The One About Win Probability

by Dave Studeman

December 27, 2004

I talk a lot about Win Shares, because they do something I think is really valuable — they estimate the contribution each player has made to his team's wins. This is an entirely different way of thinking about players, stats and value — because it measures every baseball event within the context of the ultimate goal: winning games.

But Win Shares are not the only way to skin this cat. There is another process that goes by many names and has been "introduced" to the public many times. In fact, **a 2003 Business Week article** claimed that "a novel way to evaluate baseball players" had been invented by people who wanted to bottle, patent and sell it. Too late. It was first introduced by the Mills brothers in the early 1970's, and it's been done many times since.

As I said, it goes by many names: "Player Win Averages" (Mills brothers), "**Player Game Percentage**" (Bennett), "Win Probability Added" (Drinen), "**Win Expectancy**" (Tangotiger), "**Game State Wins**" (Rhoids website), "**Player's Win Value**" (Ed Oswalt) and WRAP (Lonergan and Polak). I'm sure there are other people who have done the same thing and given it a name that I've not acknowledged here. I apologize if you're one of those people.

For purposes of this article, I'm going to use the Drinen name, "Win Probability," because I think it's the most descriptive.

Here's the basic idea. An average team, at any point in a game, has a certain likelihood of winning the game. For instance, if you're leading by two runs in the ninth inning, your chances of winning the game are much greater than if you're leading by three runs in the first inning. With each change in the score, inning, number of outs, base situation or even pitch, there is a change in the average team's

probability of winning the game.

Christopher Shea has invented a **"Win Expectancy Finder"** to look up the actual Win Probability of every base/out, inning and score combination of all Major League games from 1979 to 1990. Chris used **Retrosheet** data that had been compiled by **Phil Birnbaum**, and his WE Finder simply looks up the percent of times a team in a given situation went on to win the game during those years. Next time you watch a ballgame, use it to track the ups and downs of the game. It will change the way you watch baseball.

Here's an example: Bottom of the ninth, score tied, runner on first, no one out. The home team has a 71% chance of winning according to the Win Expectancy Finder (in this situation, the home team won 1,878 of 2,631 games between 1979 and 1990). Let's say the batter bunts the runner to second. Good idea, right? Well, after a successful bunt, with a runner on second and one out, the Win Probability actually decreases slightly to 70% (home team won 1300 of 1,848 games), according to the WE Finder. The bunter hasn't really helped or hurt his team; his bunt was a neutral event.

If you're managing a team, or even following the game, you might want to know this sort of thing. Of course, the application of actual strategy (should he bunt or not?) depends on a lot of other factors, such as the skills of the batter, the pitcher and the baserunner, the following batters in the order, the game conditions and probably a number of other things. But Win Probability sets the baseline for evaluating each event on the field.

To really have fun with this system, you can take it one step further and track something Drinen calls "Win Probability Added" (WPA).

Once again, the concept is simple. Let's say our batter in the bottom of the ninth hits a single to put runners on first and third with no outs. This increases the Win Probability from 71% to 87%, for a gain of 16%. So, in a WPA system you credit the batter +.16 and debit the pitcher/fielder -.16. If you add up every positive and negative event from the beginning to the end of a game, you wind up with a total for the winning team of .5, and a total for the losing team of -.5. And the player with the most points will have contributed the most to his team's win.

By the way, that 87% with runners on first and third in the bottom of the ninth is on the low side for reasons I'll discuss in a minute.

If you were to track an entire season in this manner, you would have a Win Contribution metric that is more accurate than Win Shares, because it is based on how much each event actually contributed to the team's wins. In a way, WPA is the ultimate baseball statistic. And in a way, it is not.

Like Win Shares, WPA is not a good predictive statistic because it's not necessarily a good representation of a player's true talent. If a player hits a home run in the ninth inning of a 1-0 game, he is credited with more WPA points than if he hits a home run in the first inning of a 1-0 game. The talent is the ability to hit the home run; when it happens in a game is something that is pretty random. When you are thinking of acquiring a player for your fantasy team, you should rely more on the traditional sabermetric stats, like Linear Weights, Runs Created, DIPS, etc. etc.

Also, WPA measures the impact of an event while the game is in progress, not after the game is over. After the game is over, the score is 1-0, and it doesn't matter when the batter hit the home run. But during the game, it matters a lot. Good managerial strategies, for instance, are based on an implicit understanding of Win Probabilities. And if there is such a thing as clutch performance, WPA might unearth it.

The most interesting and useful application of Win Probability Added — the one that Drinen, Tangotiger and others have spent a lot of time on — is the evaluation of relief pitchers and the managers who call on them. We all know that closers are important, even though they pitch less than 100 innings a year, right? Why? Because they pitch the most important innings.

Using the WE Finder again, if a pitcher gives up a bases-empty home run in the first inning of a tie game, his team's Win Probability decreases about 10%. If he does the same thing in the eighth inning, it decreases about 25%, because his team has less time to come back. In this context, the eighth inning is about 2.5 times more important than the first inning. And if you apply this sort of analysis to every appearance made by a relief pitcher, you can quantify the importance of all of his innings pitched.

More from The Hardball Times

A Hardball Times Update
by RJ McDaniel
Goodbye for now.

Tangotiger developed a system called "Leveraged Index" that measures and sums the potential "Win Probability Added" for each pitcher's appearance. Doug Drinen tracked a similar measure, called "P," in the Big, Bad Baseball Annual. Though the math behind each system differs, they are both constructed to measure the importance of relief innings. You might particularly enjoy Tango's **Crucial Situations article and chart**.

Win Shares, by the way, includes an approximation of WPA for relief pitchers, based on each pitcher's saves and holds. I've been playing with a similar system myself, and I hope to roll out some analysis in the next few weeks.

There are a couple of reasons the Win Expectancy Finder isn't the best source of Win Probabilities. First, it's based on the years 1979 through 1990, when there were fewer runs scored per game than in the past few years. A one-run lead was safer back then. Also, there are sample size issues with some of the situations. For instance, there were only 220 games with runners on first and third for the home team in the bottom of the ninth with the score tied. That's not a large enough sample size. So you shouldn't take the numbers in the WE Finder as "gospel."

The better way to develop your Win Probability table is to develop something called "Markov Chains." I won't go into all the math here (because I doubt I can really explain it well), but suffice to say that a proper Win Probability table is something a good mathematician can concoct, based on the probability of scoring a certain number of runs for each base/out situation.

And there are still a lot of Win Probability issues to be resolved. For instance, Win Probability tables really should be altered based on the home park. To track WPA on a regular basis, you need play-by-play data, so you can't create it for most of baseball history. And Win Probability doesn't solve the sticky issue of splitting credit between pitching and fielding (something Lonergan and Park admit).

Win Probability is a complicated subject, and there's so much more I could say. But I hope this article serves as a good introduction to a topic I plan to return to in the future.

References & Resources
Here's **an example of a game I tracked** with my own Win Probability tables.

Cyril Morong has a **nice review of Win Probability hitting stats** on his website. If you're a Baseball Prospectus subscriber, you can find **a table of Win Probabilities based on 2004 games** in their stats section.

I recommend Alan Schwarz's "**The Numbers Game**" for a very nice history of the evolution of Win Probability (including the critical role of George Lindsey). My thanks go to Tangotiger, Doug Drinen and Jon Daly for their support and education in this subject.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Comments are closed.

What WPA Can Tell Us About Teams

by Dave Studeman

July 13, 2006

Win Probability Added (or, just WPA to its friends) has received a lot of good and bad publicity lately. After Alan Schwarz covered WPA in **his Sunday *New York Times* column**, there was a **lively debate between Tangotiger and MGL on The Book Blog** (the Cliff Notes version: MGL hates it and finds it utterly useless and Tango thinks it's fascinating). David Pinto also weighed in with **a negative opinion of ranking players by WPA**.

You may not be surprised to learn that I am in the "fascinated" camp, though I do acknowledge WPA has limits. WPA is **a very simple idea** (which is part of its appeal to me): calculate the odds of a team winning a game as the game progresses, based on calculations and historical baseball stats, and assign responsibility for changes in WPA to the players involved in each play. I think WPA is a **great way to track a ballgame**, evaluate in-game tactics and assess the contribution of relievers. I also think it may have some value in identifying "clutch" performers and as a secondary consideration for picking MVP candidates.

Having said that, I've never really played with a full season's worth of WPA stats, or even a half season's. Thanks to the tremendous effort of David Appelman at **Fangraphs.com**, we now have a half-season's worth. So let's learn a little bit more about WPA. In fact, let's not even step into the player evaluation brouhaha, let's just talk about teams. I hope to show you that WPA can help us solve one of the more vexing questions baseball fans often ask.

Team Leverage Index

Leverage Index (LI) was invented by Tangotiger. It's used to measure **the criticality of each plate appearance**, and it's perhaps the best tool to come out of WPA. LI is set so that the overall average of a plate appearance is

1.0. Ace relievers have often achieved an LI of 2.0 or more, meaning that their appearances were twice as critical as average. So I wondered, are there differences between teams too?

To find the answer, I used the LI of each player on a team and weighted it by each player's plate appearances (or batters faced) for a team LI. I calculated two separate LI's, one for the team's batters and one for the team's pitchers, and found that the average LI on both sides ranges roughly from 1.1 to 0.9. Here's the list of this year's teams:

Pitching LI		Batting LI	
Team	LI	Team	LI
NYN	1.07	PIT	1.08
TB	1.07	ATL	1.07
OAK	1.05	SD	1.07
FLA	1.03	OAK	1.06
HOU	1.03	MIL	1.05
MIL	1.02	WAS	1.03
DET	1.02	PHI	1.03
STL	1.01	NYA	1.02
ATL	1.01	NYN	1.02
PIT	1.01	BAL	1.00
PHI	1.01	SF	1.00
SF	1.00	SEA	1.00
NYA	1.00	KC	1.00
SD	0.99	COL	0.99
ARI	0.99	STL	0.99
WAS	0.99	ARI	0.99
LAN	0.98	BOS	0.98
COL	0.98	LAA	0.98
SEA	0.98	LAN	0.97
CIN	0.98	FLA	0.97
TEX	0.98	CIN	0.97
KC	0.98	TEX	0.97
CHA	0.97	MIN	0.97
BOS	0.96	HOU	0.96
BAL	0.95	TB	0.96
LAA	0.94	CHA	0.96
TOR	0.94	CHN	0.94
MIN	0.92	TOR	0.94
CLE	0.91	DET	0.90
CHN	0.91	CLE	0.89

In general, teams that have played close games will have the highest LI and those that have played in the most runaway games will have the lowest. For instance, the Athletics rank high in both pitching and batting LI, and they have played more close games (50 games won by two runs or less) than any other major league team.

