

A New Wrinkle

by Rod Keefer

PITCH A BASEBALL STAT TO CONTINGENCIES....Funny you should ask - I was just thinking about this topic over the past few months, as the "value" of various players has been analyzed and performance measures, Hall of Fame credentials, etc., have been debated.

My first proposed statistic is what I call "run production index" (RPI). It equals (runs + RBIs - home runs) divided by (gross number of at bats). It measures the overall capacity of an offensive player to produce runs, which after all is what wins games. For example, if a player bats 550 times and scores 100 runs, has 100 RBIs, and hits 35 home runs, his RPI is $(100+100-35) / 550 = 0.300$ or 30.0%. That is, he contributes 3/10 of a run, on average, each time he appears at the plate.

There are a few things that are a bit peculiar about this statistic. First, walks and all other plate appearances not normally considered in a batting average would be counted in the denominator, as otherwise one could walk and score a run, thus contributing one to the numerator but none to the denominator. Secondly, the adjustment for home runs avoids someone getting credit for two runs when he produces just one (by driving himself in, that is). Finally, this measure corrects one of the deficiencies of traditional batting average, which arguably overstates the value of a "singles hitter" who does not drive in many runs relative to a batter with the same average who contributes more to overall run production. Yet it awards those hitters who are able to steal bases, advance on sacrifices, etc., to help manufacture runs.

A possible criticism of this statistic is that it favors hitters who either have strong hitters in the line-up behind them (increasing their chances at scoring runs) or ahead of them (increasing their chances of driving in runs). A second statistic, "relative run contribution" (RRC), would offset this weakness by dividing an individual's RPI by the team's RPI. So if the 0.300 RPI hitter above is on a team with an RPI of 0.200, his RRC would be $(.3/.2)$ or 1.5 or 50% above his "average" teammate. If he is on a team with a 0.250 RPI, then his RRC is 1.2 or 20% above the average.

Finally, I would propose a statistic that would credit, among other things, the act of sacrificing to move runners into scoring position, etc. It would be referred to as "bases advanced per at bat" (BAAB) and measure total bases reached by the player PLUS bases advanced by runners on base at the time divided by total number of at bats. So a sacrifice bunt that moves a runner to second would score $1/1$ or 1.00. A grand slam home run would score $(4+3+2+1)$ or 10.00.

Thanks for asking – we don't get many opportunities to tie our overzealous analytical energies with our passion for sports.

Bill James' response:

The statistic "runs produced" (runs scored + RBI – home runs) has been around at least since the 1950s, by that name . . . putting a denominator underneath it is, as far as I know, a new wrinkle.

I have never liked the statistic, because it discriminates unfairly against a home run hitter. Your explanation for this is that "the adjustment for home runs avoids someone getting credit for two runs when he produces just one (by driving himself in, that is)." But the real situation is that *all runs are counted twice EXCEPT home runs*. What kind of sense does that make?

It would be much MORE logical to take the home runs and add them back in again (that is, runs + RBI PLUS home runs), rather than taking the home runs out. Suppose that you take two situations, one in which a player hits a solo home run, and another in which Player A, Player B and Player C hit consecutive singles, leading to one run. Player A is credited with a Run Scored; Player C is credited with an RBI. The guy who is getting screwed here is Player B, whose contribution to the run sequence—uniting the front end with the back end—is uncredited.

There is a reasonable argument that the player who hits a home run should be credited with THREE—one for scoring the run, one for driving it in, and one for uniting the front end and the back end of the sequence. But comparing the solo home run to the sequence, you not only wish to deny the home run hitter the credit for uniting the portions, you wish to deny him the simple credit for providing both ends of the sequence. This, to me, is just indefensible.

Let's compare the 1965 Milwaukee Braves to the 1962 Chicago Cubs. Both teams played 162 games; the '65 Braves scored 708 runs and drove in 664; the '62 Cubs scored 707 runs and drove in 662.

By your method, the '62 Cubs have a much better offense:

	Cubs	Braves
Runs Scored	707	708
RBI	662	664
Home Runs	92	196
At Bats	5514	5542
Walks	620	408
Miscellaneous Plate Appearances:	163	127
RPI	.2028	.1935

But that's just not right; the Cubs don't have a better offense. If they had had a better offense they would have scored more runs.

With regard to the other suggestion . . . the idea of counting and crediting batters for bases that they advance other runners is not a bad one. I personally would not suggest doing so in the context of counting other things. If I wanted to know how many bases a batter had advanced runners, I would count that by itself, rather than mix it in with a count of total bases or place it in the context of at bats. It is much easier to count things by themselves and then combine them into meaningful combinations than it is to count things in combinations and then figure out how to make sense of them.