Wrestling with Byes at the EIWA Tournament: A Discrete Mathematical Analysis of a Sports Problem

T. S. MICHAEL[1]
Mathematics Department, United States Naval Academy, Annapolis, MD 21402
tsm@usna.edu
Version: 31 August 2005
Any updates will be posted at: `http://www.usna.edu/Users/math/tsm/`

Overview

The Eastern Intercollegiate Wrestling Association (EIWA) holds a championship tournament (the "Easterns") every year. About 13 colleges enter one wrestler in each of ten weight classes. In each weight class several byes must occur in the first round, and these are distributed among the top eight (seeded) wrestlers. The method currently in use distributes the byes at random among the eight seeds. This approach is fast, easy to implement, and perfectly fair to the individual athletes. However, since the random selections for byes are performed in each weight class independently, by sheer chance it sometimes happens that one team receives many more byes than another team, even though both teams have the same number of seeded wrestlers. This situation arose in 2004, when one Ivy League school had 9 seeds and received 0 byes, while a rival had 10 seeds and 7 byes. The two teams thus began the tournament under very different conditions—with a potential impact on the final team placings.

We show that a modification to the current scheme greatly reduces the chances of such an unbalanced situation occurring—while remaining fast, easy to implement, and fair to the individuals. The idea is to generate several complete draws, and then choose the "best" one. We use a precise mathematical notion (least squares) to measure how good a draw is.

This report is written with two audiences in mind. First, analogies, graphs, tables, and a small number of formulas are used to explain the advantages of the proposed method in a manner the non-mathematically inclined reader can follow. Second, some technical sidenotes are aimed at the more mathematically literate (or curious) reader. Enough of the wrestling context is provided so that a mathematician who is not familiar with the sport can appreciate the complexity of the problem.

[1]This report summarizes the results of the author's own investigations; no endorsement by the Navy wrestling program or coaches is implied or intended.

Summary

We propose a new method, BARRELS-(3), for selecting byes at the Eastern Intercollegiate Wrestling Association championship tournament. (BARRELS is an acronym whose meaning will be explained later.) The BARRELS-(3) method is specified below, and the anticipated consequences are sketched. A careful justification of the method, a discussion of the consequences, and computer-simulated "backtesting" using data from recent tournaments appear in the remainder of this report.

BARRELS-(3) Method[2]

We assume that the coaches have already selected the 8 seeds in each of the 10 weight classes, and that the number of byes in each weight class is known.

1: Find a random draw that distributes byes randomly and independently for all 10 weight classes in the usual manner.[3]

2: Compute the "inequitability penalty to teams" for the random draw.

3: Carry out Steps 1 and 2 three times to generate three complete random draws and their penalties.

4: Select the complete random draw with the lowest penalty.[4]

The inequitability penalty is a single number that measures the difference between the number of byes the teams *expect* to receive in a random draw and the number of byes the teams *actually* receive in a given draw. Smaller penalties correspond to draws that treat the teams more equitably (relative to the number of seeds)

[2]In the event of hardware or software failure, Steps 1 and 2 can be carried out manually. The random draws may be chosen by selecting pieces of paper from a hat, for instance, and the penalty for each draw is easy to compute by hand. (Only addition, subtraction, multiplication, division, and counting are involved.) Of course, a computer is preferred for both making the random selections and computing the penalty.

[3]In other words, select a draw by the method currently in use.

[4]In the unlikely event of a tie for the lowest penalty among several draws, select the draw generated later.

Consequences

• *Draws that assign byes inequitably among the teams (relative to their number of seeds) will occur much less frequently using the BARRELS-(3) method than under the current method.*

For instance, by one measure, a draw as disproportionate (or worse) than the one in 2004 can be expected about once every 29 years under the current method, but only once every 24,000 years under the proposed method.

• *Draws that assign byes very equitably to teams (relative to their number of seeds) will occur much more frequently than under the current method.*

For instance, a draw as equitable as the one in 2002 can be expected only once every 40 years under the current method, but will occur three times as often using the BARRELS-(3) method of draw.

• *Large gaps in the bye distribution among top teams occur less frequently.*

Under the current method of selecting byes, in years when *two* similarly situated teams are contending for the team title (e.g., Cornell and Lehigh in 2005), there is a greater than 25% chance that one team receives *at least three more byes* than the other.

Under the proposed BARRELS-(3) method, that chance drops to less than 16%.

When *three* similarly situated teams are contending for the team title (e.g., Cornell, Lehigh, and Penn in 2004), one team receives *at least four more byes* than another about once every 4 years under the current method.

Under the proposed BARRELS-(3) method, that disparate occurrence drops to less than once every 10 years.

• *A seeded wrestler in a weight class with 4 byes has a 50% chance exactly of receiving a bye under the current method. In the proposed method the probability of a bye falls within a narrow range centered at 50%. The probabilities fall between 49.5% and 50.5%. (Similar assertions hold for weight classes with other than 4 byes.)*

These tiny deviations in individual bye probabilities are the only geniune drawback to the proposed BARRELS-(3) method. Someone's decision to support or reject the BARRELS-(3) method should be based on his answer to the following question: The BARRELS-(3) method is substantially more likely to produce draws that treat the teams more equitably (relative to number of seeds); are the tiny deviations in individual bye probabilities an acceptable price for the equitable treatment of teams?

In most cases the individual bye probabilities are actually restricted to a narrower range (about 49.75% to 50.25%). Moreover, it is extremely difficult to predict or determine each wrestler's probability of a bye within this narrow range; no individual can reasonably claim his probability of a bye differs from 50% by an amount he can quantify. Even determining whether an individual's bye probability increases or decreases from 50% is difficult.

• **In a nutshell:** There is a trade-off between perfect fairness to individual wrestlers and an equitable treatment of the teams with respect to distribution of the byes at the EIWA tournament. The current method is perfectly fair to individuals, but can treat the teams inequitably or very inequitably. The proposed method gives up a very tiny amount of fairness to individuals and achieves a far more equitable treatment of teams. The deviation from perfect fairness to individuals is so small that it is difficult to measure even with a computer. Theoretically, it would take tens of thousands of years for the lack of perfect fairness to individual wrestlers to be noticed. In practice, an individual wrestler is far more likely to be treated unfairly by other matters inherent in conducting a tournament. For instance, the coaches are unable to seed the wrestlers with perfect accuracy in each weight class. Also, the clocks are not stopped and started with complete accuracy during matches.

CONTENTS

1. Background
2. The BAR Method and the 2004 Draw
3. Bye-Spreads and Computer Simulations
4. Fairness: What Does It Mean?
5. Equitable Treatment of Teams
6. The Inequitability Penalty for Teams
7. What Happened in 2002?
8. Least Squares
9. Least Squares and the Penalty
10. There is No Perfect Method to Select Byes
11. Randomness and Proportionality
12. Forced Equitability is Unfair to Individuals
13. Time-Out: A Game of Darts
14. The BARRELS Method
15. Are We Using Quotas?
16. BARS and BARRELS: Backtesting
17. The Individual Bye Probabilities for BARRELS-(3)
18. How Big Is 0.5%
19. Confidence Intervals: 40,000 Years Explained
20. FAQ
21. Recapitulation

1 Background

The Eastern Intercollegiate Wrestling Association (EIWA) currently consists of 13 teams: Brown, Columbia, Cornell, Harvard, Princeton, American, East Stroudsburg, Franklin & Marshall, Lehigh, and Rutgers Universities; the University of Pennsylvania; and Army and Navy. A conference championship tournament has been held every year for decades. The member colleges enter one wrestler in each of ten weight classes. Several byes must occur in the first round of each weight class, and these are distributed among the top eight (seeded) wrestlers. The method currently in use distributes the byes at random among the eight seeds in each weight class.

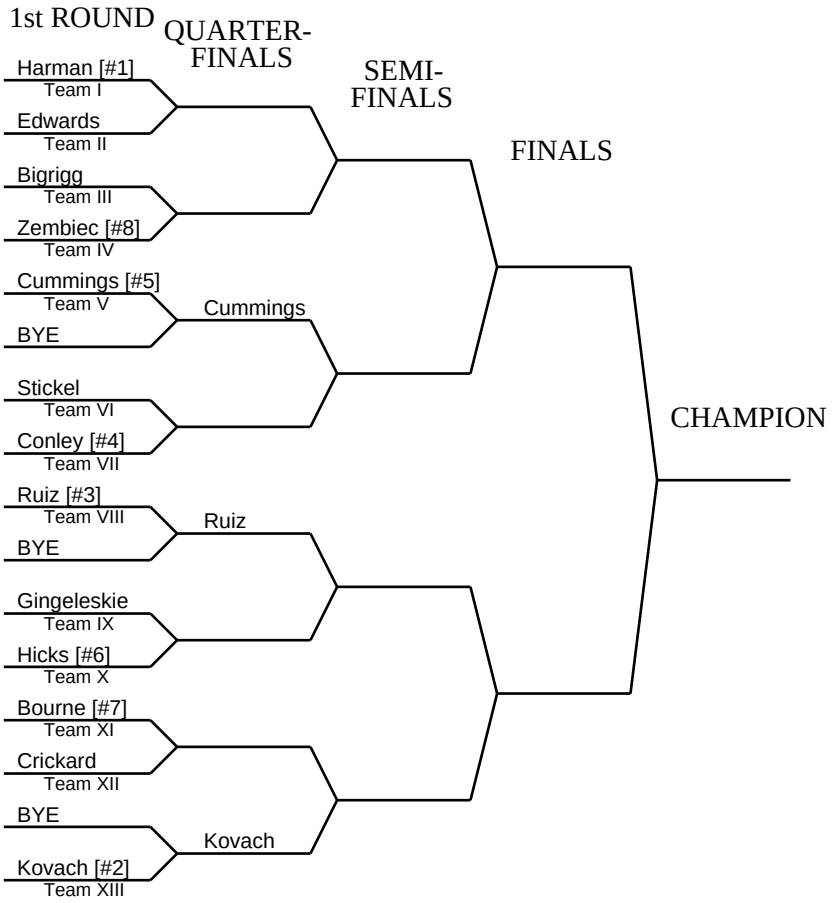

Figure 1: A bracket for a weight class with a draw that includes 3 byes.

Figure 1 shows a typical *bracket* for one weight class, including the *draw*–the pairings of first-round opponents—and an assignment of byes. The eight seeded wrestlers are placed far apart from one another in the bracket in the usual manner so that if no upsets occur, the quarterfinals will see a match-up between the #1 and #8 seeds; the #2 and #7 seeds; and so on, culminating in a championship match between the top two seeds. In Figure 1

the random byes evidently were given to the #2, #3, and #5 seeds; these wrestlers have already been advanced from the first round to the quarter-finals in the bracket. In most years some weight classes have more than three byes because not every team is able to field an entrant.

We should also mention that a wrestler who is defeated (before the championship match) continues wrestling in a separate *consolation bracket* to determine third through sixth places. The details of the consolation bracket are not important to us here.

In this study we are concerned solely with the process of selecting the byes from among the eight seeded wrestlers. This is the essential problem in determining the draw, and the one with the most interesting mathematical slant. In a weight-class with three byes, each of the eight seeded wrestlers should have a 3-out-of-8 chance (37.5%) of getting a bye; the current method, which assigns byes randomly, achieves this goal perfectly. Unfortunately, since the random selections for byes are performed in each weight class independently, by sheer chance it sometimes happens that one team receives many more byes than another team even though both teams have about the same number of seeded wrestlers.