The Mets pitchers are at the top of the pitching list, they have faced more/bigger critical situations than any other team. Meanwhile, the Pirates batters have faced the most critical batting situations. There are also some big differences between batters and pitchers on the same teams. For instance, the Tiger and Devil Ray batters don't rank highly, but their pitchers have faced a relatively high number of critical situations. I guess the Indians and Cubs have played the most boring games, judging by the low LI's for both their batters and pitchers.

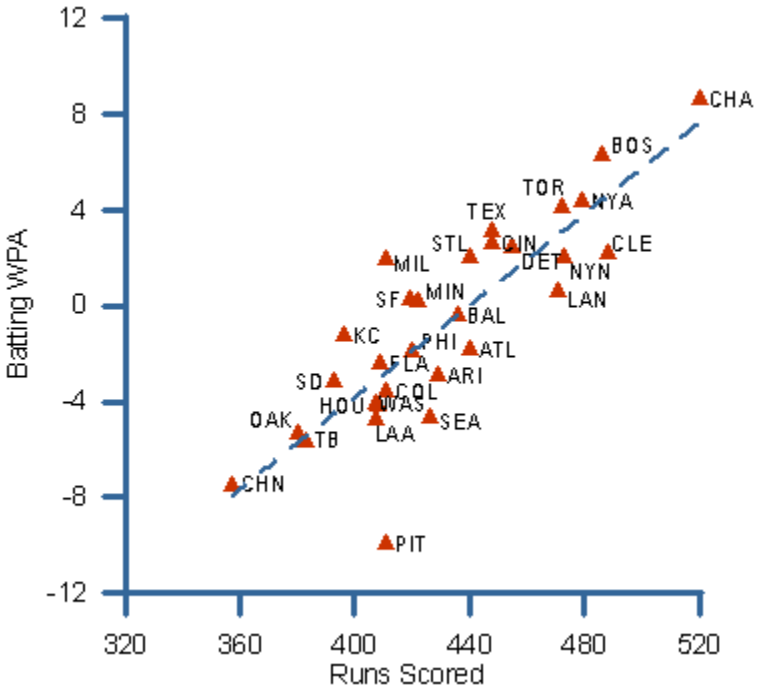
Team WPA

Next, let's look at each team's WPA rankings. As you can imagine, batting WPA and pitching WPA closely follow total runs scored and runs allowed. But there are some differences, as the following table shows:

Batting WPA			Pitching WPA		
Team	Total	RS	Team	Total	RA
CHA	8.57	520	DET	12.61	328
BOS	6.26	486	SD	7.70	369
NYA	4.34	479	NYN	6.51	404
TOR	4.10	472	OAK	6.41	394
TEX	3.08	448	CHA	4.43	415
CIN	2.59	448	COL	4.10	399
DET	2.39	455	MIN	3.87	396
CLE	2.14	488	LAA	3.81	416
NYN	1.99	473	BOS	3.74	413
STL	1.98	440	SEA	3.25	421
MIL	1.91	411	STL	2.85	425
LAN	0.57	471	HOU	2.74	435
SF	0.21	419	NYA	2.66	406
MIN	0.13	422	ARI	1.86	450
BAL	-0.46	436	LAN	1.43	416
KC	-1.30	396	TOR	0.90	432
ATL	-1.89	440	SF	0.29	407
PHI	-1.97	420	TB	0.18	457
FLA	-2.48	409	PHI	-1.53	454
ARI	-2.96	429	TEX	-2.09	427
SD	-3.20	393	CIN	-2.09	463
COL	-3.60	411	CHN	-2.48	448
WAS	-4.11	407	FLA	-2.52	420
HOU	-4.24	408	ATL	-2.61	449
SEA	-4.75	426	WAS	-2.89	470
LAA	-4.81	407	MIL	-2.91	485
OAK	-5.41	380	BAL	-3.53	501
TB	-5.68	383	PIT	-5.00	474
CHN	-7.52	357	CLE	-5.64	443
PIT	-10.00	411	KC	-11.20	528

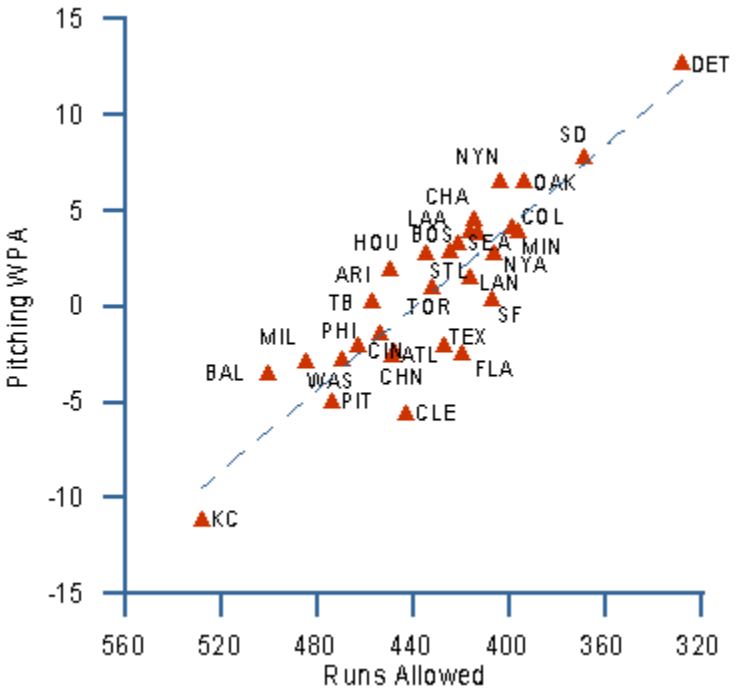
Well, that doesn't really work. I wanted to show you all the data, but the table is kind of overwhelming. Time for a graph; here's a picture of how each team's

runs scored compares to its batting WPA. Teams above the line have gained relatively more wins with their bats compared to total runs scored, while teams below the line have contributed relatively less of their runs to winning. Check out those Pirates, who have faced more critical situations than any other team's batters. As the graph shows, they haven't delivered:



Why do we see variances? Two reasons, I think. One, some teams simply have more opportunities to impact a game than others. Second, some teams actually deliver more in crucial situations than other teams do. LI measures the opportunities. WPA reflects both the opportunities and the actual production. For instance, the Pirates are only batting .220 in "close and late" situations. Combine that with their high LI and you get a really bad WPA.

Here's the same graph, except for the pitchers. In this graph, I inverted the "runs allowed axis" so that teams above and below the line will have the same impact on their teams' probability of winning as in the previous graph:



When it comes to pitching WPA, the bullpen and its management have a big impact. For instance, the Indians have had the lowest relative win impact from their runs allowed because their bullpen WPA is the worst in the majors at -5.09. That's partly because their bullpen LI is also the lowest (0.88) and mostly because they've been lousy (4.86 ERA).

Pythagorean Breakouts

This isn't just an academic graphing exercise. In fact, we can get something quite useful out of this stuff. You know the **Pythagorean Formula**? Invented by Bill James, it projects a team's won/loss record from its runs scored and allowed and it's typically very accurate. Baseball analysts like to track teams that vary from their pythagorean formula to see why and how those teams win more or less than predicted. WPA gives us a new way to approach that problem.

Here's how. First, I used regression analysis to derive formulas that would predict each team's batting WPA based on its runs scored and pitching WPA based on its runs allowed. As you can imagine, the R squared between WPA and runs is high (between .7 and .8) but not perfect. Next, I ran that formula for each team to see how much the team deviated from its predicted batting and pitching WPA. When I combined the two differences I got a number that is almost exactly each team's variance from its Pythagorean Formula.

Let me see if I can put that in English. WPA gives us a way to assess how teams are exceeding or falling short of their predicted performance (based on runs allowed and runs scored). Specifically, it allows us to allocate the difference to each team's offense and defense. The following table is a list of each team's batting and pitching pythagorean contribution (listed in the second and third columns) based on the WPA analysis. The column labeled "Tot" is the total of the previous two columns, and the column labeled "Pyth" is the actual pythagorean variance for each team.

	Bat	Pitch	Tot	Pyth
MIL	4.68	1.93	7	6
CHA	0.91	1.85	3	3
BOS	1.85	0.95	3	3
STL	1.97	1.33	3	3
DET	0.95	0.81	2	2
NYN	-1.18	2.77	2	2
OAK	0.33	1.60	2	2
TB	-0.24	2.06	2	2
CIN	1.81	0.42	2	2
BAL	-0.09	3.00	3	2
SD	1.29	0.25	2	1
MIN	1.84	-0.72	1	1
HOU	-1.19	2.28	1	1
ARI	-1.91	3.00	1	1
TOR	1.03	0.12	1	1
LAA	-1.66	1.34	0	0
NYA	0.60	-0.87	0	0
PHI	-0.07	0.03	0	0
COL	-0.84	-0.17	-1	-1
SF	2.21	-3.14	-1	-1
TEX	2.31	-3.39	-1	-1
CHN	0.42	-1.56	-1	-1
WAS	-0.96	0.36	-1	-1
KC	2.90	-1.80	1	-1
SEA	-3.42	1.30	-2	-2
LAN	-2.41	-1.04	-3	-3
FLA	0.47	-4.56	-4	-4
ATL	-1.90	-1.58	-3	-4
CLE	-2.46	-5.25	-8	-8
PIT	-7.23	-1.32	-9	-9

For example, the Pirates have the worst pythagorean variance in the majors and are on a pace to set the record for most one-run losses ever. As you can see in this table, the issue lies with the team's batters, who are a whopping 14 wins (or seven games) below what their offense "should" have contributed to the team's win-loss record.

The Indians have the second worst pythagorean variance, and you can see that their problem is the aforementioned bullpen. Bullpens are often cited as

culprits in pythagorean variances but that isn't always the case. The team with the highest pythagorean variance, the Brewers, has been led by its batters. They've added 4.7 games to the variance, while the pitchers have added just 1.9.

This is just one use of WPA that I found while playing with the stats, and I'm sure there are many more. Pretty fascinating, huh?

References & Resources

As always, many thanks to Tangotiger for educating us about WPA and making his LI tables available to Fangraphs. Also, thanks to Keith Woolner and **Baseball Prospectus** for their research in this area. And a huge thanks to **David Appelman and Fangraphs**.

What WPA Can Tell Us About Players

by Dave Studeman

July 24, 2006

Last week, we took a little time to review this new-fangled stat called **Win Probability Added**, and we compared WPA to the number of runs scored and allowed by teams. Today, I'm going to do the same thing, but I'm going to do my best to extrapolate runs to individual players and analyze some of the key differences between WPA and runs created/avoided for individual batters/pitchers. In the end, we'll have an interesting take on which players have "stepped up" with the game on the line.

