

There have been multiple attempts to quantify a team's performance based only on wins, losses, and strength of schedule. The most well-known example is RPI, the system used by the NCAA in basketball and other sports. Other examples include the Colley Matrix, used by college football, and Elo, used most popularly in Chess. Each of these has its advantages, but none of them are truly objective.

The Rating Percentage Index (RPI)

The [RPI formula](#) is the easiest to find fault in. Developed by Jim Van Valkenburg in 1981, it has a very [interesting history](#). The advantage of RPI is that it's not a complex calculation, though at the time it was developed, it was still a large task for computers to perform. The development of a better system was limited by the computing power at the time. As a result, the formula was very simple. It's been tweaked some over the years, but the current baseline formula is:

$$RPI = \frac{1}{4}(Team's\ win\ \%) + \frac{1}{2}(Opponents'\ win\ \%) + \frac{1}{4}(Opponents'\ opponents'\ win\ \%)$$

However, despite its simplicity, there is no reason to believe that this method would give a trustworthy rating, as it ["lacks theoretical justification from a statistical standpoint"](#). It's nothing more than an arbitrary formula that appeared to give ratings that made some intuitive sense, and was capable of being calculated in the 1980's. It's only still being used because it's been grandfathered in. [Many others](#) have gone into [great detail](#) on its [specific shortcomings](#), so there's no need to belabor this point here.

The Colley Matrix

The [Colley Matrix](#) is a system developed by Wes Colley. He makes a good attempt at a truly unbiased ranking system. His explanation of why margin of victory and other factors need to be ignored is [very good](#). However, there are two problems with [his math](#).

Here's Colley's method:

| | | |
|--------|-------------------------------------|---|
| Line 1 | $p = \frac{1 + s}{2 + n}$ | This is the Laplace Rule of Succession. n = number of events s = number of successes p = probability of a success Assumptions: $0 \leq s$ and $s \leq n$ |
| Line 2 | $r = \frac{1 + n_w}{2 + n_{tot}}$ | Applying the Rule of Succession to game results n_{tot} = total number of games played n_w = number of wins r = probability of a success (in this case, a success is a win). Also represents a team's rating because a higher probability of winning represents a better team, earning it a better rating. Assumptions: $0 \leq n_w$ and $n_w \leq n_{tot}$ |
| Line 3 | $r = \frac{1 + n_w}{2 + n_w + n_l}$ | Total games played equals the number of wins plus the number of losses n_l = number of losses |

So far so good. Changing gears a bit:

| | | |
|--------|---|---|
| Line 4 | $n_w = \frac{n_w - n_l}{2} + \frac{n_w + n_l}{2}$ | If you simplify this, the n_l terms will cancel, so this is true. |
|--------|---|---|

Also:

| | | |
|--------|--------------------------------------|--|
| Line 5 | $\frac{\sum r}{n_w + n_l} = \bar{r}$ | This is just calculating an average $\sum r$ = sum of ratings of all the opponents that a team faces \bar{r} = average rating of all the opponents that a team faces |
|--------|--------------------------------------|--|

Now, if we make the assumption that the average rating of opponents is equal to 0.5, it follows that:

| | | |
|--------|--|-----------------------------|
| Line 6 | $\frac{\sum r}{n_w + n_l} = \frac{1}{2}$ | Assumption: $\bar{r} = 0.5$ |
| Line 7 | $\sum r = \frac{n_w + n_l}{2}$ | Rearranging |

Now, if we take the equation from Line 4, and sub in what we know from Line 7, we get:

| | | |
|--------|--------------------------------------|-----------------------------|
| Line 8 | $n_w = \frac{n_w - n_l}{2} + \sum r$ | Assumption: $\bar{r} = 0.5$ |
|--------|--------------------------------------|-----------------------------|

From there, Colley uses the equation in Line 8 to find a team's effective number of wins, then plugs that into the numerator of the equation in Line 3 to come up with a team's rating.

The first problem with this derivation is that it isn't a valid proof that can trace its way back to the [LaPlace Rule of Succession](#). There is a problem at Line 6. Basic algebra requires that in order to make a legal substitution (in this case, replacing \bar{r} with 0.5), it first needs to be proven that both expressions are equal. When actually running the calculation, \bar{r} is not necessarily equal to 0.5, so this is not a legal substitution mathematically. Colley's method is only a method, it's not a valid proof. When that substitution is made, it loses its basis on the Rule of Succession; it becomes nothing more than another equation that ranks teams, just as arbitrary as RPI.

Another assumption that is violated is on Line 2. The Rule of Succession requires that the effective number of wins (i.e., number of successes) must be zero or greater, but when Colley runs his calculations, [that's not always the case](#). If you use negative numbers, then you won't get a valid answer from the Rule of Succession. The other assumption in Line 2 is that the effective number of wins cannot exceed the number of games played, which the Colley Matrix also violates. Intuitively, this makes sense, you can't have fewer than zero wins, and you can't have more wins than games played. Colley cannot claim his algorithm has a basis in the Rule of Succession because he violates its initial assumptions. There are even situations where teams can be punished for winning a game, and rewarded for losing.