Team Points

An individual wrestler earns points for his team in three ways:

- **Placement points:** A top-six finish in a weight class earns team points: 12 points for first place; 10 points for second place; ... .
- **Bonus points (Margin of victory):** Briefly, a win by major decision; technical fall; or pin/forfeit, respectively, earns 1.0, 1.5 and 2.0 team points.
- **Advancement points:** A wrestler receives 1 point for advancing to the next round in the championship bracket. However, a wrestler with a first-round bye must actually win his quarterfinal match to receive the 1 advancement point for reaching the quarterfinal.

The random assignment of byes influences the potential for bonus points. A team that receives a disproportionately large number of byes has fewer opportunities to earn bonus points in the many mismatches that occur between seeded and unseeded wrestlers in the first round. In 2004, for instance, the random draw gave Penn 0 byes among 9 seeds, while Cornell got 7 byes among 10 seeds. The draw may have played a substantial part in Penn edging Cornell for second place in the team race. (We should point out that the issue of whether receiving more byes is a disadvantage as far as team-scoring is not entirely clear-cut. Advantages of a bye include decreased chance of injury to a wrestler and more energy for subsequent bouts.) In any case, it is reasonable to prefer a method of selecting byes that is less likely to position two similar teams in such different circumstances at the start of the tournament.

2 The BAR Method and the 2004 Draw

Table 1: Seeds and byes for the 2004 EIWA tournament

TEAM	Leh	Pen	Cor	Rut	Arm	Nav	Har	Bro	Col	Pri	ESU	F&M	Ame	TOTAL
SEEDS	9	9	10	9	8	9	5	3	6	2	4	4	2	80
BYES RECEIVED	3	0	7	3	5	3	1	1	4	0	4	2	0	33

Table 1 shows the number of seeds and byes for each of the 13 teams in the 2004 EIWA tournament. Every year the coaches determine the eight seeded wrestlers in each weight class, and we will always take the number and distribution of seeds among the teams as given. Note that there are 80 total seeds in Table 1, as there should be. Our focus is on the distribution of the 33 byes in the draw, as shown in the last row of the table. There were 33 byes total, instead of 30 because some teams did not field entrants in all ten weight classes, and three of the weight classes had 4 byes instead of 3. Similar variations occur nearly every year, which will complicate our analysis somewhat.

In recent years the draw at the EIWA tournament is determined by selecting the byes randomly in each weight class, independently of the other weight classes. We refer to this as the *BAR method* (for **B**yes **A**ssigned **R**andomly).

An inspection of Table 1 reveals a confluence of what appear to be unlikely events. All 4 of ESU's seeded wrestlers received byes, whereas only 3 of Lehigh's 8 seeded wrestlers got byes. None of Penn's 9 seeds got byes, while 7 of Cornell's 10 seed got byes.

Question. *Just how unlikely was the 2004 draw?*

This is not a good question to ask because it is ambiguous; the answer depends on your point of view. In grappling with this question we'll introduce some basic ideas of probability and discrete mathematics.

Let us ask ask some concrete questions first.

The Number of Draws

Question. *How many draws were possible at the 2004 tournament altogether?*

Answer: The total number of draws is

$$\underbrace{56 \times 56 \times 56 \times 56 \times 56 \times 56 \times 56}_{7} \times \underbrace{70 \times 70 \times 70}_{3} = 56^7 \times 70^3.$$

In case you're curious, here's the reasoning: There are 56 ways to distribute[5] 3 byes among 8 seeded wrestlers in seven of the weight classes—giving us the seven factors of 56—and there are 70 ways to distribute 4 byes among 8 seeded wrestlers in the other three weight classes—giving the three factors of 70.

So the total number of possible draws is a huge number. We'll call it H. As an approximation

$$H = 56^7 \times 70^3 \approx 6 \times 10^{17}.$$

The large magnitude of H is difficult to comprehend. One way to think about conducting the BAR method of selecting byes is to print each possible draw on a single sheet of paper. There will be H such sheets, and we could stack them on top of one another to form a "deck of cards." Then we could select one piece of paper at random from the deck to determine the draw for the EIWA tournament. One slight difficulty arises. The deck of all possible printed brackets would be about 30 billion miles high! For comparison purposes, the distance from Earth to Pluto is less than 5 billion miles.

The upshot: Any proposed method of selecting byes that requires us to examine all H possible draws can be immediately dismissed as impractical. Even a powerful computer would have difficulty generating all draws and manipulating them in some mathematical manner to select the "best" one somehow. The number H is way too big![6]

We may now provide one (not very satisfying) answer to our earlier question: How unlikely was the 2004 draw? Well, it is just one of a total of H draws, each of which is equally likely. So the particular draw can be expected to occur just one every 6×10^{17} years. This answer is not helpful, however.

Some Team Probabilties

In the BAR method each of the 8 seeded wrestlers in a weight class with 4 byes has a 50% chance (4-out-of-8) of receiving a bye. In a weight class with 3 byes, the chance is 37.5% (3-out-of-8). We use these facts repeatedly in answering the questions below.

Question. *All 4 of ESU's seeds received byes in 2004. What was the probability of that?*

[5] You can list all 56 possibilities directly. Or if you know about Pascal's Triangle, use the binomial coefficient formula $\binom{8}{3} = 56$. Similarly, $\binom{8}{4} = 70$.

[6] The astute reader may have noticed that we have not truly accounted for *all* possible draws. Once the byes have been selected in each weight class, the wrestlers without byes must be paired up with one another for first-round bouts. There are about 3×10^{36} draws if these pairings are accounted for. (You can figure out how many galaxies the deck of draws would reach.) But we are only concerned with the selection of byes, not the pairings, and so we only need to worry about the "much smaller" number H.

Answer. The probability is about 3.5%. In other words, if the draw were conducted under the same circumstances repeatedly, ESU would get 4 byes about once every 29 years.

> **Reasoning:** The probability is $(50\%)(50\%)(37.5\%)(37.5\%) \approx 3.5\%$.
>
> Two of ESU's seeds were in weight classes with 4 byes—each with a 50% chance of getting a bye— and two are in weight classes with 3 byes–each with a 37.5% chance of a bye.

Question. *None of Penn's 9 seeds received byes in 2004. What was the probability of that?*

Answer. The probability is about 0.75%. In other words, if the draw were conducted under the same circumstances repeatedly, Penn would get 0 byes about once every 134 years.

> **Reasoning:** By similar reasoning as above, the probability is $(50\%)^3(62.5\%)^6 \approx 0.75\%$. We use 62.5% here because there are 5-out-of-8 chances of not getting a bye for each of Penn's 6 wrestlers in the 3-bye weight classes.

Question. *Of Cornell's 10 seeds, 7 got byes in 2004. What was the probability of Cornell getting* at least *7 byes?*

Answer. The probability of at least 7 byes is under 5%. (The computation is more difficult and is omitted.)

Note that we asked for the probability of *at least* 7 byes— not exactly 7 byes—because the answer is more meaningful to us.

> Here's an analogy to clarify the last point. Suppose you are wondering if a coin is fair. You flip it 99 times. What is the probability you get *exactly* 50 heads? The answer is well under 10%. This should not be surprising, since we expect that 48 or 49 or 51 or 52 heads to also be reasonably likely. However, the probability of *at least* 50 heads is exactly 50%. In the Cornell situation, we are more curious about Cornell's receiving 7 *or more* byes than exactly 7 byes.

3 Bye-Spreads and Computer Simulations

We could go on asking and answering questions similar to those in the last section from each team's point view, giving ourselves a workout in the type of calculations one learns in a junior-level discrete mathematics or combinatorics course. However, what we really

want is an *overall* measure of how unlikely the 2004 draw was—not just the probability for one team's outcome. We'll mention two such overall measures: The first comes from a simulation, and the second from a *least squares* calculation. Simulations and least squares will play an important role in all our later work.

Bye-Spreads

In the 2004 EIWA tournament there were three legitimate contenders for the team title: Lehigh, Cornell and Penn. These teams had comparable number of seeds (9, 10, and 9, respectively), but received wildly different numbers of byes (3, 7, and 0, respectively). We say the actual draw for the 2004 tournament had a *bye-spread* of 7 for the three contending teams because one team received 7 more byes than another. Many wrestling fans were uncomfortable that the three contending teams would start the tournament under such different circumstances.[7]

What are the chances of such a large bye-spread in the number of draws? Although the exact probability is nearly impossible to compute, we may use a *computer simulation* to give a very accurate estimate. Here is how the simulation is performed. First, the computer is given the schools of all 8 seeds in all 10 weight classes. Second, the computer conducts a random draw, assigning the correct number of byes in each weight class independently at random—in precisely the manner in which the actual draw was performed. Finally, the computer finds the bye-spread (using just the three contending teams) for the random draw and records the answer. This process is repeated n times for some large number n.

Table 2 summarizes the results of such a simulation for n =100,000 draws. (A desktop machine executes the entire simulation in a few minutes.) We see that a bye-spread of 7 occurred only 671 times out of the 100,000 random draws. Also, a bye-spread of at least 7 occurred only about 800 times ($800 \approx 671 + 126 + 11 + 1$). With respect to bye-spread the actual draw for the 2004 tournament was rare—occurring less than 1% of the time. The cumulative frequencies in the last row are the "running totals" and give us a quick way to determine percentages for bye-spreads. For instance, the number 21,698 in the last row shows that about 21.7% of the draws had a bye-spread of 0 or 1.

We will make use of such computer simulations several times in this report. The number n does not need to be as large as 100,000 for most of these; a simulation with $n = 10,000$ works just fine.[8]

[7] The discomfort was greater in Ithaca, New York, than in Philadelphia.

[8] How do we know that n =10,000 works well? We could try several simulations with n =10,000, or try a simulation with a much larger value of n and see if we get essesntially different distributions in the bye-spread frequency. All of our computer simulations are backed up by many such experiments, some with n as large as 1,000,000; smaller values of n are used in the report for convenience with no threat to our analysis or conclusions.

Table 2: Distribution of bye-spread for a simulation of the 2004 EIWA tournament with n =100,000 random draws (BAR-method)

bye-spread	0	1	2	3	4	5	6	7	8	9	10
frequency	3577	18,121	27,216	24,425	15,654	7579	2619	671	126	11	1
cumulative freq.	3577	21,698	48,914	73,339	88,993	96,572	99,191	99,862	99,988	99,999	100,000

From Table 2 we also see that the *median* bye-spread is 3, and the *mean* bye-spread is 2.58. The median is just the "middle" one. If all 100,000 draws were listed in order from smallest bye-spread of 0 to the largest bye-spread of 10, then the 50,000th draw has a bye-spread of 3. You can determine the median by scanning along the last row of the table until you hit 50,000.

The mean is just the ordinary average. It's a litte more difficult to compute, but here are some details:

$$\text{mean bye-spread} = \frac{(3577 \times 0) + (18121 \times 1) + (27216 \times 2) + \cdots (11 \times 9) + (1 \times 10)}{100,000} \approx 2.58$$

On average, the bye-spread is 2.58. The fact that the actual bye-spread of 7 differs so greatly from the median of 3 and the mean of 2.58 confirms that the draw for the 2004 was a rare occurrence. In fact, we see from the last row of the table that a bye-spread of 7 or more occurs about 0.8% of the time—about once every 124 years.

An Overall Measure: The Penalty

The preceding discussion about bye-spreads focused on three teams and ignored the rest. Later on, we will define a single number, a parameter we call the *penalty*, that takes *all* teams into account. We'll give the precise mathematical formula for the penalty later; its computation requires only counting and basic arithmetic. Roughly, a larger penalty means that teams received byes inequitably relative to their number of seeds, while a small penalty means that the byes were in close proportion to the number of seeds.