But first, a couple of sidebars:

Starters and Relievers

Here are the total innings pitched and pitching WPA allocated to each type of pitcher (starter or reliever, based on my sorting pitchers into each bucket). I've also included the average Leverage Index of each type of pitcher:

POS	IP	WPA	LI
SP	15,643	3.93	0.96
RP	7,897	20.95	1.04
Total	23,539	24.87	0.99

As you can see, relievers received about 17 WPA points (or 34 wins) more than starters, despite pitching in only half as many innings. Also, relievers have pitched in more crucial situations, on average, with a Leverage Index of 1.04. So few numbers, so much information.

Relievers have a number of advantages over starters in WPA world:

- It's easier to relieve than start. As ***The Book*** has documented, relievers probably have as much as a 0.80 ERA advantage over starters.
- Managers choose when to bring in specific relievers. Bad relievers tend to appear in blowouts (when they can't have much effect on a game) and good relievers appear in close games, when they can have a big impact. There is no such contrast among starters—the best and worst starters enter games at the very same time, the first inning, regardless of the score.
- WPA only looks at games above and below .500, not total wins and losses.

We can mathematically adjust WPA for all three of these issues; today, let's tackle the last one.

Thanks to my **good buddy and THT colleague David Gassko**, we were able to mathematically estimate the total number of wins and losses that would be allocated to pitchers by WPA. I've put the specific calculation in the footnote; I'll just post the totals here:

POS	W	L	W%	WPAB
SP	429	422	.505	174
RP	254	212	.545	114
Diff	175	210	---	60

Now you can see one of the problems with looking at just wins above average. Starters actually have 175 more wins than relievers; the problem is that they also have 210 more losses and a lower winning percentage. If you take a slightly different approach and calculate Win Probability Above Baseline (I used a .300 winning percentage as a baseline), you find that starters contribute over 50% more WPAB than relievers (174 vs. 114).

By the way, this is the same issue that plagues Win Shares. Ideally, you really need to know Loss Shares, or use an alternative baseline, to get a clear idea of how much a player contributed to his team's won/loss record. That's why we like to use **Win Shares Above Bench** in our Win Shares articles.

This new table still doesn't resolve the first two WPA issues noted above; it's relatively easier to relieve and managers "mix and match" relievers with the situation. We'll save those adjustments for another day.

All Players

The above stats might make you wonder how all positions rank by WPA. Here are the WPA totals by position (based on which position each player has played most often). Starting pitchers have a much lower WPA total in this table because their batting performances are included.

POS	WPA
C	-13.60
1B	14.42
2B	-9.34
SS	-5.66
3B	9.23
OF	21.07
DH	2.48
SP	-36.21
RP	18.76

The key thing to remember is that Fangraph's WPA doesn't reflect fielding contributions, just the offensive contributions of the everyday positions. In general, this table reflects the relative fielding value of each position (the higher the WPA, the more emphasis on batting for that position), but look at how poorly designated hitters are doing. Make that an exclamation point!

Individual Pitchers

Here is the WPA leaderboard for all pitchers. I've included the 23 pitchers with at least 2 WPA. Note that the list is pretty evenly divided between starters and relievers, though relievers are at the top of the board.

Pitcher	Team	IP	ERA	WPA	LI
Jonathan Papelbon	BOS	46	0.59	3.68	1.76
B.J. Ryan	TOR	43	0.84	3.36	1.62
Brandon Webb	ARI	139	2.65	3.29	0.99
Bobby Jenks	CHA	41	2.83	3.22	2.04
Trevor Hoffman	SD	35	1.03	3.01	1.91
Joe Nathan	MIN	36	1.75	2.90	1.54
Jason Schmidt	SF	126	2.78	2.85	1.13
Jeremy Bonderman	DET	120	3.46	2.83	1.02
Joel Zumaya	DET	43	2.08	2.59	1.74
J.J. Putz	SEA	43	2.11	2.57	1.38
Francisco Liriano	MIN	88	1.83	2.55	0.82

Roy Halladay	TOR	129	2.92	2.37	0.97
Chris Young	SD	110	3.12	2.32	0.96
Chris Ray	BAL	37	3.19	2.30	1.82
Roy Oswalt	HOU	120	3.15	2.30	1.04
Chris Carpenter	STL	111	2.92	2.28	1.05
Johan Santana	MIN	131	2.95	2.15	0.95
Brad Penny	LAN	108	2.91	2.07	0.98
Mariano Rivera	NYA	46	1.76	2.06	1.80
Chris Capuano	MIL	129	3.21	2.03	0.97
Francisco Rodriguez	LAA	37	2.89	2.02	1.93
Duaner Sanchez	NYN	48	2.45	2.01	1.53
Scott Kazmir	TB	116	3.27	2.00	1.10

These are rankings as of the All-Star break. I'd say that **Brandon Webb** is more likely to retain his ERA (2.65) than **Jonathan Papelbon** (0.59) or **B.J. Ryan** (0.84) are, so you may see a starter at the top of the list by the end of the year.

Individual Batters

Here are the individual batting WPA leaders. To no one's surprise, **King Albert** is at the top of the list. But you may be surprised to see that **Jermaine Dye** is third and even more surprised to see that **Barry Bonds** is fifth. I've also included each batter's Base Runs Above Average (see my explanation of BRAA in the footnote).

Batter	WPA	BRAA
Albert Pujols	5.52	32
David Ortiz	3.40	19
Jermaine Dye	3.26	24
Derek Jeter	3.18	14
Barry Bonds	2.86	14
David Wright	2.62	23
Bobby Abreu	2.57	17
Travis Hafner	2.52	33
Manny Ramirez	2.51	23
Jim Thome	2.43	29
Jason Giambi	2.41	23
Nomar Garciaparra	2.36	23
Chase Utley	2.32	17
Ryan Zimmerman	2.31	5
Todd Helton	2.26	12

Lance Berkman	2.24	24
Frank Catalanotto	2.21	8
Curtis Granderson	2.21	9
Melvin Mora	2.21	1
Justin Morneau	2.20	16

More from The Hardball Times

A Hardball Times Update
by RJ McDaniel
Goodbye for now.

"Clutch" Batters

In the last article, we used WPA to identify whether batting or pitching was driving each team's **pythagorean variance**. I'd like to apply the same thinking to individual batters. The end result will be a list of which batters have contributed more WPA than their general offensive stats would predict.

To predict WPA, I regressed each batter's Base Runs Above Average against WPA (R-Squared of .6 for you math types) and calculated the difference between the actual WPA and predicted WPA for each batter. The difference is the result of a batter's relative performance in high vs. low leverage situations. You might call high-leverage situations "clutch," so I'm calling the difference "Clutch WPA" or ClWPA in the table. That's not a suggested label, by the way. Just one to use for this article.

Surprise! The best clutch batter, by this definition, is Milwaukee's **Geoff Jenkins**.

Batter	Team	WPA	BRAA	ClWPA	LI
Geoff Jenkins	MIL	2.05	-3	2.43	1.07
Albert Pujols	STL	5.52	32	2.28	1.06
Melvin Mora	BAL	2.21	1	2.10	0.96
Mark Loretta	BOS	1.24	-6	1.86	1.00
Ryan Zimmerman	WAS	2.31	5	1.79	1.12
Derek Jeter	NYA	3.18	14	1.76	1.07
Ken Griffey Jr.	CIN	2.01	4	1.68	1.07
Barry Bonds	SF	2.86	14	1.45	0.98
David Ortiz	BOS	3.40	19	1.44	0.97

Frank Catalanotto	TOR	2.21	8	1.43	1.03
Michael Young	TEX	1.45	1	1.36	0.92
Jeff Francoeur	ATL	0.82	-5	1.33	1.12
Ty Wigginton	TB	0.83	-4	1.31	0.93
Ramon Hernandez	BAL	1.78	5	1.31	1.03
Curtis Granderson	DET	2.21	9	1.30	0.89
Orlando Cabrera	LAA	1.67	4	1.29	0.99
Gabe Gross	MIL	1.43	2	1.27	1.10
Jason Varitek	BOS	0.13	-10	1.21	1.07
Jay Payton	OAK	0.61	-5	1.20	1.09
Alex Cintron	CHA	0.77	-3	1.12	1.05
Hank Blalock	TEX	1.16	1	1.11	1.00
Todd Helton	COL	2.26	12	1.08	0.90
Jacque Jones	CHN	1.83	8	1.08	0.98
Nick Swisher	OAK	2.17	11	1.07	1.08
Paul Lo Duca	NYN	0.93	-1	1.04	0.97
Marcus Giles	ATL	0.81	-2	1.03	1.02

Jenkins has had some key hits in high-leverage situations. For instance, he hit a two-run single in the bottom of the ninth off **Bob Wickman** to beat the Indians, 3-2 on June 17. That game alone was worth .43 WPA, almost an entire win. According to STATS Inc., Jenkins has batted .377/.426/.604 in "close and late" situations. His WPA reflects that.

This is certainly a very interesting list of batters. Remember that this list reflects the **difference** between predicted WPA and actual WPA, which means that these leaders have performed relatively better in high-leverage situations. It doesn't mean these are the best batters you want at bat in key situations. Personally, I'd still pick Pujols in clutch situations over Jenkins.

Individual Pitchers

For a final list, let's calculate the same "clutch" performance for pitchers. In this analysis, we'll used **Runs Saved Above Average** instead of Base Runs. RSAA presents some issues when comparing relievers and starters, but it seems like the best alternative in a pinch. We'll also include Leverage Index in the initial regression because there are significant differences between pitchers in leverage opportunities. For you mathematicians, the R-Squared for the regression is .73.

Here's the list of all pitchers with at least 0.75 "ClWPA." As you can see, the top of

the list consists entirely of the game's best relievers.

Pitcher	Team	WPA	LI	ClWPA
Bobby Jenks	CHA	3.22	2.04	2.02
Joe Nathan	MIN	2.90	1.54	1.55
Trevor Hoffman	SD	3.01	1.91	1.49
Chris Ray	BAL	2.30	1.82	1.48
Jonathan Papelbon	BOS	3.68	1.76	1.45
B.J. Ryan	TOR	3.36	1.62	1.41
J.J. Putz	SEA	2.57	1.38	1.20
Joel Zumaya	DET	2.59	1.74	1.19
Francisco Rodriguez	LAA	2.02	1.93	1.14
Jorge de la Rosa	MIL	-0.02	0.83	1.13
Jason Marquis	STL	0.08	0.86	1.10
Jim Brower	TOT	-0.46	0.39	1.08
J.C. Romero	LAA	0.20	0.70	1.05
Tom Martin	COL	1.06	0.68	1.01
Jeremy Bonderman	DET	2.83	1.02	0.99
Juan Rincon	MIN	1.93	1.28	0.98
Billy Wagner	NYN	1.88	1.99	0.95
Duaner Sanchez	NYN	2.01	1.53	0.95
Horacio Ramirez	ATL	0.76	1.26	0.86
Chad Harville	TB	0.86	0.99	0.84
Fernando Rodney	DET	1.02	1.66	0.81
Oliver Perez	PIT	-1.17	0.96	0.79
Dave Borkowski	HOU	0.87	0.41	0.78
Roberto Novoa	CHN	0.14	0.42	0.76
Carlos Silva	MIN	-1.62	0.86	0.75

Mariano Rivera? He's 74th on the list, with a ClWPA of 0.43.

