WHO'S REALLY #1?

Choose your math and get the rankings you want

By IVARS PETERSON

It happens every fall. Fierce arguments erupt over which college football team is the best in the country. As the season progresses, this frenzy of head scratching and navel gazing mounts until the climactic bowl games on New Year's Day finally settle the issue.

Or do they?

Even as the mud on the fields of battle begins to congeal, select sportswriters and coaches vote to rank the top 25 U.S. college teams and award the unofficial national championship. But in most years, when the final polls appear, grumbling echoes throughout the land. Allegations of bias, unfairness, cronyism, and petty politicking taint the result.

This season, those cries of anguish and disgust sounded early. In an epic confrontation on Nov. 13, Notre Dame defeated Florida State, 31 to 24, to claim the top spot in the national polls. The Associated Press media poll dropped Florida State from first to second place, while the CNN/USA Today coaches poll put undefeated Nebraska second and Florida State third.

The media voting reflected a fanatic desire to see a rematch between Notre Dame and Florida State in the season-ending Fiesta Bowl, and that could occur only if Florida State retained its second-place standing. But such obvious maneuvering was enough to trigger outbursts of outrage from a number of commentators.

"Voters are no longer rating the best teams. They are creating their idea of the best game," puffed Malcolm Moran of the New York Times.

"Credibility will always be a problem in a business of self-interest and mutual distrust," he concluded. "But a computerized power rating ... would restore a sense of fairness."

It was a direct appeal for a clean, mathematical answer to a messy, human situation.

Similar thoughts had occurred to mathematician James P. Keener of the University of Utah in Salt Lake City after the 1984 college football season. That was the year sportswriters and coaches voted Utah rival Brigham Young University (BYU) the national title on the strength of its undefeated season. But these victories had come against generally undistinguished opponents.

Did this BYU team, despite its record, deserve to be number one? The voters in the polls had said yes.

Perturbed by this result, Keener set out to see whether a mathematical scheme, which automatically takes into account the strength of a team's opponents, would provide a more satisfactory answer. "My aim was to get some kind of fair, mathematically based ranking system that removes the subjectivity," Keener says.

Ranking places teams in an order from first to last, while rating assigns numerical values to their relative strengths. Once you have ratings, you can easily generate rankings. Keener turned to a ranking system suggested a few years earlier by Joseph B. Keller, an applied mathematician at Stanford University.

Keller had devised his ranking strategy while serving a term on the board of trustees of the Society for Industrial and Applied Mathematics (SIAM). He had become interested in ascertaining how well SIAM's published journals stacked up against others in the same field.

He decided he might rank SIAM journals by counting how many times they were cited by other publications. But this wasn't good enough. There had to be a way of adjusting the scores so that being cited in a prestigious, influential journal counted for more than being cited in some "garbage" journal, Keller noted at the time.

This sounds like a chicken-and-egg situation. It's hard to see how one can determine which journals are more influential without knowing how they rank, and ranks can't be determined without knowing how influential the journals are.

Remarkably, a relatively obscure mathematical result known as the Perron-Frobenius theorem furnishes a recipe for calculating just such a ranking. "The theorem shows that this circular reasoning actually has a solution," Keener says. "It also hints at how to arrive at that solution." Using this recipe and given the right data, a computer can establish the rankings in a matter of seconds.

Keller never got a chance to apply his scheme to journals. He did try it on baseball teams, however, using the number of times one team beat another team as the entry in the table of values at the heart of his computation of rankings.

In the 1984 National League baseball season, the Eastern Division had proved stronger than the Western Division, but teams within a division play more games against each other than against teams in the other division. Keller's scheme provided a way of ranking all 12 teams while taking into account the fact that the six teams in the stronger division faced tougher opponents more often and deserved greater credit for their victories.

College football, however, involves many more teams, which individually play far fewer games than professional baseball teams. The issue comes down to what value should go into the space reserved for the outcome of a matchup between each pair of teams in the table used for computing rankings.

In the simplest possible scheme, one can assign a single point for a win, half a point for a draw, and zero for a loss and calculate rankings on this basis. But unlike baseball teams, football teams rarely play each other more than once during a season. Moreover, all the credit goes just to the winner, whether the score is close or lopsided.