Anticipating some of our later work, we now present Figure 2, a "BAR-chart" (pardon the pun) showing the result of a computer simulation of the BAR-method wth 10,000 draws. The horizontal axis is the penalty, which ranges from about 1 to about 70. The vertical axis is the frequency. For instance, the large peak shows that a penalty between 15 and 20 occurred over 3000 times (about 30%). The median penalty is just under 18, and the mean penalty is 19. (The median and the mean were calculated by the computer, and can not be read off from the rough BAR-chart.) The penalty for the actual draw for the 2004 EIWA tournament was 36.4, which falls in the far right tail of the plot, indicating a draw that treated the teams inequitably.

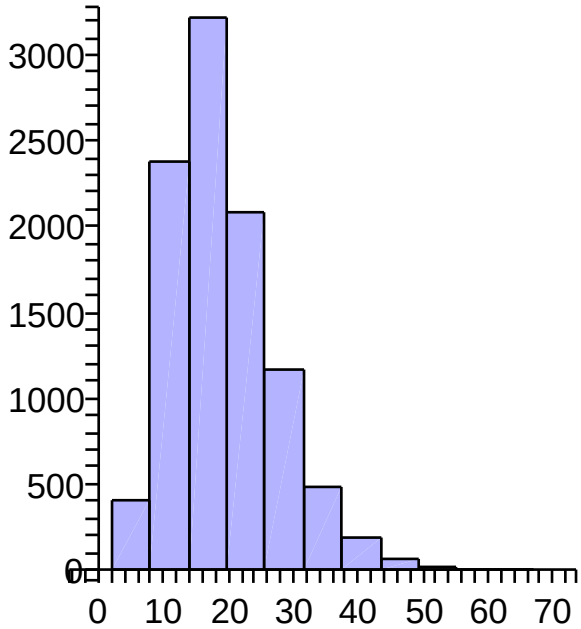

Figure 2: A frequency plot of the penalty based on 10,000 random draws (BAR-method) using the seeds for the 2004 EIWA tournament ($n = 10,000$; median $= 17.9$ mean $= 19.0$.) The actual draw for the tournament had a penalty of 36.4, which is far to the right.

4 Fairness: What Does it Mean?

What makes a method of selecting byes fair? There are two aspects of fairness here: fairness to individuals and equitable treatment of teams. The former notion is more important and easier to define–but some subjectivity occurs in our attempt to quantify it. We'll treat equitable treatment of teams in the next section.

Consider a seeded wrestler in a weight class with 4 byes. Then a method of selecting byes is *perfectly fair* to that wrestler provided he has exactly a 50% chance (4-out-of-8) of getting a bye. Of course, if there are 3 byes in the weight class, perfect fairness requires the wrestler to have 37.5% chance (3-out-of-8) of getting a bye. A particular method might be perfectly fair to some wrestlers, but not to others. Or it might be perfectly fair to all wrestlers in a weight class, but fail to be perfectly fair to some wrestlers in other weight classes. Or it might be perfectly fair to all wrestlers on one team, but not perfectly fair to some wrestlers on another team. The ideal is to have a method that is perfectly fair to all individual wrestlers.

Our concern with fairness to individual wrestlers stems from the following fact: The EIWA tournament is a qualifying meet for the national collegiate championships, the

pinnacle of college wrestling. Merely qualifying for the NCAA meet is an honor and a goal for many student-athletes. Only individuals can qualify, not whole teams. In recent years, the top three place-winners in each weight class at the EIWA tournament earn automatic berths to the national meet; several lower place-winners are granted wildcard berths by the league coaches. The BAR method places the seeded wrestlers on a perfectly equal footing going into the draw; each seed has the exact same probability of getting a bye in the first round as all the other seeds in the same weight class. A method of draw that, say, preferentially grants byes to Ivy League wrestlers would not be acceptable because individual wrestlers would have predictably different routes to earn berths to the national meet.

If a method is not perfectly fair to a wrestler, we may try to quantify the degree of unfairness. For instance, consider our seeded wrestler in a weight class with 4 byes again. Suppose that a particular proposed draw method gives him a 70% chance of getting a bye. This deviation from 50% is entirely unacceptable, and the method should be rejected outright.

What if the method gives the wrestler a 50.6% chance or a 49.2% of receiving a bye? Some people might find such a system acceptable, but we will take a very conservative view. We regard any deviation greater 0.5% as *unfair* to individuals—and therefore unacceptable.

Our proposed method is extremely fair to individuals because each wrestler's chance of a bye differs by less than 0.5% from his chance under a perfectly random draw. In a Section 18 we will explore the practical ramifications of a 0.5% shift in a bye probability for an individiual wrestler.

5 Equitable Treatment of Teams

Table 3: Seeds, byes, and the expected number of byes for the 2004 EIWA tournament

TEAM	Leh	Pen	Cor	Rut	Arm	Nav	Har	Bro	Col	Pri	ESU	F&M	Ame
SEEDS	9	9	10	9	8	9	5	3	6	2	4	4	2
BYES RECEIVED	3	0	7	3	5	3	1	1	4	0	4	2	0
BYES EXPECTED	3.75	3.75	4.125	3.75	3.25	3.75	1.875	1.25	2.5	0.75	1.75	1.75	0.75
DISCREPANCY	-0.75	-3.75	3.875	-0.75	1.75	-0.75	-0.875	-0.25	-1.5	-0.75	2.25	-0.75	-0.75

To explain equitability to teams we may study Table 3, which gives data on seeds and byes for the 2004 EIWA tournament again. The second row shows the number of seeds for each team, as finalized by the coaches in a meeting before the draw. The third row shows the actual number of byes each team received in the random draw conducted by

the BAR method. As noted earlier, Penn received 0 byes out of 9 seeds, while Cornell received 7 byes out of 10 seeds. The next row shows the *expected number of byes* for the teams. For instance, the expected number of byes for Princeton is 0.75.

Question: *A wrestler either receives a bye or he doesn't. So how can a team "expect" three-quarters of a bye, given that there is no such thing as a fraction of a bye?*

Answer: The expected number of byes is a mathematically computed quantity. It is an average.

For instance, if you toss a fair coin three times then the expected number of tails is 1.5. Of course, the coin cannot come up tails exactly 1.5 times.

Here is how the expected number of byes is computed for each team. Princeton had two seeded wrestlers in the tournament, each of whom was in a weight class with three byes. Each of those two wrestlers had a 3-out-of-8 chance of getting a bye (in a random draw), and so the number of byes Princeton expected to get was $(3/8) + (3/8) = 3/4 = 0.75$. In other words, suppose truly random draws were conducted repeatedly. In each draw Princeton would receive 0, 1, or 2 byes. However, the average number of byes for Princeton taken over all the draws would be 0.75.

The expected number of byes for the other teams is computed similarly, but we must be careful to account for the weight classes with four byes properly. For instance, Brown had three seeded wrestlers, one of whom was in a weight class with four byes. So Brown's expected number of byes was $(3/8)+(3/8)+(4/8) = 1.25$. There were seven weight classes with three byes, and three weight classes with four byes. So Cornell's expected number of byes was $\left[7 \times (\frac{3}{8})\right] + \left[3 \times (\frac{4}{8})\right] = 4.125$. To compute the expected number of byes for the other teams, you would need to know how each team's seeds were distributed across the weight classes; we have not provided that detailed information. Also, note that two teams could theoretically have the same number of seeds, but a different expected number of byes if their seeds were distributed across different weight classes appropriately. However, this did not occur in 2004.

Discrepancies

The last row of the table shows the *discrepancy*, which is just the number of byes minus the expected number of byes. A negative discrepancy indicates that a team received fewer byes than expected, while a positive discrepancy means that a team received more byes than expected. The magnitudes of the discrepancy also carry information, of course, and we regard the discrepancies as one measure of how equitably the draw distributed byes among the teams. If each discrepancy is exactly 0, then the draw assigned to each team its expected number of byes. We cannot expect a draw to produce such a perfect situation, but we can hope to have a draw that produces discrepancies close to 0 (say, between -1

and 1) for all teams. The draw in 2004 was particularly poor in this regard since Penn, Cornell, Army, Columbia, and ESU all experienced large discrepancies.

6 The Inequitability Penalty for Teams

We now present the data in Table 3 graphically so we can visualize the discrepancies for the teams.

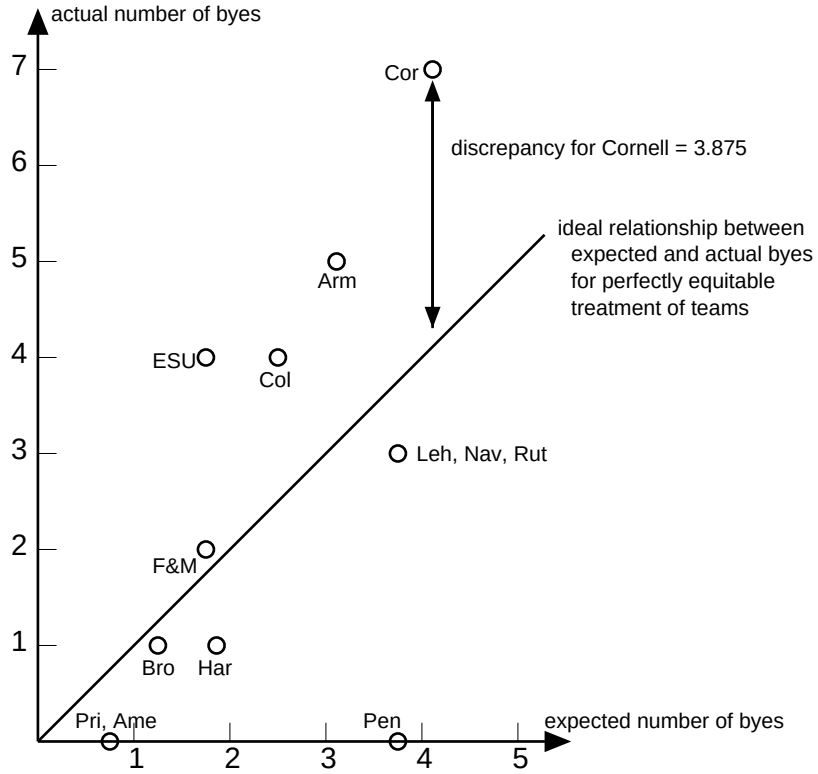

Figure 3: A plot of the actual number of byes vs the expected number of byes for each team at the 2004 EIWA tournament. The diagonal line represents the ideal locations for teams—where the actual and expected number of byes agree.

Figure 3 plots the actual number of byes vs the expected number of byes for the draw in the 2004 EIWA tournament. The upward sloping line depicts the ideal situation, where the number of byes a team actually received is exactly equal to the number of byes it expected. Some teams fall above the line (receiving more byes than expected), while some fall below (receiving fewer byes than expected). The vertical distance—with a suitable plus or minus sign—from a team's plotted point to the line is the discrepancy. The discrepancy for Cornell is shown in the figure as an example.

> **Idea:** *A draw that places teams near the ideal line treats the teams more equitably than a draw that places teams far from the line.*

So a list of discrepancies for the teams gives us a picture of the inequitability of the draw for the teams as a whole. Scientists, engineers, economists, and statisticians will immediately recognize the proper setting of Figure 3. What we have is an excellent framework for *least squares*. Instead of listing all the discrepancies to measure the inequitability to the teams, we *square all the discrepancies and add them up* to get a single number that represents the overall inequitability of the draw to the teams.

This is a crucial definition for us, and we pause to highlight it.