The second major problem with the Colley Matrix is that it's actually a subjective system, despite Colley's claims. For the sake of argument, we'll assume the substitution from Line 8 is valid, and call it the "Colley Exception". Now, looking back at the equation on Line 4:

| | | |
|--------------------|---|---|
| Line 4 (repeat) | $n_w = \frac{n_w - n_l}{2} + \frac{n_w + n_l}{2}$ | If you simplify this, the n_l terms will cancel, so this is true. |
|--------------------|---|---|

The green term represents the component that is based on a team's winning percentage. The red term represents the component that factors in strength of schedule (after substituting with the Colley Exception). Colley arbitrarily chooses to split these in half, giving each component equal weight (arguably, a 50/50 split puts [too much weight](#) on strength of schedule). Instead of splitting 50/50, you could just as easily split it 75/25:

| | | |
|---------|---|---|
| Line 9 | $n_w = \frac{3n_w - n_l}{4} + \frac{1}{2} \left(\frac{n_w + n_l}{2} \right)$ | If you simplify this, the n_l terms will cancel, so this is true. |
| Line 10 | $n_w = \frac{3n_w - n_l}{4} + \frac{1}{2} \sum r$ | Substituting with the Colley Exception |

Or 80/20:

| | | |
|---------|---|---|
| Line 11 | $n_w = \frac{4n_w - n_l}{5} + \frac{2}{5} \left(\frac{n_w + n_l}{2} \right)$ | If you simplify this, the n_l terms will cancel, so this is true. |
| Line 12 | $n_w = \frac{4n_w - n_l}{5} + \frac{2}{5} \sum r$ | Substituting with the Colley Exception |

Or 41/59:

| | | |
|---------|--|---|
| Line 13 | $n_w = \frac{41n_w - 59n_l}{100} + \frac{59}{50} \left(\frac{n_w + n_l}{2} \right)$ | If you simplify this, the n_l terms will cancel, so this is true. |
| Line 14 | $n_w = \frac{41n_w - 59n_l}{100} + \frac{59}{50} \sum r$ | Substituting with the Colley Exception |

Or anything else. And mathematically, they're all just as valid as Colley's 50/50 split. Colley hangs his hat on the fact that his formula has no "ad hoc" or "biased" adjustments or constants. But contrary to his claim, Colley is putting an arbitrary weight on strength of schedule, which is exactly the mistake he accuses other ranking methods of making.

In the end, the Colley Matrix ranks teams no better than RPI does. It's subjective and the equations used are not self-evidently true.

The Elo rating

The [Elo rating](#) at least has a basis of statistics. It was developed by Arpad Elo in 1960's as a way to rate chess players. The concept that "[performance of each player in each game is a normally distributed random variable](#)" is the same assumption made in the Bayesian Resume Rating (BRR). But there is still a mathematical problem with applying Elo to sports seasons.

The problem with Elo is that instead of making retroactive adjustments, it uses an arbitrary "[K factor](#)" to weigh the most recent match. The [amount of debate](#) over what to choose for [reasonable K factor](#) illustrates how arbitrary it is. A lower K factor reduces the amount of points awarded to a player for winning a game (and likewise reduces the points lost when a player loses a game). If the K factor is too

low, it would take a long time for players to settle into ratings that match their abilities. On the other hand, a higher K factor increases the amount of points awarded and lost. If the K factor is too high, players' scores would jump around erratically with each game played. A good balance between these two extremes is [hard to find](#), and hotly debated, because the appropriate K factor can vary depending on a variety of factors (e.g., total experience, playing frequency).

The K factor, subjective as it is, isn't necessarily wrong, it's just not completely objective. And its subjectivity is probably an unavoidable problem in rating chess players because they don't play in seasons. If applied to chess, the BRR could only be used to rank a closed scope of players within a given time period (like a college chess club over a single school year, for example), but would not be ideal for a worldwide rating system.

In the [Bayesian Resume Rating](#), when a team plays another game, its score changes, and that change affects all previous games that it played, which in turn affects all other teams, and the calculation is iterated until equilibrium is found. In the Elo rating, if a player is found to have been overrated or underrated, the opponents he played in the past don't get their scores adjusted for that discovery. Because sports teams play in seasons, and the purpose of the BRR is to measure a team's performance within a given season, the BRR is a better rating method than Elo when a league has a defined season.

Of systems that rank teams based only on wins, losses, and strength of schedule, only the BRR is a truly objective and mathematically sound formula. The BRR makes the assumption that teams and leagues behave within a bell curve, but beyond that assumption, there are no arbitrary weighting factors, and the math is based only on laws of probability.