Utah's Keener chose to allocate the value per game between the two competing teams on the basis of the game score, and he explored various ways of making this distribution. Each method he looked at showed certain biases, rewarding or penalizing close defensive contests, wild shootouts, blowouts, and other types of game outcomes in different ways.

It's this aspect that bothers Stanford's Keller. "With baseball, it's clear what the entries ought to be — the number of times one team beats another," Keller says. "In football, you have to make up some other way to decide what number goes in as an entry. It's a little too arbitrary."

Nonetheless, once the ground rules — however arbitrary — are set, Keener's scheme produces the required rankings. No amount of politicking or favoritism can budge the results.

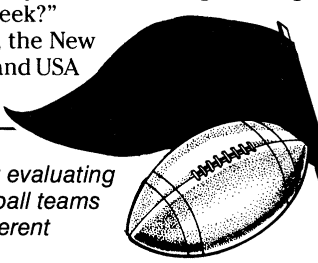
Inevitably, coaches and fans who saw Keener's computer-generated rankings found objectionable features. For example, one of his schemes clearly penalizes a team that plays a weak schedule because such a team can never earn enough points playing weaker opponents to increase its own earned score.

"Of course, this is not all bad," Keener responds. "Simply because a team is undefeated does not mean it should have the highest rank, particularly if it did not play a difficult schedule."

Indeed, when Keener took a look at Brigham Young's 1984 season, the team finished out of the top 10. "There's no objective ranking scheme that I know of that makes BYU number one or even number two," he claims.

Keener is not the only mathematician or statistician with a strong interest in the sporting world. Many have dabbled in the art of ranking, not just in football but also in tennis, basketball, hockey, soccer, racquetball (SN: 10/24/87, p.269), and other sports. Some have also delved into rating teams or players in order to predict the outcome of future games. The statistics literature, for example, includes a 1976 paper titled "Predicting the outcome of football games, or can you make a living working one day a week?"

Moreover, the New York Times and USA Today publish separate team rankings and ratings based on proprietary mathematical formulas. The ratings provide a numerical measure of a team's relative strength and have some value in predicting by how much one team would beat another if they played at a neutral site.



Different schemes for evaluating college football teams produce different rankings, as demonstrated in these results for the 1993 season based on games played through Nov. 20. 1. Associated Press media poll. 2. USA Today/CNN football coaches poll. 3. The New York Times computer rankings. 4. USA Today computer rankings. 5. Keener's method, based on the Perron-Frobenius theorem. 6. Bradley-Terry model, which statistically relates future performance to outcomes of previous games. Asterisks indicate ties in the rankings.

The fine print accompanying the newspaper charts notes that these analyses are based on each team's scores, with the emphasis on who won, by what margin, and against what quality of opposition. The New York Times computer model also "collapses" runaway scores, takes note of home-field advantage, and counts recent games more than earlier games. The USA Today computer ratings, compiled by Jeff Sagarin, incoi porate similar adjustments.

Harvard statistician Hal Stern has played with the data to see if he can tease out the mathematical procedures underlying these computer ratings. Neither appears to be based on a method as straightforward and elegant as Keener's.

"The more mathematical stuff that Keener has done is interesting and helpful if you want a mathematical foundation for ranking," Stern says. "But from a statistician's perspective, the question is which model predicts best."

Gamblers, for example, want to predict the outcomes of games. To them, a team's rating is a far more valuable piece of information than its rank because the rating suggests the potential margin of victory over a given opponent.

On the basis of his own experiments, Stern suspects that the USA Today computer rating is based on a "least squares" method. When one team plays at home against another, the game's outcome is predicted by calculating the difference between the two teams' ratings and adding some factor to account for the home-field advantage. A team's rating for a single season is obtained by minimizing the sum of the squares of the differences between the actual outcomes of games and the predicted outcomes.