> **Definition of the Inequitability Penalty to Teams:** Given a draw, we compute the expected and actual number of byes for each team, and then subtract to find the discrepancies, as in Table 3. The *inequitability penalty to teams* of the draw is the sum of the squares of the discrepancies.

We'll refer to the inequitability penalty to teams as simply the *penalty* of a draw. Again, the penalty is a single number that measures how inequitably the byes are actually distributed among all the teams relative to the number of byes the teams would expect using ideal random draws.

Example. We can use the rows of Table 3 to find the penalty for the draw in the EIWA tournament in 2004 as follows:

$$\begin{aligned}
&\text{penalty for 2004 EIWA tournament draw} \\
&\quad = \text{sum of squares of all teams' discrepancies} \\
&\quad = (\underbrace{3 - 3.75}_{\text{Lehigh discrep.}})^2 + (\underbrace{0 - 3.75}_{\text{Penn discrep.}})^2 + (\underbrace{7 - 4.125}_{\text{Cornell discrep.}})^2 + \cdots + (\underbrace{0 - 0.75}_{\text{American discrep.}})^2 \\
&\quad = 36.40625.
\end{aligned}$$

So the penalty for the draw at the 2004 EIWA tournament is about 36. Out of context, we don't know what the magnitude of this number means. We can say, however, that the terms $(0-3.75)^2 \approx 14$ and $(7-4.125)^2 \approx 8$ for Penn and Cornell account for most of the penalty. It turns out that typical draws conducted by the BAR method have penalties in the range of 17–19 when there are 13 teams. We may view a draw with penalty above 22 as bad; penalties over 30 are egregious, as far as equitable treatment of the teams. Fortunately, the draws with high penalties occur less often than those with moderate penalties. We will see that a draw as bad as the 2004 EIWA draw will occur about once every 29 years under the current BAR method.

Of course, we want draws with *low* penalties. Unfortunately, draws with low penalties are also rare using the BAR method. Look back at Figure 2 for a confirmation.

7 What Happened in 2002?

As a contrast to the 2004 draw, we now examine the draw for the 2002 EIWA tournament. Table 4 shows the seed, bye, and discrepancy data,[9] and Figure 4 depicts the data graphically, as we did before. Observe that the data points fall closer to the line for this draw, indicating that teams were treated more equitably than in the 2004 draw.

Table 4: Seeds, byes, and the expected number of byes for the 2002 EIWA tournament

TEAM	Leh	Pen	Cor	Rut	Arm	Nav	Har	Bro	Col	Pri	ESU	F&M	Buc
SEEDS	9	10	9	7	7	7	8	7	4	5	5	0	2
BYES RECEIVED	2	4	3	2	3	3	4	3	4	3	3	0	1
BYES EXPECTED	3.625	4	3.625	2.75	2.75	2.75	3.25	2.75	1.625	2.125	2	0	0.75
DISCREPANCY	-1.625	0	-0.625	-0.75	0.25	0.25	0.75	0.25	2.325	0.875	1	0	0.25

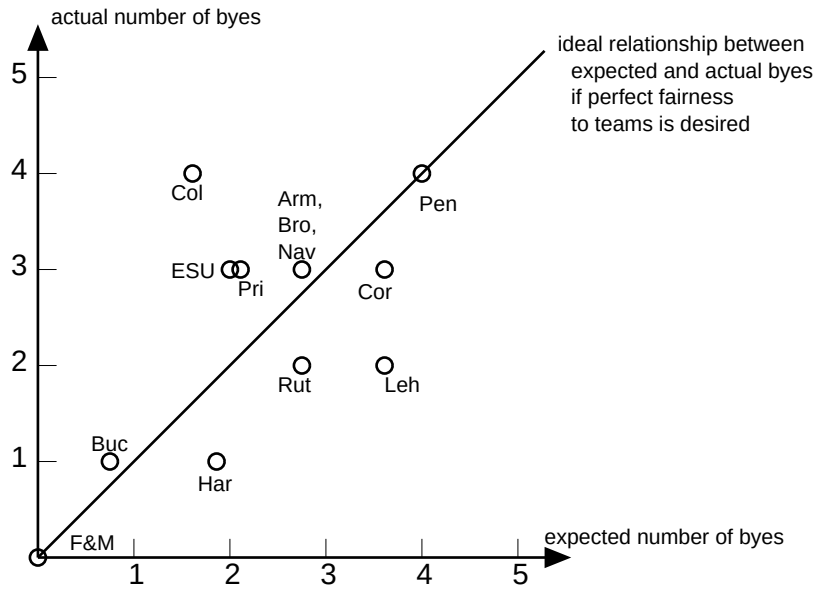

Figure 4: A plot of the actual number of byes vs the expected number of byes for each team at the 2002 EIWA tournament. The diagonal line represents the ideal locations for teams—where the actual and expected number of byes agree.

The discrepancies in the last row of Table 4 are small for the most part, and our sense is that teams were treated equitably on the whole. We can quantify this feeling by computing the inequitability penalty for teams:

[9] Bucknell University was in the league, but American University was not in 2002.

penalty for 2002 EIWA tournament draw

$$\begin{aligned} &= \text{sum of squares of all teams' discrepancies} \\ &= (-1.625)^2 + (0)^2 + (-0.625)^2 + \cdots + (1)^2 + (0)^2 + (0.25)^2 \\ &= 6.5625. \end{aligned}$$

This is a very small penalty! In fact, we will see that such a small penalty would only be expected about once every 40 years if we use a random draw. We would like such draws to occur more frequently. Our proposed method of draw makes draws as good (or better) than the 2002 draw occur three times as often.

8 Least Squares

We pause to discuss *least squares,* a powerful and versatile mathematical tool used to summarize data efficiently: This tool will be used in our proposed method of selecting byes in the EIWA tournament.

Let's look at an informal example from economics. Suppose we want to measure how education—say, number of years of schooling beyond 8th grade—is related to annual income for one's first full-time job. We collect data from many people and plot the points on a graph, as on the left in Figure 5. The horizontal axis is for education, and the vertical axis is for income. We have not included tickmarks and units on the axes, because this is just an informal example. As expected, the plotted points go upward to the right, since more highly educated people usually get higher-paying jobs. If someone asked us for a summary of the relationship between income and education, we could present them with the plotted points, but this would not be particularly helpful. The graph on the right in Figure 5 superimposes a straight line on the data and serves as a concise summary. We might even omit the plotted points and present just the line. The intercept on the vertical axis is roughly the annual income a person with an 8th grade education can expect. The slope of the line represents the increase in income for each additional year in education.

Note that the plotted data generally lie *near* the line, but not directly on it; some natural variation occurs. In fact, if all the data fell exactly on the line, we would suspect that the data-collector lied.

Now the line plotted in Figure 5 was not just eye-balled. It was obtained by the precise method known as the *least squares fit.* Here is the idea. There are many possible lines that roughly fit the data. For each such line we can compute the discrepancies, just as we did for byes in Table 3 and Figure 3. Then we add up the squares of the discrepancies to get a "penalty." A line with a lower penalty fits the data better than a line with a

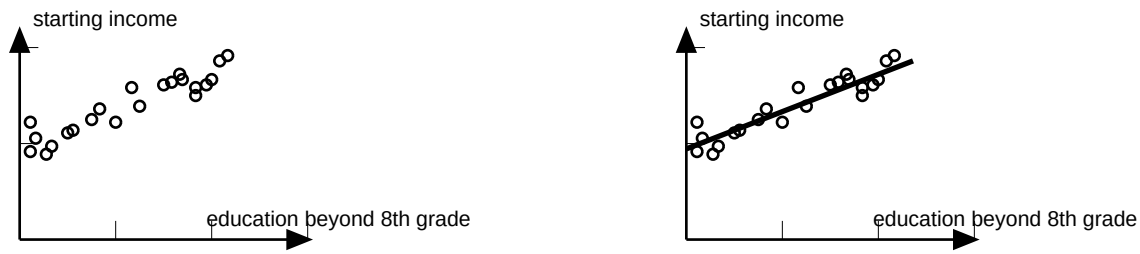

Figure 5: Data for starting income vs years of education beyond 8th grade together with the best (i.e., least squares) line to fit the data.

higher penalty. The line with the smallest penalty fits the data the best. This line is called the *least squares line*. There is a somewhat fearsome-looking formula that allows you to give the equation for the least squares line in terms of the given data. In fact, many calculators and spreadsheets have the capability to compute the least squares line for data you provide. (Often this method goes by the name *regression*.) Many college students encounter least square fit lines in their chemistry and physics laboratory courses[10]

However, the method of least squares is much more than a mathematical plug-and-chug tool for tormented college students. One of the most important scientific discoveries of the 20th century had a least squares calculation at its core. In the 1920's the astronomer Edwin Hubble plotted data for galaxies—with a galaxy's distance from Earth on one axis and speed on the other. The plotted points fell roughly on a line sloping upward and to the right, which told him that the farther galaxies were the ones traveling faster. One way to explain the data is to guess that the universe started with a Big Bang, and that the galaxies have been moving away from one another ever since. Now by looking at the slope and intercept of the least squares line, we can try to estimate how long the galaxies have been moving—the age of the universe.

9 Least Squares and the Penalty

In a typical application of least squares we must determine the straight line that best fits the given data. Our proposed method of selecting byes uses the same least squares idea in reverse. We already know the best line, namely, the line with the actual number of byes equal to the expected number of byes. However, the data (number of byes each team receives) are the unknowns. We would like to have the data fall near the best line.

[10]Midshipmen at the Naval Academy also derive the fearsome least squares formula as a homework problem in their sophomore multivariable calculus class—at least in years when the course-wide syllabus is written by the author of this report.

The penalty gives us an overall measure of how far the data points are from the line. We select three complete draws and use the one whose penalty is the smallest, that is, the draw that comes closest to the ideal equitable treatment of teams.

10 There is No Perfect Method to Select Byes

Here is a hope held by some wrestling fans: We should be able to devise a method of selecting byes that treats the teams equitably while preserving perfect fairness to individuals. Such a method might be complicated, but it could be done, perhaps with a computer.

We now state some facts that dash this hope completely.

> **Fact 1:** *The method currently in use (*BAR: *byes assigned randomly and independently in each weight class) is the only practical method of selecting byes that is perfectly fair to individuals.*
>
> **Fact 2:** *The method currently in use* (BAR) *often distributes byes inequitably among the teams. It may even produce grossly disproportionate draws.*
>
> **Fact 3:** *It is impossible to find a method of selecting byes that is perfectly equitable to teams in all situations.*
>
> **Fact 4:** *Any method of selecting byes that guarantees coming close to being perfectly equitable to teams will be unacceptably unfair to individuals.*

Many people in the wrestling community in some sense already "know" that Facts 1 and 2 are true, even though they may not have articulated them explicitly. Many people also suspect that something like Facts 3 and 4 are also true. We will give convincing evidence for Facts 3 and 4 a bit later.

> **Summary of the above facts**: *We cannot* force *perfect or nearly perfect equitability on the teams without destroying fairness to individuals. There is no method of conducting the draw that is perfectly fair to individuals, while simultaneously treating the teams equitably.*

Tycho, Hermes, and Athena

In view of the above facts one may wonder why the BAR method should be changed at all. Fairness to individuals is paramount, and since all methods of imposing perfect equitability

on the teams create unacceptable unfairness to individuals, we should leave well enough alone and stick with the BAR method. Although Tycho (Greek goddess of chance) and Hermes (god of wrestling) may conspire to produce a draw that treats some teams very differently from others every few years, we mortals should stoically accept whatever draws the gods bestow upon us. It is lamentable that byes are sometimes distributed inequitably to teams, but there's nothing to be done about it. Besides, inequitability to teams is not so terrible. The team championship is won by the wrestlers on the mats.