There is so much we can do with Win Probability. This article, though it contains a lot of information, only scratches the surface. But you'll be happy to know that David Appelman is working on a number of upgrades to the **Fangraphs site** that will permit you to view WPA in even more detail. And this year's Hardball Times Annual (which is **now available for preorder**) will contain a veritable plethora of WPA stats and analysis. Yes, there is more WPA to come!

References & Resources

Many thanks to **Fangraphs** for calculating WPA each day, and to **Tangotiger** for the underlying math.

To calculate pitcher wins and losses, David and I used LI*IP/9*.51 to calculate "expected wins." Actual WPA wins were calculated as Expected Wins plus WPA and losses were calculated as Expected Win minus WPA. We used .51 instead of .5 (remember, one win equals .5 WPA) because the tables behind Fangraphs' WPA are based on a slightly higher run environment than the one we've experienced so far this year. That's also why total pitching WPA doesn't equal zero. If we wanted to show totals in which pitching and hitting contributed equally to a team's wins, we would have used .255.

Tangotiger reminded me about a site that **ranks all pitchers in WPA from 1972-2002**. As you can see, top starters lead top relievers on the list over entire careers, which is illustrative of how an entire career's stats may be necessary to properly use WPA. This list makes you wonder why **Goose Gossage** isn't in the Hall of Fame yet. Some folks might also be interested in **Bert Blyleven's** ranking.

I calculated Base Runs, created **many moons ago by David Smyth**, for batters. Base Runs is similar to Runs Created in many ways, but it has a number of advantages for certain analyses. For instance, in this case I set the "run scoring multiplier" on a team-specific basis, and then calculated the team's runs with and without each specific batter. In that way, my Base Runs totals equal the team's total runs scored. That makes this analysis more consistent with last week's article. By the way, it took ten Base Runs to "predict" one WPA, which is very consistent with other run/win ratios.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Comments are closed.

10 Lessons I Have Learned about Win Probability Added

by Dave Studeman

April 29, 2014

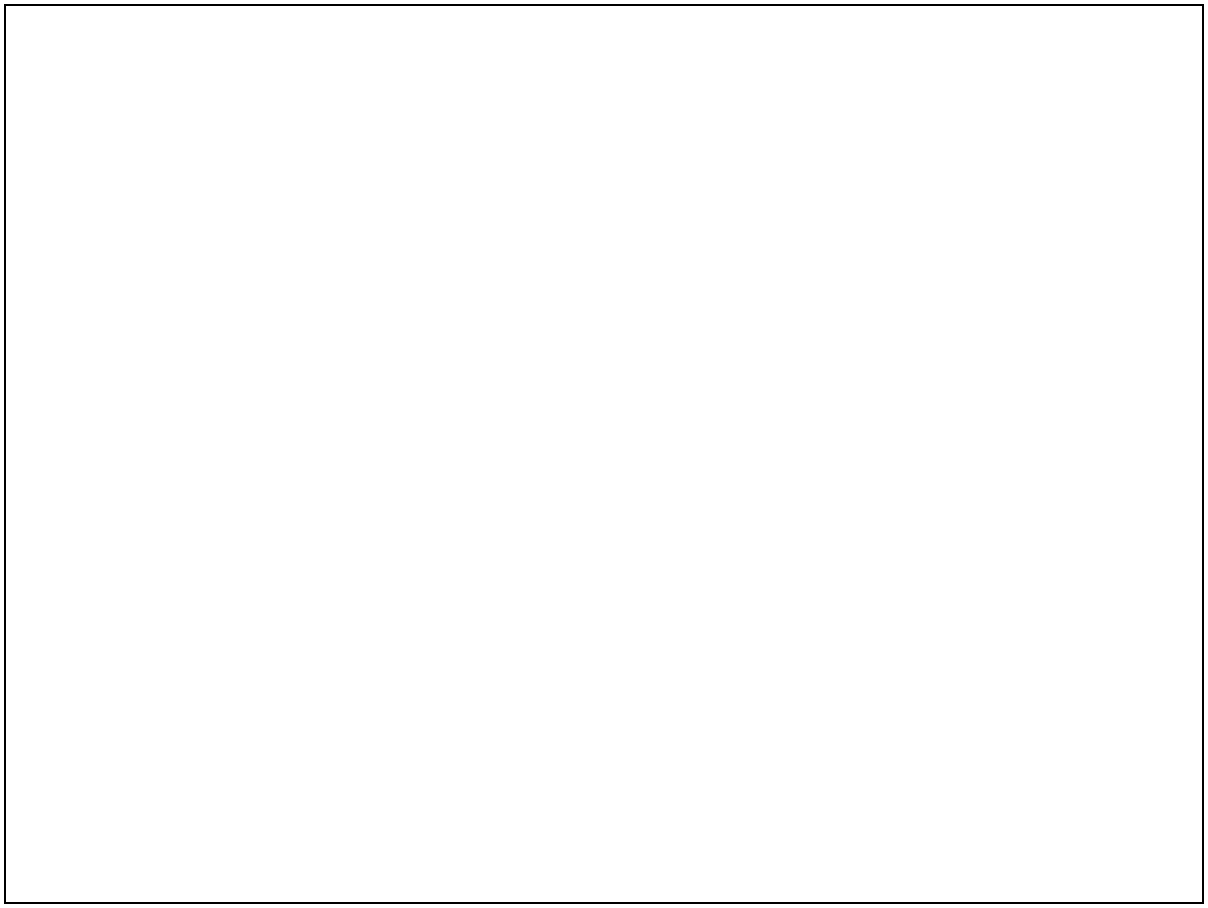
Jim Johnson and the Orioles' bullpen couldn't replicate their great 2012 season (via Keith Allison & Howell Media Solutions).

Editor's Note: This is the second post of "10 Lessons Week!" For more info, click ***here****.*

Ten years ago, I wrote an article called **The One About Win Probability** (*Friends* was a big hit on TV at the time), and it has turned out to be the most-read article in THT's history. I've written about Win Probability many times since, most notably

every year in the *Hardball Times Annual*, and I've learned a few things about WPA along the way. Allow me to share.

Lesson #1: Win Probability Graphs are Cool



Source: **FanGraphs**

The box score is perhaps the most beautiful display of data in the known universe. It contains so much terrific information in a compact space. But it lacks one thing: the dynamic of the game. Box score information is static, but games happen in real time.

Exhibiting runs scored by inning helps capture the dynamic of individual innings, and Game-Winning RBIs helped capture the dynamic of individual hits (which, I think, was the primary appeal of that otherwise irrelevant stat). But WPA graphs are much better at capturing the dynamics of the total game. They are a natural accompaniment to the box score.

The above graph is from a game played on April 23 this year, when the Angels had a 4-1 lead on the Nats in the ninth inning. Their biggest hit had been **Albert Pujols**' double in the top of the sixth. But thanks primarily to **Jayson Werth**'s double and then **Adam LaRoche**'s single, the Nats staged a statistically improbable comeback

in the ninth to win the game, 5-4.

Everything I just told you came from looking at the graph. If you hold your mouse over any point in the game, you'll see much more … exactly what happened on each play and how much it impacted the game. If you had perused just the box score, you might have missed what made this game special.

Lesson #2: Win Probability is the Story Stat

Win Probability is the ultimate quantification of a game. It captures the specifics of each situation—the score, inning, base situation and number of outs—and reduces them all into one number, the likelihood that one team or the other is going to win the game.

It's important to remember that Win Probability isn't a forecast. It's a measure by which to judge the outcome of each play and the game. WP assumes that every play after the one in question will be average. Each player will be average, and each team will be average. In that way, we don't prejudge a play's impact.

When you look at the change in Win Probability from one play to another—which we call Win Probability Added, or WPA—you quantify how important that particular play was in the story of the game. This accounts for the swings on the graph.

If you sum up a player's WPA score in a single game, you get a sense of how involved he was in the key plays of the game. And if you want to rank games by sheer excitement, you simply have to add up the total number of absolute swings in Win Probability throughout the game.

For the past two years, I've published **a list of the most watchable games of the season** by combining WPA and LI (read more about LI below) into a ranking system. WPA makes it easy to decide how to spend your time watching MLB.tv in the offseason.

Shane Tourtellotte has developed **his own take on the system**, called Win Percentage Sum. Shane uses WPA to emphasize the top plays as well as the last play of the game. There are a lot of ways you can go at the task. They key thing is that WPA gives us a tool to measure the excitement, to measure the story, of the game.

Lesson #3: Win Probability is Just an Extension of the Run

Expectancy Chart

Most baseball analysts have no problem using run expectancy charts to measure run contributions, but many seem to have a problem with the Win Probability approach. Let's discuss.

More from The Hardball Times

A Hardball Times Update
by RJ McDaniel
Goodbye for now.

I recently **wrote about wOBA** and how linear weights have become the fundamental methodology behind many, if not most, of our most advanced stats. As you may know, linear weights are an extension of the run expectancy chart, as Tom Ruane **most memorably recorded at Retrosheet**.

For reference, I'm putting Tom's run expectancy chart for the 1992 American League here. These are the number of runs the average 1992 team scored in an inning after reaching each specific situation:

Men OnNumber of Outs

Tom Ruane's 1992 American League Run Expectancy Chart			
FST	*0*	*1*	*2*
—	0.482	0.258	0.096
x–	0.853	0.51	0.211
-x-	1.095	0.646	0.293
xx-	1.494	0.907	0.423
–x	1.356	0.94	0.377
x-x	1.804	1.151	0.47
-xx	2.169	1.418	0.598
xxx	2.429	1.549	0.745

There are two elements to the run expectancy chart, the runners on base and the number of outs that have been recorded. You can think of the number of outs as an

inning timer—when there are two outs in an inning, time is running out and the impact of events is heightened.

For instance, a single that scores a runner from second is worth 0.758 runs with none out (1.000 plus 0.853 minus 1.095) and 0.918 runs with two outs (1.000 plus 0.211 minus 0.293). The fact that time is running out in the inning makes the event worth 21 percent more.

Win Probability tables add two more elements to this logic, runs scored in the game (or the difference between the two ... the Run Differential) and the inning (the game timer). In fact, it's relatively easy to construct your own Win Probability tables from Run Expectancy data. I did so in an Excel spreadsheet **that I published eight years ago.** (Man, I feel old.)

A single that scores a runner on second with two out in the bottom of the first of a tie game is worth 0.09 WPA points, a ninth of a difference between a win and a loss. The same hit in the bottom of the ninth is worth 0.39 WPA, an increase of 333 percent over the first-inning hit. When you talk about WPA, the game timer has a big impact.

