "Relationship between judges' scores and dive attributes from a video recording of a diving competition" *PLoS One*: <https://doi.org/10.1371/journal.pone.0273374>.