My view is different. First, I claim that a draw that treats teams inequitably may also create an artificial unfairness to the individuals. In the first rounds of the EIWA tournament, a wrestler may adopt strategies and tactics to maximize his chances of earning bonus points for his team (or to avoid conceding bonus points to a superior opponent), even though these actions might deviate from his preferred style of wrestling—or from his own best interest if he only sought to advance through the bracket efficiently. This observation is evident to anyone who watches an EIWA tournament unfold.[11]

> **Claim.** *If the team competition is important enough to influence the strategy and tactics adopted by individual wrestlers in their matches, then draws that distribute byes inequitably to teams should be discouraged. In general, we do not like to see a team receive byes in flagrant disproportion to the number of byes it would expect to receive in a random draw based on seeds. Such disproportionate draws may artificially place individuals in situations where they alter their tactics and strategies.*

Second, there is a more powerful member of the Greek pantheon eager to join Tycho and Hermes at the EIWA wrestling tournament. Athena (goddess of noble competition, fair play, ... and mathematics) is perturbed by the pranks sometimes played by Tycho and Hermes in the draw. She knows that by mildly altering the random method of selecting byes much of the inequitability to teams can be avoided.

We will demonstrate mathematically that a modification to the BAR method will produce draws that are far less likely to treat the teams inequitably—with negligible loss of fairness to the individual wrestlers.

[11] For instance, a wrestler will also continue a bout after sustaining an injury to avoid defaulting a match (surrendering bonus points to a rival team), or in the hopes of being able to maintain a lead for the win somehow. In the 2005 EIWA tournament, there were close races for final team placements up and down the leaderboard, and there were several instances of wrestlers "taking one for the team," as described.

11 Randomness and Proportionality

We have stated that we prefer draws that assign byes to teams in rough proportion to their number of byes. The word *proportion* has a controversial connotation to many in the wrestling community, and we pause to explain the context here.

Randomness is a good and appropriate device to assure fairness in sports. In a wrestling match a coin flip helps determine which wrestler chooses the starting position for the second period; in the third period the other wrestler gets the choice. However, too much randomness is not good. In a collegiate wrestling dual meet, the choice of positions for the second period alternates between the two teams in successive weight classes. A devoted advocate of randomness would insist that a separate decision be made for each weight class. In the long run, each individual wrestler and each team would get the second-period choice of position half the time, but in a particular dual meet it might happen that one of the two teams would get the second period choice in nine or ten matches. The rule to alternate the choice between the teams guarantees that such unbalanced meets are avoided. The price is small: an individual wrestler's chances of getting the choice of position in the second period are correlated with the chances of his teammates.

A similar state of affairs holds for the BAR and the BARRELS methods of selecting byes for the EIWA tournament. The randomness of the BAR method of draw is its strongest feature. It guarantees perfect fairness to indivdual wrestlers in a simple manner. It also guarantees that teams are treated perfectly equitably, *but only in the very long run.* In the short run, large discrepancies occur. If we take the seeding data for the 2004 EIWA tournament, for instance, and repeatedly conduct random draws by the BAR method, then *on average* Cornell receives 4.125 byes and Penn receives 3.75 draws—their expected numbers, i.e., the numbers that are proportional to their seeds. It just so happened that the *particular* draw selected in 2004 deviated greatly from the average. The BARRELS method allows us to avoid nearly all of the situations where the number of byes is disproportional to the number of seeds for a team. The price is again small: the individual bye probabilities differ from their ideal values by a tiny amount.

> **Note for mathematicians and computer scientists.** Although the BARRELS-(3) method is not perfectly random in its treatment of individual wrestlers, it comes very close to being random. Someone who insists on absolute perfect randomness in the draw may be surprised to know that the BAR method itself is likely not perfectly random. Although the byes are determined by a computer's "random number generator", in fact, such generators are frequently not perfectly random; they are only approximately random. We may aspire to conduct a perfectly random draw, but we may, in fact, be using just an excellent approximation.

12 Forced Equitability is Unfair to Individuals

If you believe that better draws are ones in which teams receive close to their expected number of byes, then a natural question arises: Is there a way to conduct a draw in such a way that the teams are *forced* to have small dscrepancies? In other words, can we arrange it so that draws similar to the one in 2002 always occur, while forbidding draws like 2004? Yes, such draws are certainly possible. We can envision a two-step process:

First step: Determine the number of byes each team will receive.

Second step: Distribute the allotted number of byes to the wrestlers on each team in some manner.

We will show how to carry out the first step—at which point gross unfairness to the individual wrestlers will become apparent, and so we will not bother with the second step.[12]

Here is how to carry out the first step. Ideally, we arrange to have all discrepancies equal to 0 so that the penalty is 0, corresponding to a perfectly equitable distribution of the byes to the teams. However, the expected number of byes for a team is usually not a whole number, so there is no way to get the discrepancy equal to 0. The best we could hope for is to make the discrepancies as small as possible.

In Table 5 we repeat Table 3, but add a new row in which we *assign* a reasonable number of byes to each team. to make the discrepancies and penalty small. In our first attempt (not shown) we simply rounded the expected number of byes for each team to the nearest whole number. However, that approach gave us a total of 35 byes, instead of the desired 33. So we opted to round Columbia's 2.5 down to 2, and then (somewhat arbitrarily) rounded Lehigh's 3.75 down to 3; choosing other adjustments to achieve 33 byes will lead to problems similar to the ones described below.

Table 5: Seeds, byes, and an *assigned* number of byes for the 2004 EIWA tournament

TEAM	Leh	Pen	Cor	Rut	Arm	Nav	Har	Bro	Col	Pri	ESU	F&M	Ame	TOTAL
SEEDS	9	9	10	9	8	9	5	3	6	2	4	4	2	80
BYES EXPECTED	3.75	3.75	4.125	3.75	3.25	3.75	1.875	1.25	2.5	0.75	1.75	1.75	0.75	
BYES ASSIGNED	3	4	4	4	3	4	2	1	2	1	2	2	1	33
NEW DISCREPANCY	-0.75	-0.75	0.875	0.25	-0.25	0.25	0.175	-0.25	-0.5	0.25	0.25	0.75	0.25	

The last row of Table 5 lists the discrepancies for our assignment of byes. If we square these discrepancies and add, we do indeed get a very small inequitability penalty for

[12] We point out that considerable combinatorial difficulties would sometimes arise in the second step anyway. Experts in combinatorics will recognize that the second step is equivalent to finding a certain factor in a graph wtih restrictions on the vertex degrees; the necessary "Marriage Theorem" conditions might not be met in a particular draw.

teams:

$$\text{penalty} = (3 - 3.75)^2 + \cdots + (1 - 0.25)^2 = 3.17125$$

This small penalty is no surprise, since we are *assigning* byes to force the actual number of byes to be close to the expected number for each team.

A Tale of Two Tigers

However, we should not be overly pleased with ourselves for having created a situation that treats the teams equitably. *We have created gross unfairness to the individual wrestlers.* Consider Princeton's two seeded wrestlers. Each seeded Tiger occurred in a weight class with 3 byes, and so should have a 37.5% chance of getting a bye. But we are now *mandating* that the Princeton team as a whole receives 1 bye, which implies that one of Princeton's two seeds must have at least a 50% chance of getting a bye. The gap between the ideal of 37.5% and an actual probability of (at least) 50% is too large to accept if we care about fairness to individual wrestlers.

A similar argument shows that wrestlers on other teams are also treated unfairly; we illustrated the process with Princeton because it was easier to exhibit the difficulty with just two wrestlers.

A Question

Was the 2002 EIWA tournament draw unfair to individuals? The discrepancies were small (See Table 4 and Figure 4), and so the teams were treated very equitably. We just demonstrated that when the teams receive byes equitably, the draw must be unfair to individuals.

Answer. No. The draw in 2002 was perfectly fair to individuals (because it was chosen by the BAR method). It is only when we *require* teams to receive byes equitably that unfairness to individuals arises. No such requirement was in place in 2002; the draw was random. It just so happened that the draw distributed the byes equitably. We certainly don't want to forbid such a fortunate occurrence. Our proposed BARRELS-(3) method of draw makes such occurrences much more likely than the BAR method.

13 Time-Out: A Game of Darts

We leave the wrestling world for a moment to discuss an unusual type of darts game, whose behavior mimics our proposed method of selecting byes. The dart-board is partitioned into 100 regions, numbered 1, 2, ... 100. Because the regions all have the same area, but are irregularly shaped and positioned; a person throwing a dart at the board has a 1% chance of hitting any particular number; there is no skill involved in this purely

mathematical game. The two players are Bar and Barbarrelsa. The winner is the one whose dart hits the *lower* number.

Bar steps up and throws a dart. On average, he'll score a 50, so let's say he gets exactly 50 for the moment. Then it's Barbarrelsa's turn. She gets to throw three darts, but only her best (lowest) one counts. Clearly, the odds favor Barbarrelsa. Each of her darts has a 50% chance of beating (or tying) Bar's score of 50. The only way Barbarrelsa can lose if all three of her darts score over 50. The probability of three such misses is $(1/2) \times (1/2) \times (1/2) = 1/8 = 12.5\%$. In other words, Barbarrelsa's chances of winning this skewed game are $100\% - 12.5\% = 87.5\%$.

What if Bar's dart happened to hit a relatively low number, say, a 20? This is a tougher score for Barbarrelsa to beat. With three darts her chances are $1-(4/5)\times(4/5)\times(4/5) = 49\%$. Despite Bar's good score of 20, Barbarrelsa retains a decent chance of winning, merely because she has a few more darts to throw.

> **Mathematical note:** In general, if Bar hits the number p on the dart board, then Barbarrelsa's chances of winning with three darts are $1-(1-p)^3 = 3p - 3p^2 + p^3$ (in percent). If we account for all values of p from 1 to 100, then we find that Barbarrelsa's overall chance of winning is 75%, and her median score (for her lowest dart) is about 17, which is much better than Bar's median score of 50.

What if we allow Barbarrelsa to throw many darts, say, 500? Then it is extremely likely that at least one of her darts will hit the 1, the lowest possible number, guaranteeing her a win—or at worst a tie—no matter how lucky Bar is. In fact, we would be very surprised if Barbarrelsa failed to a hit a 1. (With 500 darts, the chances that Barbarrelsa fails to get a 1 are $(99\%)^{500}$, which is less that 1%.)

14 The BARRELS Method

Here is our idea to improve the current method of seleting byes.

> **Strategy:** Make several complete draws using the usual BAR method. Then choose the one with the lowest inequitability penalty to teams.

We call this approach the *BARRELS method*, which stands for

Byes **A**ssigned **R**andomly. **R**epeat. **E**xecute **L**east **S**quares.

"Execute Least Squares" refers to our use of the inequitability penalty for teams. A precise protocol for the BARRELS method was outlined on page 2 of this report.

When we use k complete draws and choose the one with the lowest penalty, we have the *BARRELS-(k) method.*

Note: The BARRELS-(1) is just the usual BAR method; if there is only one complete draw to select from, then it must have the lowest penalty among the draw(s) we selected. So we should view the BARRELS method as an extension of the current BAR method. The BARRELS method enjoys nearly perfect fairness to individuals, but greatly improves equitability to teams—with little extra effort in actually conducting the draw.

Our analysis shows that using $k = 3$ complete draws works well in achieving equitability to teams while preserving fairness to individual wrestlers.