A lot of people are uncomfortable with this aspect of WPA. A 333-percent jump is just too big a number to be reasonable; it takes more than four hits in the first inning to equal the impact of the same type of hit in the ninth. Even though the logic is similar to that of Run Expectancy, the Win Probability impact just feels too extreme.

I understand why people feel this way. Because of the extreme nature of WPA and the fact that opportunities aren't evenly divided among players, WPA is not the best stat for rating players. However, it seems to me that ranking games by WPA, or using it to quantify key plays or even players within a game, is a legitimate use of the stat. If you remember that WPA measures the story and not the value, you'll be on solid ground.

Lesson #4: The Difference in Impact Based on Game Run Differential

The most common complaint about WPA is that, once the game is over, it doesn't matter when the run scored. A single run scored in the first inning of a 1-0 game yields the same result as a single run in the ninth inning of a 1-0 game.

Okay, but what about the difference between a 1-0 game and a 10-0 game? Are you willing to say that a home run in a 1-0 game is worth more than a home run in a 10-0 game? After all, WPA captures this difference, too. Perhaps you want to capture this situational aspect of a player's contribution, but you feel that WPA has too much other baggage. If so, I have a solution for you.

I first introduced this concept in an article I called **Long Live Baseball Analysis**. As you might expect, WPA shows that events in close games have a bigger impact than those in not-so-close games. Here is a table of how the impact of an event varies by the final score of the game. The data are from 2006, but other years provide similar results.)

Relative WPA Impact of Events by Final Margin of Game	
Margin	*Impact*
1	1.38
2	1.13
3	0.97
4	0.86
5	0.76
6	0.66
7	0.63
8	0.57
9	0.51
10	0.47

The numbers aren't as round as I would like, but you can see that a home run in a 1-0 game (or any event in a 1-0 game) is worth nearly three times as much as a home run in a 10-0 game (or any event in a 10-0 game). The difference of 1.38 to 0.47 is a relative ratio you can use to judge such a thing.

I've wanted to invent a stat—let's call it Game-Adjusted wRC—that takes this into account. I haven't had the time, but the lesson I've learned is that there is something here worth capturing. A stat that captures the average impact of a player's hits in both runs scored and games won can be a viable and useful way of saying that Player A "contributed" more than Player B. And there is a way to develop such a stat that avoids the extreme impacts of WPA.

Lesson #5: It Defines Critical Situations

Perhaps the most useful thing to come out of Win Probability is Leverage Index. LI is a scale developed by Tom Tango that quantifies the importance of a situation based on the range of potential outcomes from that situation.

Tom introduced his concept in a three-part series here at THT:

- **Part One**
- **Part Two**
- **Part Three**

The key to Leverage Index is to measure the range of potential outcomes from any single plate appearance. The wider the range of potential outcomes, the more critical the plate appearance is. It's a straightforward concept, but the math is tricky. Most of us rely on Tom's numbers to look up the Leverage Index of a situation.

However, I can tell you a shortcut. You can mimic the relative scale of Leverage Index by calculating the difference in WPA between the current situation and the WPA that would result after a strikeout. The only exception is with a runner on third and fewer than two outs, for some fairly obvious reasons (it's the same reason we have the sacrifice fly).

I sometimes hear people say that WPA doesn't "work," that Leverage Index, for instance, doesn't really measure the most critical situations. I find this hard to believe, so I decided to pull together some data points from the 2012 season. Consider the following…

- If a team got an out in a situation with the LI between 1.5 and 2.0, it eventually won the game 60 percent of the time.
- If a team got an out with the LI between 2 and 2.5, it won the game 62 percent of the time.
- If it got an out with the LI between 2.5 and 3, it won 68 percent of the time.
- If it got an out with the LI between 3 and 4, it won 69 percent of the time.
- If it got an out with the LI between 4 and 5, it won 71 percent of the time.
- If it got an out with the LI over five, it won 76 percent of the time.

The more critical the situation, the more impact its outcome had on the outcome of the game. I don't know how to make the case more strongly: Leverage Index does a very good job of measuring the criticality of a situation. There may be better stats for this purpose, but I haven't seen one yet.

One more thing: Leverage Index is a very good tool for determining optimal bullpen usage. In general, you want your best pitcher on the mound when the game is most critical. Now, LI doesn't measure everything related to bullpen usage. The strength of the opposing batters (and of those coming later) is key, as is the lefty/righty match-up and the status of the pitcher. (Is he rested? Has he warmed up sufficiently?) But it's a natural starting place, a foundation for a decision.

Lesson #6: It's Great for Rating Bullpens

A natural use of WPA and Leverage Index, and probably the most accepted use of WPA in general, is its relevance to the bullpen. WPA, with its heightened focus on close games and late innings, is a valid way to measure the contribution of a reliever. We may shy away from ranking batters based on WPA, and there is a definite difference between starting pitchers and relief pitchers that negates direct WPA comparisons. But WPA gets to the nub of the reliever's job: perform well in high-leverage situations.

The 2012 Orioles bullpen was the best of all time, as measured by WPA. Their relievers' total WPA was 13.9, far ahead of the No. 2 team, the Detroit Tigers of 1984 (which featured AL MVP **Willie Hernandez**). I wrote about the Orioles' bullpen in detail in the 2013 *Annual*, and every aspect of its performance was tremendous. Baltimore's relief men stood out in most of the ways a bullpen can rack up high WPA points:

- The Orioles' relievers pitched more innings than most other teams' in 2012 (their innings pitched was the fourth-highest figure in the majors).
- They pitched very well in general. Their WPA/LI (more about WPA/LI below) was second-highest in the majors.
- Their overall Leverage Index was higher than average, though not outrageously so.
- Most importantly, they "pitched to the situation." They were at their best when Leverage Index was highest.

It's funny how things change. In 2013, the Orioles' bullpen WPA was 0.42, 21st in the major leagues. Why? Well, take a look at their performance in their 10 most critical appearances of the year. In seven of the 10 appearances, the Oriole reliever posted a negative WPA total.

Top 10 Most Critical 2013 Orioles Bullpen Appearances

Top 10 Most Critical 2013 Orioles Bullpen Appearances			
Game Date	*Pitcher*	*MaxOflli*	*WPA*
2013-07-05	Jim Johnson	9.0	-0.805
2013-05-26	Tommy Hunter	7.8	0.043
2013-05-18	Jim Johnson	7.5	-0.820
2013-05-26	Jim Johnson	7.4	-0.955
2013-08-14	Jim Johnson	7.1	-0.313
2013-04-24	Jim Johnson	6.9	-0.279
2013-08-09	Jim Johnson	6.4	-0.329
2013-09-20	T.J. McFarland	6.4	0.249
2013-09-29	Jim Johnson	5.8	0.077
2013-06-10	Tommy Hunter	5.5	-0.026

The Orioles had the guy they thought was their best reliever (**Jim Johnson**) on the mound in most of their high-leverage situations last year. Unfortunately, Johnson was not as brilliant as he had been in 2012. Truth be told, most of the bullpen wasn't.

By the way, WPA provides ways to measure a reliever's performance and contribution that are far better than saves and holds. In addition to WPA and LI, FanGraphs carries the number of each pitcher's shutdowns (any time a pitcher adds at least .06 WPA points to his team) and meltdowns (any time a pitcher loses at least .06 WPA points for his team). I'm looking forward to the day when we quote shutdowns instead of saves.

Lesson #7: It Identifies Clutch Performances

Does clutch hitting exist? Well, that is a question sure to make any sabermetrician spin in his computer desk chair. But at least WPA gives us a way to measure clutch performance. By comparing a player's WPA/LI in general to WPA/LI for each specific play, we can measure how well he batted in high-leverage situations compared to how well we would have expected him to bat.

The Clutchiest batters of the past 10 years have been:

Clutch Leaders, 2004-2013	
Batter	*Clutch*
Willie Bloomquist	6.2
Yadier Molina	5.6

Clutch Leaders, 2004-2013	
Jimmy Rollins	5.2
Marcus Giles	4.7
Ryan Howard	4.5

Yeah, **Willie Bloomquist** has risen to the occasion the most in the past 10 years. Now, this doesn't mean that Bloomquist is the No. 1 guy you want at the plate in a clutch situation. Even clutch Bloomquist is worse than a lot of other batters. You would probably opt for **Ryan Howard**, **Jimmy Rollins** or just about anybody from this list. It does mean that, somehow, Bloomquist has managed to exceed expectations when the game is in critical condition.

Said differently, Bloomquist has managed to positively insert himself into the game story more often than you'd expect based on his nominal stats. He may be overhyped in some other ways, but in this way Bloomquist deserves our respect.

Lesson #8: You Can Apply WPA Logic to a Season

Having called Win Probability a story stat, I should say that it is really more of a method (methodology?) than a stat. What's more, it's a method that you can apply to other facets of a baseball seasson.

For instance, if we say that plays towards the end of an inning are more important than those at the beginning, and innings at the end of a close game are more important than those at the beginning, how about late games in a season for a contending team?

Independent of each other, Sky Andrecheck and I asked ourselves this very question a few years ago. I developed a system called Drama Index, and he called his system Championship Leverage Index, but they amount to the same thing.

We both looked at each team's probability of making the postseason at each point of the season based on the standings before each game, as well as the number of games left on each team's schedule. We used these two figures to calculate a Season "Leverage Index," in which each game is rated in terms of its criticality.

There probably has been no better season for Championship LI in recent memory than 2011, when four critical games were played on the last day of the season … and three of those games were WPA humdingers.

Tampa Bay played in one of those games. The Rays had been out of the pennant race most of the year, with a CLI between 1.0 and 2.0 for most of the season until they seemingly dropped out of pennant contention in August. They managed to scramble back into things in September, however, and posted a CLI of 9.7 on the last day. That game was nearly 10 times more critical than an average major league game.

Tampa Bay wound up with 192 CLI points in 2011, the third-highest total in the American League. The Rangers (who made the postseason) and the Angels (who didn't) were the only two teams with higher scores.

Calculating Championship Leverage Index is a challenge, and MLB's recent changes to Wild Card qualification has made our old code obsolete. As a result, we haven't published any results in a couple of years. It would be fun to make this available again.

Lesson #9: You Can Apply WPA Logic Even to the Postseason

I have a lot of fun applying WPA logic to the postseason. It seems to me that the postseason is more about the story than the stats, which makes WPA a natural fit. In fact, one of the early uses of WPA was an article by Jay Bennett, who used it to make a case that **Joe Jackson didn't throw the 1919 World Series**.

What's more, the postseason offers a pretty easy way to describe how the WPA method works. It all comes down to the range of possible outcomes of a game (i.e. a win or loss). When two teams play a seventh game of a World Series, there is one full world championship at stake. Think of the possible outcomes: One team wins and is 1-0 in championships; the other team loses and is 0-1 in championships. One minus zero is one, so we give the seventh game a criticality value of one.