DISORDER ON THE FOOTBALL FIELD

Team	Record	1	2	3	4	5	6
Florida State	10-1-0	1	2	1	1	1	1
Nebraska	10-0-0	2	1	2	4	6	11
Auburn	11-0-0	3	—	5	13	11	22
Notre Dame	10-1-0	4	4	9	3	4	4
West Virginia	10-0-0	5	3	7	12	9	19
Tennessee	8-1-1	6	5	4	2	2	2
Florida	9-1-0	7	6*	3	5	5	6
Texas A&M	9-1-0	8	6*	6	6	3	3
Miami (Fla.)	8-2-0	9	9	8	7	7	5
Wisconsin	8-1-1	10	8	16	22	8	21
Boston College	8-2-0	11	12	13	20	—	—
Ohio State	9-1-1	12	10	11	9	17	10
North Carolina	9-2-0	13	13	14	16	21	18
Penn State	8-2-0	14	11	15	11	13	8
U.C.L.A.	8-3-0	15	14	10	8	16	9
Oklahoma	8-2-0	16	15	12	19	15	13
Alabama	8-2-1	17	17	21	14	14	16
Colorado	7-3-1	18	18	18	17	25	25
Arizona	8-2-0	19	16	25*	21	20	15
Kansas State	8-2-1	20	19	24	—	—	—
Indiana	8-3-0	21	21	—	—	18	20
Virginia Tech	8-3-0	22	20	19	25	—	—
Michigan	7-4-0	23	22	17	10	10	7
Clemson	8-3-0	24	23	—	—	—	—
Michigan State	6-3-0	25	24	25*	—	—	—
Southern California	7-5-0	—	25	23	15	24	12
Fresno State	8-3-0	—	—	20	—	12	—
Louisville	7-3-0	—	—	22	—	—	23
Washington	7-4-0	—	—	—	18	19	14
Arizona State	6-4-0	—	—	—	23	—	17
Virginia	7-4-0	—	—	—	24	23	24
Memphis State	6-4-0	—	—	—	—	22	—

"I've used a fairly standard least squares method, and I come close to what's in USA Today," Stern says. "The small differences, I think, probably have to do with weighting games differently as the season progresses or using information from previous seasons."

But most people can't quite bring themselves to trust the computer-generated rankings. After the Notre Dame-Florida State game, the New York Times rankings left Florida State, despite its loss, in first place and relegated winning Notre Dame to third. The USA Today computer poll ranked Florida State first and Notre Dame second. Things got even more confused in subsequent weeks.

Fans find such counterintuitive results a little hard to swallow. Besides, the computer polls don't even agree among themselves, they complain.

College football is not the only sport in which rankings and ratings play a role. When the National Collegiate Athletic Association (NCAA) chooses the participants for its annual basketball tournament, it uses a rating scheme to help determine which teams should fill unclaimed spots in the field of 64.

"They have a couple of fudge factors, and they play with those factors to make it come out right," Stern says. "The problem is that they now seem to be using the same fudge factors in other sports." Those factors may distort the selection process. Responding to complaints from Harvard's hockey coach, Stern is trying to obtain data to assess whether the same fudge factors that apply to basketball teams would also work for hockey.

"It's a very tricky problem," Stern admits. "The key point is to try and define carefully what you would like to do [with the ratings]."

For example, if the ratings are to be used to choose teams for participation in a championship playoff, it's inappropriate to use information from prior seasons, he argues. If prediction of future games is the ultimate goal, then information from prior seasons may be useful.

In the end, when it comes to rating or ranking teams, the value of any particular method resides in the mind of the beholder.

"There is no unique way to devise a ranking scheme," Keener says. "The different ranking schemes give different rankings because they weigh important factors differently." The same conclusion applies to rating methods.

And there's always the temptation to fiddle with the mathematics to obtain a result more in line with human thinking, in effect enshrining preconceived notions of how things should turn out.

"Each of the schemes proposed . . . has strengths and weaknesses," Keener adds. "But invariably when a method is tweaked to get rid of some 'undesirable' feature, another 'counterintuitive' result shows up."

Nonetheless, mathematical ranking schemes do remove some — though not all — of the subjectivity that plagues the national polls. And there's a lot of entertainment value in these methods.

"I find that the few minutes I spend each week during the fall collecting and entering data for my computer program are justified by the increased student interest in the mathematics of the methods," Keener writes in the March SIAM REVIEW. "It is not difficult for students to write their own computer programs to test some of these ideas on their favorite competition."

So let the games begin, and may the best numbers win. □