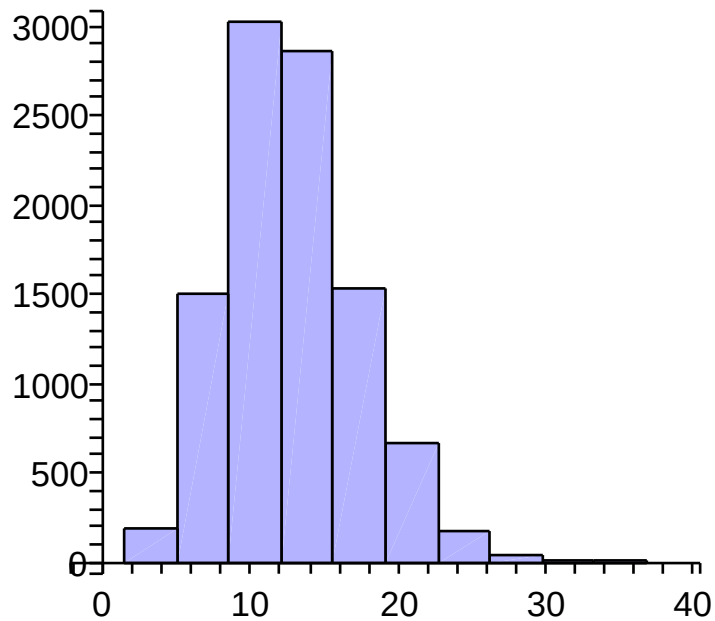

Figure 6: A frequency plot of the penalty based on 10,000 random draws (BARRELS-(3) method) using the seeds for the 2004 EIWA tournament ($n = 10,000$; median = 12.4 mean = 12.8.)

We now discuss the BARRELS-(3) method and compare it to the BAR method of selecting byes. The bar-graph (or perhaps we should say BARRELS-graph) in Figure 6 shows the frequency of penalties for a simulation of 10,000 draws using the BARRELS-(3) method based on the draw for the 2004 EIWA tournament.[13] Comparing Figure 2 to Figure 6, we see that the penalties are skewed to smaller values in the BARRELS-(3) method, ranging from 1 to about 38, with a peak near 10. The median penalty for BARRELS-(3) was 12.4, and the mean was 12.8. Recall that for the BAR method the

[13]The simulation performed the following process 10,000 times: produce three complete draws; compute their penalties; choose the draw with the smallest penalty.

median was 17.9, and the mean was 19.0. The actual draw for the 2004 EIWA tournament had a penalty of 36.4. Figure 6 shows that such a penalty would almost never occur under the BARRELS-(3) method; we would have to produce three random draws in a row, each with a penalty of at least 36.4, for the BARRELS-(3) method to produce a draw as inequitable to the teams as the actual 2004 draw.

That the BARRELS-(3) method generally gives draws that treat the teams more equitably than the BAR method is not surprising. The advantage of three attempts to get a lower penalty is the same advantage Barbarellsa had over Bar in the fictitious game of darts discussed in the previous section.

BARRELS and Bye Spreads

Table 6: Distribution of the bye-spread for simulations of the 2004 EIWA tournament with n =100,000 random draws using the BAR method and the BARRELS-(3) method

bye-spread	0	1	2	3	4	5	6	7	8	9	10
frequency (**BAR**)	3577	18,121	27,216	24,425	15,654	7579	2619	671	126	11	1
cumulative freq.	3577	21,698	48,914	73,339	88,993	96,572	99,191	99,862	99,988	99,999	100,000
frequency (**BARRELS**)	5725	27,546	34,618	22,336	8065	1553	149	7	1	0	0
cumulative freq.	5725	33,271	67,889	90,225	98,290	99,843	99,992	99,999	100,000	100,000	100,000

The BARRELS-(3) method also performs better than the BAR method with respect to bye-spreads. Table 6 shows the results of two simulations with 100,000 draws based on the seeds for the 2004 EIWA tournament. Recall that in 2004 there were three teams vying for the team title, each with about the same number of seeded wrestlers. The bye-spread is the maximum difference in the number of byes among these three teams. (The two rows for the BAR method in Table 6 occurred earlier in Table 2.) The two rows for the BARRELS-(3) method show that smaller bye-spreads will occur much more frequently. From Table 6 we see that the bye-spread will be 0, 1, or 2 more than two-thirds of the time using the BARRELS-(3) method, but less than half of the time using the BAR method. Also, a bye-spread of 3 or greater can be expected less than once every 10 years using BARRELS-(3), but more than once every 4 years using BAR.

By any measure of equitability to teams, the BARRELS-(3) method performs significantly better than the BAR method. Our discussion in this section has focused on data for the 2004 tournament, but the analysis leads to similar results for other years. See Section 16.

15 Are We Using Quotas?

A reader of an earlier draft of this report asked a natural question: Is the BARRELS-(3) method based on quotas?

Answer. No, the BARRELS method is not based on quotas. The expected number of byes is not a quota. The method does not *require* the number of byes of each team to be equal to or near some prescribed quota and does not *forbid* the number of byes to be far from some quota. As we've shown, any draw method based on strict or almost-strict quotas is guaranteed to be unfair to individuals. The BARRELS method discourages the number of byes from being far from the expected number for each team. It turns out that just a little nudge in this direction (selecting three complete draws randomly and choosing the best one) leads to huge increases in equitability among the teams. The price to pay is small. An individual wrestler's probability of a bye changes by less than 0.5%. We'll discuss the ramifications of this small change later.

The Case of the Capricious Professor

In general, when absolute randomness is desirable, but extreme outcomes are to be avoided, a quota method is too "rigid," and is not random at all. A BARRELS-type scheme preserves the randomness while minimizing the chances of an unusual outcome, like the draw at the 2004 EIWA tournament. The following anecdote relies on the same issue.[14]

Once upon a time a math professor got a lot of flak from students and the dean at his college for his harsh grading policies. The dean monitored the professor's grades for years. One semester the professor responded by secretly instituting a new grading scheme. He'd give half his students A's and half B's. The grades would be assigned randomly, independent of any work or performance by the students. The day before semester grades were due, the professor would write 'A' on 20 slips of paper and 'B' on twenty slips of paper and withdraw them one at a time from a hat to determine the grades of his 40 students. Of course, the professor did not announce his new grading scheme; the dean could have revoked his tenure and fired him. Nonetheless, the students at the school quickly figured out what was happening. Within a year the whole student body knew they could take his class and get either an A or a B without doing any work at all. Everyone was happy. Not surprisingly, the students did not complain to the dean. The professor explained to the dean that his new curve gave A's to the top half of each class and B's to the bottom half. The dean was satisfied because the inflated grades were in line with those assigned by other professors at the college. The professor got no flak from anyone, and the random manner of assigning grades appealed to his mathematical (and lazy) nature.

[14] A few details of this true story have been changed; the unnamed professor has retired.

Everything worked well for years. But then one semester the random assignment of grades happened to give his 20 female students all A's and his 20 male students all B's. The professor faced a dilemma. He knew the students compared grades at the end of each semester (to make sure the unspoken bargain was still in effect), and that the male students would collectively accuse him of discrimination and bring the matter to the dean. In fact, the dean was likely to notice the sex-biased grading pattern independently of the students.

The professor considered two methods to preserve his unspoken, secret agreement with the students. One method used quotas: Each semester he would give half the women A's and half the women B's; he would do the same for male students. The quota method would avoid the charge of discrimination. However, the professor knew the dean would notice the neat pattern within a few years, bringing the grading scheme to light. Instead, the professor adopted a second method—-akin to our BARRELS-(3) method of selecting byes at the EIWA tournament. He would make three complete assignments of grades each semester—and then choose the one that distributed the A's and B's among the men and women most fairly (i.e, in proportion to the number of male and female students in his class). The new method would still be largely random, but now the chances of nearly all the women getting A's would drop dramatically. The randomness inherent in the method guaranteed that there would be no neat, repeating pattern for the dean to notice.

16 BAR and BARRELS: Backtesting

In the following pages we backtest the BARRELS-(3) by comparing the following items using the actual seeding data from the EIWA tournaments from 2003 to 2005:

- the expected performance of the BARRELS-3 method;
- the expected performance of the BAR method;
- the actual performance of the BAR method.

In essence, we are comparing what was likely to happen under the BARRELS-(3) and BAR methods with what actually happened. We also "forward-test" the BARRELS-(3) and BAR methods for future tournament scenarios with 14 teams.[15]

[15]The EIWA tournament is expected to increase to 14 teams in 2006 when Bucknell re-joins the league

The 2005 EIWA Tournament

There were 13 teams and 37 byes.
Table 7 shows the seed and bye distributions.

Table 7: Seeds, byes, and the expected number of byes for the 2005 EIWA tournament

TEAM	Leh	Pen	Cor	Rut	Arm	Nav	Har	Bro	Col	Pri	ESU	F&M	Ame
SEEDS	9	7	10	4	7	10	5	7	6	1	7	2	5
BYES RECEIVED	5	4	6	4	2	5	3	3	0	0	4	0	1
BYES EXPECTED	4.125	3.375	4.625	1.875	3.25	4.625	2.375	3.375	2.875	0.5	3.125	0.875	2.0
DISCREPANCY	0.875	0.625	1.375	2.125	-1.25	0.375	0.625	-0.375	-2.875	-0.5	0.875	-0.875	-1.0

Penalty data:

	BAR	BARRELS-(3)
median penalty	18.3	12.6
mean penalty	19.4	13.0

The penalty for the actual draw was 20.8 (somewhat higher than expected).

Using the BAR method, we expect to get a draw that treats the teams more equitably than the actual draw 63% of the time.

Using the BARRELS-(3) method the chances of a more equitable draw increase to 95%.

Bye-spread data: (A comparison for BAR and BARRELS-(3) methods based on the seeding data for the 2005 tournament.)

There were two main contenders for the team title (Lehigh and Cornell). The BAR method assigns one of those two teams at least 3 more byes than the other 25% of the time. For the BARRELS-(3) method this chance drops to 16%. A bye-spread of 4 or greater occurs about once every 9 years with the BAR method, but only once every 20 years with BARRELS-(3).

	BAR	BARRELS-(3)
Probability of bye-spread ≥ 2	50%	40%
Probability of bye-spread ≥ 3	25%	16%
Probability of bye-spread ≥ 4	11%	5%
Probability of bye-spread ≥ 5	4%	$< 1\%$

The 2004 EIWA Tournament

There were 13 teams and 33 byes.
Table 8 shows the seed and bye distributions.

Table 8: Seeds, byes, and the expected number of byes for the 2004 EIWA tournament

TEAM	Leh	Pen	Cor	Rut	Arm	Nav	Har	Bro	Col	Pri	ESU	F&M	Ame
SEEDS	9	9	10	9	8	9	5	3	6	2	4	4	2
BYES RECEIVED	3	0	7	3	5	3	1	1	4	0	4	2	0
BYES EXPECTED	3.75	3.75	4.125	3.75	3.25	3.75	1.875	1.25	2.5	0.75	1.75	1.75	0.75
DISCREPANCY	-0.75	-3.75	3.875	-0.75	1.75	-0.75	-0.875	-0.25	-1.5	-0.75	2.25	-0.75	-0.75

Penalty data:

	BAR	BARRELS-(3)
median penalty	17.9	12.4
mean penalty	19.0	12.8

The penalty for the actual draw was 36.4 (very high).

Using the BAR method, we expect to get a draw that treats the teams more equitably than the actual draw 97% of the time.
Using the BARRELS-(3) method the chances of a more equitable draw increase essentially to 100%

Bye-spread data: (A comparison for BAR and BARRELS-(3) methods based on the seeding data for the 2004 tournament.)