When two teams play the sixth game of a World Series, there are two possible outcomes. The team that is ahead in the series might win, which would give it a championship. Or it could lose, which would result in a seventh game … which the team has a 50 percent chance of winning. The difference is one minus 0.5, or 0.5.

Conversely, the trailing team could win, which would give it a 50 percent probability of winning the seventh game, or could lose and lose the championship overall. The difference is 0.5 minus one, or 0.5, same as the leading team.

The sixth game is half as critical as the seventh.

In all cases, a postseason game is equally critical to the leading and trailing team.

You can use this approach for all games in a series. You can even apply it to previous series. For instance, the final game of a League Championship Series will be worth 0.5 world championships, because the winner has a 50 percent probability of winning the World Series while the loser gets zero world championships. So you see that the final game of a league championships series is as critical as the sixth game of the World Series.

For your reference, here's a table of the Championship Leverage Index for all potential games in the postseason.

Championship Leverage Index of each Postseason Game										
Series	*0-0*	*1-0*	*1-1*	*2-0*	*2-1*	*2-2*	*3-0*	*3-1*	*3-2*	*3-3*
Wild Card	0.125									
Divison Series	0.094	0.094	0.125	0.063	0.125	0.250				
Championship Series	0.156	0.156	0.188	0.125	0.188	0.250	0.063	0.125	0.25	0.500
World Series	0.313	0.313	0.375	0.250	0.375	0.500	0.125	0.250	0.500	1.000

Since we're dealing with just two results here, win or lose (as opposed to WPA, which has to deal with home runs, stolen bases, outs, etc.), it's fairly straightforward to show the relationship between the Championship Added and the Championship Leverage Index. In fact, they're (almost) the same thing.

At the beginning of a World Series, we assume that each team has a 50 percent chance of winning the Series. If a team wins the first game, it has a 65.6 percent chance of winning three of the next six. But if it loses, it has a 34.4 percent chance of winning four of the next six games. This is based on simple probability.

The difference—.656 and .344—is what's at stake in the first game. .656 minus .344 is .313 (okay, I'm rounding here), which is the championship value of game one. The Championship Added of winning that first game is .656 minus .500, or half of .313: .156 (rounding!). In every case, the Championship Added of winning a game is exactly half of its Championship Leverage Index.

When you multiply each game's championship value by each play's WPA, you find some really interesting things. For instance, **Hal Smith**, not **Bill Mazeroski**, swatted the biggest home run in Pittsburgh Pirate history; for that matter, in major league history. And **Willie McCovey** faced **Ralph Terry** in the most critical at-bat

of all time.

There is almost no end to the things you can research and write about using WPA and its offshoots. Baseball history is rich, and there are many stories to tell.

Lesson #10: WPA/LI Might Be the Ultimate Stat

Think about it for a second. The value of winning a postseason game is always exactly half of its Championship Leverage Index. So when you divide one by the other, you get an equal value for all wins. Dividing the "Added" part of the system by its LI averages things out and makes each win equal in importance.

This is how you make a run in the first inning equal to a run in the final inning of a close game: divide by Leverage Index. Leverage Index is the great equalizer in the WPA fabric.

This is kind of esoteric, but when you divide the WPA of a specific play by its Leverage Index, you get a number that indicates who "won" the at-bat, and by how much. WPA/LI, which is sometimes called situational wins, is a funky stat. It is impossible to describe and its units are unclear. Let's try to get at this with an example.

In 2013, **Mike Trout** led the majors in WPA/LI despite being ninth in WPA. According to WPA, his biggest hit of the season was a run-scoring single in the bottom of the eighth of a tie game. Given what we know about WPA, this makes sense, right? But WPA/LI says that his biggest hit of the year was a home run in the top of the first with two out. This also makes sense. Why, you ask?

Well, first of all, home runs are the biggest hits of all, so a home run should be at the top of most "best hit" lists. Secondly, Trout hit his home run at a most opportune time because there were two out and no one was on base in a tie game. Isn't a home run the best "response" to that situation? When you lead off an inning, a single, walk, double, etc. are all reasonable alternatives to a home run. But with two out and no one on, a home run is far and away better than all other alternatives. WPA/LI measures how well the player "won" the situation.

I don't know how to better describe what WPA/LI does, but I hope you can see how useful it can be. A few years ago, I used it to **assess the impact** of all those unintentional walks to **Barry Bonds**. I have a feeling that I have just scratched the

surface of its usefulness.

So WPA/LI has a lot to recommend it. It evens out opportunities between all players so that no one benefits from having more high-leverage situations. It quantifies how well each player performed within the context of the situation. WPA/LI might just prove to be the “ultimate stat.”

Try to keep an open mind.

References and Resources

Tango’s blog and website have many excellent sources for learning more about WPA, Leverage Index and WPA/LI. Here’s **a link to his website** and here’s **a link to the run/win expectancy posts** on his old blog.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Talking Situational Wins

by Dave Studeman

June 3, 2014

Are there situations in which stolen bases are more valuable? (via Keith Allison)

I am totally fascinated by WPA/LI, even though I can't really tell you what it is. The title says what it is: **Win Probability Added** divided by **Leverage Index**, but that doesn't really help. We are probably better off calling it Situational Wins, which is a less geeky name for a very geeky concept. I tried my best to describe Situational Wins **in this article (see the end)**, in which I called it "a number that indicates who 'won' the at-bat, and by how much." Hence, the name. Situational Wins.

A vague explanation isn't the only issue with Situational Wins. Here are some others.

- Some people just don't like WPA, and using WPA/LI (which corrects some of the things people don't like about WPA) feels like jumping further into the rabbit hole.
- Total WPA/LI for a team doesn't equal the team's won/loss record. WPA does (which is one of WPA's main attractions).
- Because of rounding and the infrequency of some situations, you sometimes get results that are a bit off. I'll show you a couple of examples below, but these small differences shouldn't make much difference for players over the course of a season. Still, they exist.
- WPA/LI is calculated by comparing two tables that Tangotiger has derived: WPA and LI by situation. We all trust Tango, but are we certain his tables are completely correct?
- WPA/LI has **a bias in favor of home runs**, because not all baseball events are randomly distributed across situations (personally, I don't think this is a bias; it's just a fact. But I thought I'd mention it.)

Having said all that, my intuition is that WPA/LI works. To me the proof is in the postseason scenario (as described **in my previous article**) in which the added "championship value" of winning a game has a constant relationship to its leverage index. As a result, when you divide each game's championship outcome by its leverage index, you get the same number. Each game is equal in importance. Really, **read the article**.

Situational Wins do the same thing to plate appearances. They ensure that each plate appearance is treated the same as every other plate appearance regardless of how important it is to the game. The result is a measure of how successfully the batter or pitcher approached the situation.

This is not just an exercise in baseball math, by the way. Situational Wins are useful; they do something few other stats do. To quote Tango:

> "The key point of Situational Wins is best described in this extreme situation: with the bases loaded, tie game, bottom of the 9th, Situational Wins (and WPA) are the ONLY metrics around that will give equal contributions to the walk as it does to the homerun.

So what's stopping us from using this useful stat? To me, there is a bottom line issue with Situational Wins. When you neutralize the criticality of a situation (divide by Leverage Index), what's left? When you take out the critical impact of the score,

inning and base/out situation, what is left for WPA/LI to consider? What are the key elements of the "situation" in Situational Wins? And how are they calculated?

...

Today, I'd like to experiment with Situational Wins to see if we can get a better handle on this issue. I plan to calculate the WPA/LI of several different situations and outcomes just to see what we see. I will discuss the results and encourage you to add your own comments. By working together, perhaps we can come to a better understanding of what this elusive stat is.

I'm going to use the WPA Inquirer for my results, which you can find below (you can also find it at **our WPA Inquirer page**). You're welcome to play with the WPA Inquirer, add your own scenarios to the conversation and leave them in the comments.

In all cases, I will assume a run environment of 4.5 runs per game. All values will be expressed from the batting team's perspective but we will also try to incorporate the pitching team's perspective.

First up: What happens when the score changes?

Value of Leadoff Event in Bottom of Seventh				
Home Team Score Diff	*Out*	*Single*	*Double*	*Home Run*
Down by Two	-0.025	0.044	0.065	0.094
Down by One	-0.026	0.040	0.069	0.119

Value of Leadoff Event in Bottom of Seventh				
Tie Score	-0.025	0.037	0.072	0.135
Up by One	-0.025	0.038	0.071	0.131
Up by Two	-0.026	0.038	0.072	0.134

An out has the same negative impact regardless of the score. How can this be? Doesn't an out hurt the team more when you're behind than when you're ahead? Yes it does, but that is what WPA captures. Situational Wins count a bases-empty out as the same value, regardless of the score, inning or number of outs. Bases-empty outs are the constant in the Situational Win universe.

Try it yourself. Plug a bases-empty out into the Inquirer in lots of different situations to see what you get. Remember that one- and two-point differences are insignificant and that rare opportunities are affected by rounding in the tables. As an example, try a seven-run lead in the bottom of the eighth. The Leverage Index is so low in that situation (0.01) that you know it's rounded. The WPA/LI value won't line up as well.

Anyway, does this make sense? Should bases-empty outs have the same value regardless of the situation? Well, when a pitcher gets a bases-empty out from a batter, he has won the contest and **moved the game clock forward** by the same amount each time. Situational Wins should be consistent here, and the fact that WPA/LI works this way is a confirmation of the system. Things are different when there are runners on base; I'll explore that in a minute.

...

There's another wrinkle here: a single is valued more highly when a team is down, but a double and home run are valued more highly when a team is tied or ahead. Playing with the WPA Inquirer, we find that this effect is much less pronounced in earlier innings and more pronounced in later innings. So game time is a factor here, but the trend is the same across all innings. What's up with that?

Remember that we're talking about leadoff situations in this example. Hitting a single at the beginning of an inning when you're down by two is better than when you're ahead because you need to score runs. By hitting a single, you keep the inning going and set up the potential for more runs.

On the other hand (from the pitcher's perspective), the output is saying that giving

up a solo, leadoff home run is more harmful when you're behind and trying to catch up than when you're ahead and trying to maintain a lead. The same is true for a double, but to a much lesser extent.

I've thought about this a lot, but I haven't been able to find the ideas or words that tease out what is going on here. Why does the single follow a different pattern than the extra-base hit?

Next, let's first see how Situational Wins change when the number of outs change:

Value of Event with Runner on Second in the Top of the Fourth of a Tie Game				
# of Outs	*Out Without Moving Runner*	*Bunt Runner to Third*	*Run-scoring Single*	*Home Run*
Zero	-0.034	-0.011	0.058	0.107
One	-0.028	-0.024	0.069	0.127
Two	-0.028	-0.028	0.082	0.154

First of all, the negative impact of not moving the runner over to third is larger with none out than with one or two out. This makes sense, of course, because a runner on third can score on a sacrifice fly with one out but not with two out. Conversely, bunting the runner to third with none out (which is a negative play nonetheless) is a much better play than doing the same thing with one or two outs. Once you think about the sacrifice fly, this will make perfect sense to you.

On the other hand, the value of a positive batting event, such as singling in the runner or hitting a home run, increases as the number of outs increases. This makes sense, too, because teams have more time to score runners when the inning is still young. When the inning is running out, however, run-scoring events are more meaningful.

Think of it from the pitcher's point of view. With a runner on second and no outs, you sort of expect that some runs are going to score. But with a runner on second and two outs, you're hoping no runs will score. So giving up the run-scoring hit hurts more–is a worse result for the situation–than earlier in the inning.

...

Finally, how do things change as the game progresses?

Value of Event with Runner on First, One Out in Tie Game in Top of...

Value of Event with Runner on First, One Out in Tie Game in Top of...				
Inning	*Stolen Base*	*Caught Stealing*	*Double Play*	*Single Runner to Third*
First	0.014	-0.035	-0.044	0.055
Third	0.014	-0.034	-0.043	0.054
Fifth	0.014	-0.034	-0.043	0.055
Seventh	0.016	-0.034	-0.044	0.056
Ninth	0.019	-0.034	-0.045	0.061

Once again, outs have the same value regardless of the inning. The new insight is that these aren't outs that occur with no one on base, but outs that finish with no one on base. So we can update our previous finding to say that all outs that finish with no runners on base have the same value. Of course, caught stealing has a bigger negative value than a bases-empty out because it eliminates a baserunner.

Now positive offensive events, such as stolen bases and singles, go up a bit in the ninth inning (and a bit less in the seventh). Similar events in Win Probability Added increase dramatically in the late innings of a game. It appears that the WPA/LI value of positive batting events also increase in the late innings of a game, but to a lesser extent.

This is another tough one for me to tease out. It's obvious that batting events have more positive value as the game clock runs down. Is that all that is going on here? If so, how is this quantified in a way that differentiates it from WPA? If not, what else is being considered?

...

I'm afraid I've posted more questions that answers today. Hopefully, this will be one of those articles in which the best insights are left in the comments. Got any?

References and Resources

Kincaid's article about WPA/LI is worth reading because it demonstrates that the average WPA/LI values of events line up well with their linear weight values.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic

tweets **@dastudes**.

What a no-hitter tells us about WPA/LI

by Dave Studeman

June 10, 2012

I posted an idea yesterday, a made-up stat just for fun. I called it **No-Hitter Added** because it was a "real time" stat that apportioned credit among Mariners' pitchers for their Friday no-hitter. **Tom Wilhelmsen** was given a lot more credit for the no-hitter than **Kevin Millwood**, because Wilhelmsen pitched the critical ninth inning when the no-hitter was most in reach … despite the fact that Millwood faced six times as many batters.

As a reminder, if you were to break out credit for the no-hitter based on batters faced (which is obviously the best way to do it), it would look like this:

```
K Millwood          60%
T Wilhelmsen        10%
S Pryor             10%
C Furbush           10%
B League             7%
L Luetge             3%
```

The No Hitter Added stat was just for fun, an exercise to show how WPA works, as well as its strengths and its flaws. kds posted a great comment on the article, noting that an approach that divides the No Hitter Added stats by the Leverage Index would be fairer. And that got my head spinning.

See, it's easy to calculate the Leverage Index of each batter faced in this situation, because the only outcomes we care about are a hit or not. For instance, with two out in the ninth, the Leverage Index is 1.0, because an out will result in a no-hitter (a 1) but a hit will result in no chance of a no-hitter (a 0). 1-0 equals one. In another example, the Leverage Index of the first batter is 0.0007, because a hit results in a 0, but an out leads to the next at-bat, at which point the team has a 0.07% probability of throwing a no-hitter (not giving up a hit to the next 29 batters). 0.0007 minus zero is … well, you get the idea.

In all cases, the "No-hitter Leverage Index" of an at-bat is the probability of throwing a complete-game no-hitter starting with the next at-bat.

So now we have No Hitter Added, No-Hitter Leverage Index, and one divided by the other (NHA/LI). Let's apportion credit among Mariner pitchers for that no-hitter using NHA/LI:

```
K Millwood          60%
T Wilhelmsen        10%
S Pryor             10%
C Furbush           10%
B League             7%
L Luetge             3%
```

Yeah, the results are exactly the same as those using batters faced. In other words, NHA/LI works.

So what, you say? Did I just waste your time? Was that just a bag of gas for nothing? Well, yes, it mostly was except for one thing. Now you know why WPA/LI works.

Most people get **WPA** and **LI**, but WPA/LI is maddeningly difficult to explain. Tangotiger has tried several times, **as have I**, but it's hard to do. People have just had to take it on faith that it works.

But No Hitter Added is a much simpler concept (and the math is much simpler) than Win Probability Added. And here you can see that dividing the "Added" stat by the Leverage Index results in the exact numbers it should. I'm not going to try to explain why it works; I've tried that before. I'm hoping you can see that the math, which works perfectly in the no-hitting example, will work just as well within the framework of winning games.

WPA/LI is available at both **Fangraphs** and **Baseball Reference**.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Barry Bonds' situational wins

by Dave Studeman

November 23, 2010

When should you walk **Albert Pujols**? **Josh Hamilton**? **Cesar Izturis**? There's no easy answer because the decision depends not only on the base/out situation, the inning and the score; it depends on the quality of the batters coming to bat next, the nature of the pitcher/batter matchup and maybe even the phase of the moon. Even Izturis was intentionally walked once last year (with two out, runners on second and third and the pitcher up next; the moon was in its first quarter.).

Tom Tango, MGL and Andy Dolphin wrote the definitive chapter on intentional walks in **The Book: Playing the Percentages in Baseball** (a must-have for all serious baseball fans) and I'm going to dive into the intentional walk question by taking one of Tango's best (and least understood) statistical concepts—Situational Wins—and applying it to the all-time leader in intentional walks, **Barry Bonds**.

Do you know what Linear Weights are? They're the weights that can be applied to events on the field in order to properly credit players for their impact on the game. A single is typically worth about .45 runs, a double .80 runs, a triple 1.0 runs and a home run 1.4 runs. There are several ways to create your linear weights, but my favorite is the "Value Added" approach, which works like this:
{exp:list_maker}Figure out the 24 basic out/base situations (three outs—0, 1 and 2—and eight base situations—bases empty, runner on first, etc.).
For each base/out situation, go through the season and figure out how many runs subsequently scored during that inning. Call that the Run Expectancy Matrix.
Go through the season again and figure out how much each batting event changed the team's place on the Run Expectancy Matrix (easy example: a single with no one on and one out results in a runner on first and one out).
Use that average change to assign a Linear Weight to each type of event.
{/exp:list_maker}The very best on-line example of this methodology is **Tom Ruane's work at Retrosheet**. If you want more detail about the method, read

Tom's article.

Linear Weights assigns an average weight to each type of batting event, depending on how often each one occurred in each type of situation. This can lead to some inefficiencies in the system. For instance, pitchers rarely walk batters with the bases loaded. Wild pitchers who do give up walks with the bases loaded, however, have hurt their teams more than their Linear Weights suggest. This is where Situational Wins comes to play.

Situational Wins goes two steps beyond Linear Weights:
{exp:list_maker}It takes the **specific** base/out situation of each event into account. This is only possible thanks to **Retrosheet's fantastic data collection efforts**. It adds the inning and score to the context because it measures impact on wins, not runs.{/exp:list_maker}
Let me give you another example to show the difference. Bases are loaded in the bottom of the ninth, tie score. If the batter makes an out, the game goes into extra innings. If the batter gets a hit, home team wins. It doesn't matter whether the batter hits a single or home run; the home team will win regardless. So, each type of hit has the exact same value to his team.

Linear Weights treats the single and the home run differently. Situational Wins doesn't.

And how do we compute Situational Wins? Easy. It's Win Probability Added (WPA) divided by Leverage Index (LI). It's so simple to calculate that both **Fangraphs** and **Baseball Reference** include it in their player "win stats," but most people probably don't understand what it is or what it measures. Now you know: It measures the contribution each batter made to his team's wins, based on contributing specific batting events in specific situations.

The reverse is also true for pitchers, by the way. I've just decided to describe the system from the batter's perspective because it fits my topic of the day. If you want to read more detail about Situational Wins, try **this article from Tom Tango**. If you want even more detail, **read this one**.

I pulled the event files from 2001-2004 and calculated the average Situational Win of each type of batting event. These were my findings:

Event	SW

Single	0.040
Double	0.068
Triple	0.095
Home Run	0.142
Walk	0.030
Intentional Walk	0.006
HBP	0.030
K	-0.027
Ground out	-0.025
Fly Out	-0.026
DP	-0.054

Notice anything familiar about these numbers? They're roughly equal to 1/10th of each event's Linear Weight ... which makes sense when you consider that it takes about 10 runs to turn a loss into a win. (By the way, **here is a mathematical proof**, based on the Pythagorean Formula, that a win must equal two times the number of average number of runs scored per game and team.)

Anyway, this is an indication that Situational Wins work, that they make sense. And now I'm going to show you how we can use them.

It seems like ancient history, but Barry Bonds had four amazing years from 2001 to 2004. Here are a few key stats for each year:

Year	PA	HR	NIBB	IBB	SO	BA	OBP	SLG
2001	664	73	142	35	93	.328	.515	.863
2002	612	46	130	68	47	.370	.582	.799
2003	550	45	87	61	58	.341	.529	.749
2004	617	45	112	120	41	.362	.609	.812

73 home runs in 2001 is ridiculous, of course (and leave your comments about steroids for another article). Perhaps even more ridiculous, however, was the fact that he received more intentional walks than non-intentional walks in 2004. Did Bonds deserve such a high level of respect? Is there any way we can put this many free passes in perspective?

Situational Wins to the rescue. Let's start with Bonds' average Situational Wins per at-bat for each year in question. This is how much he contributed, on average, when the opposing pitcher didn't walk him or hit him with a pitch:

Year	Avg SW
2001	0.017
2002	0.016
2003	0.012
2004	0.013
Total	0.015

Every time he was allowed to swing away, Bonds added 0.015 wins for his team. That's a very good number. Keep in mind that the sum of all Situational Wins is zero—all the positive and negative events, all the wins and losses, cancel each other out—and that walks are a positive event for the batter. That means that the average at-bat, not including walks, is a slightly negative event. Bonds' at-bats were positive by a good margin.