There were three main contenders for the team title (Lehigh, Penn, and Cornell). Using the BAR method one of these teams receives *at least four more byes* than another about once every 4 years. Under the proposed BARRELS-(3) method, that disparate occurrence drops to once every 10 years.

	BAR	BARRELS-(3)
Probability of bye-spread ≥ 2	79%	67%
Probability of bye-spread ≥ 3	51%	32%
Probability of bye-spread ≥ 4	26%	10%
Probability of bye-spread ≥ 5	11%	$< 2\%$

The 2003 EIWA Tournament

There were 13 teams and 32 byes.
Table 9 shows the seed and bye distributions.

Table 9: Seeds, byes, and the expected number of byes for the 2003 EIWA tournament

TEAM	Leh	Pen	Cor	Rut	Arm	Nav	Har	Bro	Col	Pri	ESU	F&M	Ame
SEEDS	9	9	10	7	6	9	8	3	6	3	7	1	1
BYES RECEIVED	4	2	4	1	4	3	3	1	4	1	3	1	1
BYES EXPECTED	3.625	3.625	4.0	2.875	2.50	3.625	3.0	1.25	2.5	1.125	2.875	0.375	0.375
DISCREPANCY	0.375	-1.625	0.0	-1.875	1.50	-0.625	0.0	-0.25	-1.5	-0.75	-0.125	0.625	0.625

Penalty data:

	BAR	BARRELS-(3)
median penalty	17.5	12.0
mean penalty	18.5	12.4

The penalty for the actual draw was 12. (somewhat low).

Using the BAR method, we expect to get a draw that treats the teams more equitably than the actual draw about once every 5 years.
Using the BARRELS-(3) method a more equitable draw will occur every 2 years.

Bye-spread data: (A comparison for BAR and BARRELS-(3) methods based on the seeding data for the 2003 tournament.)

The four teams with the most seeds were Lehigh, Penn, Navy, and Cornell.

The actual bye-spread was 2.

The table compares the bye-spread probabilities and demonstrates a big advantage for the BARRELS-(3) method. For instance, one team receives at least 4 more byes than another (bye-spread ≥ 4) every 2 out of 3 years in the BAR method (67%), but only 1 time out of 6 years in the BARRELS-(3) method (16%). Bye-spreads of 5 or greater occur every 2 out of 5 years (39%) with BAR and only about 1 time out of 37 years with BARRELS-(3).

	BAR	BARRELS-(3)
Probability of bye-spread ≥ 2	99%	92%
Probability of bye-spread ≥ 3	90%	46%
Probability of bye-spread ≥ 4	67%	16%
Probability of bye-spread ≥ 5	39%	$< 3\%$

The 2002 EIWA Tournament

There were 13 teams and 35 byes.
Table 10 shows the seed and bye distributions.

Table 10: Seeds, byes, and the expected number of byes for the 2002 EIWA tournament

TEAM	Leh	Pen	Cor	Rut	Arm	Nav	Har	Bro	Col	Pri	ESU	F&M	Buc
SEEDS	9	10	9	7	7	7	8	7	4	5	5	0	2
BYES RECEIVED	2	4	3	2	3	3	4	3	4	3	3	0	1
BYES EXPECTED	3.625	4	3.625	2.75	2.75	2.75	3.25	2.75	1.625	2.125	2	0	0.75
DISCREPANCY	-1.625	0	-0.625	-0.75	0.25	0.25	0.75	0.25	2.325	0.875	1	0	0.25

Penalty data: BAR and BARRELS-(3) methods compared.

	BAR	BARRELS-(3)
median penalty	17.8	12.3
mean penalty	18.8	12.7

The penalty for the actual draw was 6.6 (very low).

Using the BAR method, we expect to get a draw that treats the teams more equitably than the actual draw only once every 40 years. Using the BARRELS-(3) method a more equitable draw occurs about once every 14 years.

Bye-Spread Data: (A comparison for BAR and BARRELS-(3) methods based on the seeding data for the 2002 tournament.)

The three teams with at least 9 seeds were Lehigh, Penn, and Cornell. Using the BAR method one of these teams receives *at least three more byes* than another about every other year. Under the proposed BARRELS-(3) method, this occurrence drops to about once every three years.

	BAR	BARRELS-(3)
Probability of bye-spread ≥ 2	79%	67%
Probability of bye-spread ≥ 3	52%	33%
Probability of bye-spread ≥ 4	27%	10%
Probability of bye-spread ≥ 5	12%	$< 2\%$

EIWA Tournaments with 14 Teams

In anticipation of Bucknell's re-joining the League in 2006, we created some seeding data for an EIWA tournament with 14 teams and compared the expected performance of the BAR and BARRELS-(3) methods.

With 14 teams there could be as few as 20 byes (2 per weight class), but we assume that—as usual—some teams are unable to enter ten wrestlers. Simulations were performed for several different scenarios.[16] The numbers given below are representative of those that occurrred.

Penalty data:

	BAR	BARRELS-(3)
median penalty	17.3	11.3
mean penalty	18.8	11.9

Bye-spread data: (A comparison of the BAR and BARRELS-(3) methods.)

Assume there are three main contenders for the team title. Using the BAR method, one of these teams receives at least 3 more byes than another about 55% of the time. Also, using the BAR method, one of these teams receives at least 4 more byes than another about 30% of the time.

Under the proposed BARRELS-(3) method, these undesirable large bye-spread occur only 16% and 4% of the time, respectively.

[16] One scenario: Eight of the teams each have 8, 9 or 10 seeds; the remaining six teams each have one seed, scattered across five weight classes; 30 byes total, and each weight class has 3 byes. Other scenarios use fewer byes and different distributions of seeds among the teams and byes across the weight classes.

17 Individual Bye Probabilites for BARRELS-(3)

The BARRELS-(3) method assigns byes to teams far more equitably than does the BAR method (relative to seeds). However, this is a price to pay: The bye probabilities for individual wrestlers are slightly altered from their ideal values. In the discussion below we confine our attention to wrestlers in a weight class with 4 byes to make the arguments easier to follow. Essentially the same comments apply to a weight class with a different number of byes.

With 4 byes, each of the 8 seeded wrestlers should ideally have a 50.00% chance of receiving a bye. With the BAR method the chance of a bye is indeed exactly 50.00%. Recall that the BARRELS-(3) method selects three complete random draws and then chooses the "best" one, i.e., the one with the lowest inequitability penalty. The use of the penalty means that the BARRELS method is not perfectly random, and thus the bye probabilites will not be exactly 50.00%. Extensive computer simulations show that the bye probabilities for individual wrestlers typically fall between 49.75% and 50.25%. The largest deviation from 50.00% encountered during these simulations was less than 0.5%. In other words, a wrestler's bye probability in a weight class was never as small as 49.5% or as large as 50.5%.

BARRELS-(5) and Beyond

In the BARRELS-(k) method we make k complete draws and select the one with the lowest penalty. We have focused on the case $k = 3$. What happens if we try larger values of k? As k increases, the team inequitability penalty decreases, while the individual bye probabilities deviate more and more from their ideal values. If we try $k = 100$ (which is easy for a computer to execute), we find that the teams are treated very equitably; the penalty typically drops to a number less than 5. In essence, by making so many draws and choosing the best one, we are *forcing* the teams to be treated very equitably. (In our dart game scenario, when one player receives 100 darts, she is nearly certain to hit a very low number.) However, the bye probabilities for the individual wrestlers deviate more and more from their ideal values as k increases. The value $k = 3$ gets the best of both worlds: Teams are treated much more equitably than under the BAR method, *and* individual wrestlers are also treated extremely fairly. Finally, $k = 3$ is small enough to allow a human to carry out the entire BARRELS method using hand computations, if necessary.

18 How Big is 0.5%?

Exactly how big is 0.5%? That's a strange question, and the answer depends on the context. Some in-season wrestlers can predict their weight when asked on the spot to within 0.5% (1 pound on a 200-pound wrestler). Home-buyers certainly notice a difference of 0.5% in mortgage interest rates. However, a 0.5% shift in probability can be much harder to detect. In this section we explain the practical signficance of a potential 0.5% change in bye probabilities for an individual wrestler. To be specific we focus on Wrestler X in a weight class with 4 byes in the 2004 tournament.

To think about the situation probabilistically, let us assume that the 2004 tournament draw is to be conducted repeatedly with the same wrestlers, same seeds, and same number of byes to be assigned. In other words, imagine that the 2004 EIWA tournament is in a *Twilight Zone* episode, doomed to repeat the same events year after year.[17] The only difference each year is the draw, which is determined by either the BAR method, or, say, our proposed BARRELS-(3) method.

First suppose the BAR method is used. Then Wrestler X expects to get a bye exactly 50.00% of the time. So in a 200-year span he'll get 100 byes on average; we would expect the number of byes in a 200-year span to fall between 90 and 110, say. In a 2000-year span he'll get 1000 byes on average; we would expect the number of byes in a 2000-year span to fall between 950 and 1050, say.

Now suppose the meet director proposes a switch to the BARRELS-(3), and that this changes the bye probability for Wrestler X to 50.5% (or to 49.5%)—the worst-case scenarios. What does such a change mean to Wrestler X? Now he expects to get 101 (or 99) byes in a 200-year span, and 1010 (or 990) byes in a 2000-year span. Suppose that Wrestler X and his coach object to this small change—-even though implementing the BARRELS-(3) method would greatly improve the equitable distribution of byes among teams.

> **Provocative Hypothetical.** Suppose that the tournament director ignores all objections and secretly switches to the BARRELS-(3) method anyway.
>
> **Question:** How long will it take for Wrestler X to conclude (with reasonable certainty) that the director stopped using the BAR method?
>
> **Answer:** About 40,000 years!

We'll explain where the number 40,000 comes from later.

We have assumed that Wrestler X is only gathering and contemplating individual bye data (for himself, for his weight class, and for all other weight classes) as the years go

[17] We should also assume that the wrestlers, fans, coaches, and officials are in suspended animation between the annual tournaments so they don't age.

by. The fans in the stands—who are looking at the big picture and the distribution of byes among all of the teams—would know after about 25 years that the BAR method had been abandoned simply because the draws that treat the teams inequitably (relative to number of seeds) would occur so rarely. They would see year after year of tournament draws that treat the teams relatively equitably, and draws like the one in 2004 would essentially disappear.

The preceding discussion provides a powerful refutation to those who object to the BARRELS-(3) method on the basis that it fails to treat the individuals with perfect fairness. The deviation from perfect fairness is negligible—almost immeasurable from a practical point of view. It would take 40,000 years to distinguish the BARRELS-(3) method from the BAR method, as far as their treatment of individuals.

Let us emphasize that our analysis was carried out for the *worst-case* for the BARRELS-(3) method. In fact, for nearly all wrestlers in all draws the deviation from the ideal bye probabilities is much less than 0.5%, which means that hundreds of thousand years would be needed to distinguish between the wrestler bye probabilities provided by the BARRELS-(3) and the BAR methods

Wrestler X and his team are far more likely to be treated unfairly by other sources of errors in the EIWA tournament than the bye probability adjustments caused by the BARRELS-(3) method. The large red-green "coin" used by the mat referee to determine choice of positions at the start of the second period might not be fair to within 0.5%. (Someone could flip it 40,000 times to find out.) Errors in the time-keeping for matches sometimes greatly exceed 0.5%.[18] Another potential source of unfair treatment occurs in the seeding meeting; coaches are not able to seed the wrestlers in each weight class with complete accuracy.

The Direction of the Bye Probability Shift

Suppose that Wrestler X still objects to the BARRELS-(3) method because it changed his bye probability from 50.00%. The tournament director should immediately ask Wrestler X whether his bye probability increased or decreased from 50%. Wrestler X's claim of unfair treatment loses credibility if he cannot even assert in which direction his chances changed given the seeding data.

This director's question to Wrestler X is more difficult than it might appear. Of course, distinguishing between 49.5% and 50%, or between 50.5% and 50% is inherently difficult. However, one might hope that determining the *direction* of the shift would be easier. Such is not the case for some subtle mathematical reasons.

[18]For example, in the championship match for one weight class in a recent EIWA tournament one of the clocks ran forward instead of backward, thereby awarding the "riding time" of one wrestler to his opponent. This error went unnoticed and uncorrected by the meet officials.

Let us return to our two Princeton wrestlers we considered earlier in their respective 3-bye weight classes in the 2004 EIWA tournament. Under the BAR method of draw, each wrestler had a 37.5% chance of getting a bye. We saw earlier that if we *require* the team penalty to be low, then the individual bye probabilities might exceed 50% for these two wrestlers. It would be natural to assume that for the BARRELS-(3) method—which tends to lower the team penalty— the individiual bye probabilities should increase slightly from 37.5% toward 50%. However, computer simulations show that this assumption is often false; the individual bye probabilities might *decrease* slightly from 37.5%. The behavior of the individual bye probabilities is non-linear and unpredictable over small intervals due to the presence of the squares in the penalty function.

19 Confidence Intervals: 40,000 Years Explained

A magician claims to have a bag of special coins. Some of the coins are fair, but others have been rigged. Some coins come up heads more than half the time and others come up tails more more than half the time. You withdraw a coin from the bag and perform a test. You flip the coin 10 times and gets tails 7 times. Should you believe you have one of the "tail-rigged" coins? Probably not. With only ten coin flips, a perfectly fair coin would get seven (or more) tails about 17% of the time.

You continue the experiment, and get 56 tails out of 100 flips. The magician smiles triumphantly and claims you have a tail-rigged coin. You should remain skeptical, of course. A fair coin would get 56 (or more) tails about 13% of the time.

Question: The magician insists you are flipping a tails-rigged coin. How many coin flips would you need to become convinced?

Answer: It depend on how biased the coin is. A smaller bias is harder to detect and requires more flips.

If the magician claims the coin is rigged to come up tails 98% of the time, then your results after 100 flips (and just 56 tails) should convince you he is lying. However, if the claim is that the coin only comes up tails 50.5% of the time, then you'll need a lot more flips to determine whether the magician is lying or not.

The situation we have described is much more than an amusing game. Essentially the same problem comes up science, manufacturing, medicine, and other fields every day.

Here is a situation from quality-control. Say you buy a 1-pound bag of M&M's. You weigh the contents on a very accurate scale and get a weight of 0.985 pounds. Of course, you weren't expecting *exactly* 1 pound, but perhaps you feel short-changed. The next bag you buy weighs 1.001 pound. You feel slightly better, but calculate that the average of the two bags is only 0.993 pounds. Outraged (and on a sugar-high), you contemplate

a lawsuit. Of course, you need more evidence. You buy 10,000 bags of M&M's and find their average weight is 0.991 pounds. Are the manufacturers deliberately making each bag a bit lighter in order to increase profits? Do you have enough evidence for a lawsuit now?

Here is another situation of the type faced by the Food and Drug Administration. A certain serious medical condition kills 95% of its victims within a week, even with the best available treatment. A company claims that its new drug lowers the death-rate to 75%. A controlled experiment must be performed to determine whether the company's new drug works as claimed. How many patients should the experiment have?

Situations like the ones just described fall squarely in the realms of probability and statistics. Fortunately, for our wrestling problem we only need one basic result—one that tells us the likelihood of an event occuring due to sheer chance. Suppose a coin is flipped n times, where n is a large number (say, bigger than 100). Suppose the coin comes up tails T times. We expect T to be about $n/2$ if the coin is fair. The farther T is from $n/2$ the more confident we are that the coin is rigged. The following rule gives us numbers to measure our confidence.

Coin-Rigging Rule: *Suppose a coin is flipped n times (where $n \geq 100$) and comes up tails T times. If*

$$T < \frac{n}{2} - \frac{\sqrt{n}}{2} \qquad \text{or if} \qquad T > \frac{n}{2} + \frac{\sqrt{n}}{2},$$

then we can be 68% confident that the coin is rigged. If

$$T < \frac{n}{2} - \sqrt{n} \qquad \text{or if} \qquad T > \frac{n}{2} + \sqrt{n},$$

then we can be 95% confident the coin is rigged.

The essential notion here is called a *confidence interval*. The most frequently used interval is the one that gives 95% confidence. Statistical rules similar to the Coin-Rigging Rule can be applied to help determine the number of bags needed for the M&M lawsuit and the number of patients for the drug trial.

We now apply our Coin-Rigging Rule to the suspicions of Wrestler X that the tournament director has abandoned the BAR method of draw. Assume that the tournament director has in fact switched to the BARRELS-(3) method and that Wrestler X's bye probability has changed to 50.5%. How many draws would have to be conducted for Wrester X to be 95% sure that the BAR method has been replaced? In this case we associate "tails" with getting a bye. We take $T = 0.505n$ (the expected number of byes in n draws), and look at the 95% confidence interval from the Coin Rigging Rule. If

$$0.505n > \frac{n}{2} + \sqrt{n},$$

then Wrestler X can be 95% sure that the BAR method as been abandoned. A bit of algebra allows us to solve for n. We find that $n > 40,000$. With a smaller number of draws, Wrestler X will not have enough information to deduce that the tournament director switched draw methods. A similar process shows that even to be 68% confident Wrestler X would need 10,000 draws.

20 FAQ

Question. How hard it is to implement the BARRELS-(3) method on a computer?
Answer. Implementation is easy. Very little programming is required. Random draws are generated almost instantaneously (as they are now) using, say, a computer's random number generator. The computation of the penalty for a draw is also a straightforward formula, relying on counting and basic arithmetic. In fact, the penalty can be computed by hand, as illustrated by the tables in this report.

Question. Can the BARRELS method be used at the NCAA wrestling tournament?
Answer. No, at least not easily. The BARRELS method was devised and tested for the EIWA tournament. At the NCAA tournament there are many more teams, more rounds, and other issues to consider.

Question. Hey, I found a few typographical errors in some of your data! Do these mistakes invalidate this report?
Answer. No. The analysis is *robust*—a few small changes in the data do not change the conclusions. For instance, an article on the Navy website and the seedings listed by wrestlingreport.com website give the Harvard wrestler as the 8th seed in the 149-pound weight class for the 2005 EIWA tournament. The initial computer simulations to analyze byes that were conducted by the author the night of the draw were based on that information. In fact, the 8th seed was from Rutgers—and the new analysis and simulations did not alter the relative performance of the BAR and BARRELS methods. Numerous computer simulations were also performed using fictitious seeding data to try to find the worst-case performance of the BARRELS method. The conclusions stated in this report remain valid in all the cases because the most conservative (i.e., largest) estimates were made for deviations in the probabilities of byes for individual wrestlers—and then an extra margin was added. The largest observed deviation was actually well under 0.5%.

Of course, the author would appreciate having any significant typo's brought to his attention.

Other Methods: Why Don't You Just ... ?

A number of methods to select draws have been proposed.[19] We mention a few of these and their shortcomings.

1. Why don't we eliminate all byes by making sure each weight class has 16 wrestlers? With 16 entrants a beautiful, bye-free bracket is easily constructed.[20] The larger and stronger teams in the league boast many able back-up wrestlers, who are eager to wrestle in the EIWA tournament. Why not put those wrestlers in the brackets to get rid of the byes altogether?

   There are several problems with this proposal, but we bring up just two: What if the star wrestler from one team is injured (accidentally, of course) in his bout with the back-up wrestler from another team? And what if a back-up wrestler earns an automatic berth to the national tournament?

2. Why don't we distribute byes to all teams as equally as possible, regardless of seeds? (Unseeded wrestlers are eligible to receive byes, too.)

   Restricting byes to seeds—as is done now—is a good idea. It forces an unseeded wrestler to win four matches in the championship bracket to win the title in a weight class, and thereby demonstrate his superiority to incredulous coaches. With a first-round bye an unseeded wrestler has a shorter path (three matches) to the title. No unseeded wrestler should have an easier path than some seeded wrestlers.

   An unseeded wrestler with a bye would be placed directly in the quarterfinals, and a fluke injury or other mishap to his quarterfinal opponent, say, would put the unseeded wrestler in the semi-finals—with a guarantee to finish no worse than 6th place in his weight class. Any probability of such occurrences should be distributed to the seeded wrestlers, not the unseeded ones. This is especially the case if we assume that EIWA will soon be granting automatic berths to the NCAA's to the top-four finishers in each weight class.

3. Why don't we just somehow distributed byes in proportion to the number of seeds—or nearly so?

   This matter was addressed earlier. The very process of imposing nearly perfect "fairness to teams" creates unacceptable "unfairness to the individuals."

4. Why don't we distribute any byes in a weight class in order from the top seed on down? For instance, in a weight class with 3 byes, the top three seeds automatically get first-round byes.

[19] from the wrestlingreport.com message board, for instance

[20] The regional brackets in a certain NCAA sport use sixteen teams.

This method encourages "gaming" during the seeding meeting. If coaches know that the 3rd seed will receive a bye, and the 4th seed won't, then they can adjust their votes accordingly (read: "strategically") to help their own wrestler avoid a bye, say.

A random distribution of byes among the 8 seeds prevents such gaming.

5. Why don't we *cap* the number of byes for each team (at some number a bit above its expected number of byes presumably) and then distribute byes as follows? Select a weight class at random, and distribute byes at random. Repeat with the next randomly selected weight class, and so on. However, if at any stage a team reaches its cap of byes, then we exclude that team from receiving more byes later.

   This method is unfair to individuals, because the cap amounts to a type of quota. Any method that places a *quota* (explicitly or implicitly) on a team's number of byes will be unfair to individual wrestlers, as we demonstrated. It is also possible that even if we use reasonable caps, it will not be possible to complete a draw on the first attempt due to combinatorial constraints. Some back-tracking in the computer program would be required, which would complicate the process.

   Moreover, despite the random order of weight classes, the unfairness to individuals is sometimes *identifiable.* In other words, we may sometimes be able to identify which wrestlers are less or more likely to receive byes,[21] a situation we regard as undesirable. In our proposed BARRELS-(3) method of draw, the tiny amount of unfairness to individual wrestlers that occurs is identifiable only with extreme difficulty, if at all.

21 Recapitulation

The proposed BARRELS-(3) method for selecting byes at the EIWA tournament is a variation of the current method. Three complete draws are made and then the best one is chosen. We can measure the goodness of a draw using the team inequitability penalty, which is found by a standard least squares computation. Better draws have lower penalties. In such draws the number of byes each team receives tends to be close to the the number of byes the team expects to receive in a random draw. By making three draws and choosing the best one, we greatly decrease the chances of two teams with about the same number of seeds receiving very different numbers of byes. Individual wrestlers are treated fairly under the proposed method, too. Deviations from perfect fairness to individuals are so small as to be negligible.

[21] If a weight class contains seeded wrestlers from several teams that do not have many seeds, then the wrestlers from those infrequently-seeded teams are more likely to receive byes; the other teams will have a greater chance of having reached their cap earlier in the process.