These numbers give us an important baseline. The answer to "When should we walk Barry Bonds?" is simply "When the Situational Win resulting from a walk is less than his average Situational Wins in at-bats." Intentional walks are a choice, and you know that if you don't walk Barry Bonds, he will produce .015 Situational Wins, on average. If you don't want to take your chances, then you should walk him when the impact of the walk will be less than 0.015 Situational Wins.

More from The Hardball Times

A Hardball Times Update
by RJ McDaniel
Goodbye for now.

Quick aside: I'm oversimplifying things here, because when you do pitch to Barry Bonds you often wind up walking him anyway and that affects the baseline Situational Wins. I'm choosing to ignore that wrinkle for now.

Did managers follow this rule? Let's look at each year.

In **2001**, when Bonds batted third in the order and had **Jeff Kent** hitting behind him, he was intentionally walked 35 times. In 31 of those IBB's, the resulting Situational Win was less than 0.017, Bonds' at-bat baseline that year. The average for all of his intentional walks was 0.011, significantly less than 0.017. Opposing

managers were being appropriate, perhaps even conservative.

In **2002**, manager **Dusty Baker** continued to have Bonds hit third in the lineup with Kent fourth, and then flip-flopped the two for the last two months. Regardless of where he batted, opposing managers issued intentional walks to Bonds twice as often … 68 times in all. 51 of those intentional walks created Situational Wins of 0.16 or less and the overall average climbed to 0.013 WS per intentional walk. Managers weren't as conservative as they had been in 2001, but they were probably still appropriate on average.

In **2003**, Bonds primarily batted fourth and Kent's protection was replaced by several batters, such as **Edgardo Alfonzo** and **Benito Santiago**. Opposing managers continued to walk him at about the same rate, resulting in 61 intentional walks. 45 of those walks resulted in a Situation Win of 0.014 or less—an overall average of 0.011. Still appropriate, I'd say.

All hell broke loose in **2004**. Bonds got off to an otherworldly start, batting .472/.696/1.132 with 10 home runs in the first month. Managers panicked, walking him a staggering 120 times in the year with batters like Alfonso and **Pedro Feliz** batting behind him.

61 of those intentional walks resulted in a Situational Win of 0.013 or less (his average at-bat SW that year) and 69 were less than 0.015 (his average at-bat SW all four years). Just a little more than half of them were clearly appropriate; the others, not. The overall average Situational Win for Bonds' intentional walks in 2004 was 0.015, higher than his average Situational Wins in his at-bats.

Managers must have thought that walking Bonds helped their overall cause. On May 1, for example, Marlins' manager Jack McKeon intentionally walked him 4 of 5 times* and Bonds went on to bat only .250 in May. Perhaps some people saw a a connection. In fact, all the intentional walks may have had an impact on Bonds. This is the average Situational Win of Bonds' home runs each year (remembering that the overall major league average was 0.142):

```
Home Runs Only
  Year    Avg SW
  2001     0.164
  2002     0.161
  2003     0.150
```

2004	0.132

By walking Bonds in high-leverage situations, managers clearly eroded the impact of his home runs. However, the decrease in home run impact was more than offset by an increase from those intentional walks. Bonds' overall Situational Wins average rose in 2004.

All Plate Appearances

Year	Avg SW
2001	0.020
2002	0.019
2003	0.015
2004	0.017

As great as Barry Bonds was, we all lost our perspective in 2004.

By the way, these calculations are based on league-average hitters. Say what you will about the likes of Alfonso and **A.J. Pierzynski**, they weren't worse than league average at the time.

To wrap this up, here is a table of all the intentional walks given to Bonds during these four years, broken into three groups (Okay, Maybe and No Way).

WPA/LI	Number	Pct.
Less than 0.016	199	70%
0.016-0.019	21	7%
0.019 and Over	64	23%
Total	284	100%

39 of the 64 "No Way's" occurred in 2004.

**That game included the most egregious example of a bad intentional walk when McKeon had Bonds walked with one out and no one on base in the fifth inning, with a 3-2 lead. That represented a Situational Win of 0.037. The Giants didn't go on to score in that inning, but they did eventually win the game.*

References & Resources

You can calculate the Situational Win of a walk by using **our WPA Inquirer**.

This article was inspired by **Steve Treder's piece in walks to sluggers**.

Many, many thanks to **Tom Tango** and **Retrosheet**.

Tom also developed a WPA framework for when/when not to walk Barry Bonds. Here's one for **when the Giants are on the road** and here's one for **when they're at home**.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

What's a Most Valuable Player?

by Dave Studeman

September 15, 2015

There's more to Anthony Rizzo's NL MVP case than meets the eye. (via Ben Grey)

Let's say you play in a baseball simulation league, something like APBA or Strat-O-Matic, and you have a choice between two batter cards. The first batter, let's call him Joe Slugger, will put up amazing slugging stats—the best in the league. His On-Base Percentage (OBP) and Slugging Percentage (SLG) will be at the top of the charts, perhaps historically great.

The second batter, Eddie Situation, will post very good numbers but not as good as Slugger's. The difference is that he will excel in key situations. According to his card, his performance will increase with batters on base while Slugger's performance will

get worse with runners on base.

This is a very sophisticated game you play, and Situation's card also shows that he will perform best in close games, while Slugger's stats are more likely to be accrued in runaway contests. You know that a home run in a one-run contest is worth more than a home run in a five-run contest, so you're even more intrigued by Situation's card. When you run all the numbers, you find that Situation's contribution to your team's wins will be higher than Slugger's.

Which batter will you choose?

Me, I'd choose Situation instead of Slugger. I don't want to lead my league in counting stats—I want to lead my league in wins. Choosing the guy who hits better with runners in scoring position *and* in close games will result in more wins for my team. I don't want the *best* hitter on my team, I want the *most valuable* hitter.

See where I'm going with this?

Currently, the consensus seems to be that **Bryce Harper** is the clear choice for the National League Most Valuable Player Award. After all, he's batting .333/.463/.648 with 36 home runs, 84 RBI and 104 runs scored (all numbers as of Sept. 14). However, I don't think we should totally hang our hat on those counting stats. Allow me to offer another way to look at things.

Here are the current (as of Sept. 14, and that's the last time I'll say it) NL leaders in wRC (weighted Runs Created, which is a similar scale to Runs Scored and RBI), along with some other key stats:

2015 NL MVP CANDIDATES

Name	Team	PA	AVG	OBP	SLG	wRC	wRAA
Bryce Harper	Nationals	574	.333	.463	.648	133	68.3
Joey Votto	Reds	616	.315	.459	.555	128	58.2
Paul Goldschmidt	Diamondbacks	620	.316	.431	.553	118	48.0
Anthony Rizzo	Cubs	616	.278	.388	.523	107	37.2
Andrew McCutchen	Pirates	601	.299	.401	.509	103	35.4

As you can see, Harper leads the league in wRC, five ahead of Joey Votto and 15 more than Paul Goldschmidt. He also leads in wRAA, which is the same thing as

wRC, but it's expressed in runs above average instead of total runs. You can see Harper moves ahead in wRAA by posting the highest wRC total but in fewer plate appearances. Votto and Anthony Rizzo, who have the exact same number of plate appearances, also have the exact same difference in both wRC and wRAA. I like to look at both stats as a starting point.

Not all hits have the same impact, however. A single with a runner on third and two outs is worth more than a single with no one on base and two outs, for just one example. The best research I've seen that captures this phenomenon was posted by **Tom Ruane at Retrosheet many years ago**. Tom calls it Value Added Batting Runs, but it's called RE24 at FanGraphs and Baseball-Reference. Shorter names fit better on teeny columns.

You calculate each batter's Value Added Runs—or RE24—by calculating the difference in expected runs from before his plate appearance to after his plate appearance, adding in the number of runs that scored. You can calculate the difference in expected runs by…well, just read the article.

RE24 is a real gem of a stat and vastly underused. It tells you very different things about batters. Let's look at the RE24 of our top five MVP contenders and compare it to their wRAA.

2015 NL MVP CANDIDATES, WRAA & RE24 COMPARISON

Name	Team	wRAA	RE24	Diff
Bryce Harper	Nationals	68.3	68.06	-0.24
Joey Votto	Reds	58.2	64.35	6.15
Paul Goldschmidt	Diamondbacks	48.0	51.41	3.41
Anthony Rizzo	Cubs	37.2	55.47	18.27
Andrew McCutchen	Pirates	35.4	47.69	12.29

Harper's RE24 is about the same as his wRAA (too many acronyms!), but every other batter increases his impact once we add in his performance in each specific situation. How does this happen? Well, here are a couple of factoids (out of many possible factoids) to consider:

- With runners in scoring position, Bryce Harper has batted .294/.476/.559. With no one on, he's batted .336/.443/.695.
- With runners in scoring position, Anthony Rizzo has batted .305/.425/.602.

With no one on, he's batted .249/.344/.488.

With runners in scoring position, Harper has actually been worse than with no one on, but Rizzo has been much, much better than his no-one-on stats. The nice thing about RE24 is that it takes all the differences in base/out situations (there are 24 of them) and sums them up in a single number.

This is called situational hitting. People tend to ignore it because it's not very likely Rizzo will repeat this breakout again. But MVP awards aren't given for repeatable performances. They're given based on what actually happened. Rizzo's RISP performance actually happened.

You probably know all of this; maybe you've looked up RE24 on baseball websites many times before. But there's another wrinkle to consider, something **I first researched in 2007**. That is, the margin of victory in a game.

This is a simple yet powerful idea, too: runs in close contests are more important than runs in blowouts. The object of the game is to win the thing, not to run up a big run total. Thanks to WPA, we can quantify exactly how much events matter to winning in close games vs. those in blowouts. For instance, a batting event in a one-run game was worth 1.38 more than average, while a batting event in a three-run game was worth 0.97 of average. I'm going to call this a Margin Factor (and that's all the detail I'm giving you here; read the article for more).

So I multiplied each batter's RE24 in each game by the Margin Factor of the game. Below you can see how our Big Five rank in wRAA, RE24 and this last stat, which I'll call Game-Adjusted RE24 (GameRE24 for short).

2015 NL MVP CANDIDATES, GAME-ADJUSTED RE24

Name	Team	wRAA	RE24	GameRE24
Bryce Harper	Nationals	68.3	68.06	56.4
Joey Votto	Reds	58.2	64.35	60.6
Paul Goldschmidt	Diamondbacks	48.0	51.41	55.5
Anthony Rizzo	Cubs	37.2	55.47	54.0
Andrew McCutchen	Pirates	35.4	47.69	43.4

Most of our batters lost ground in the transition from RE24 to GameRE24 because one-run games tend to be low-scoring affairs. However, Paul Goldschmidt actually

increased his total impact by factoring in the game situation. (Fun fact: Goldschmidt leads the majors in RE24 in one-run games at 25.4. **Kris Bryant** is second at 19.1).

In our overall GameRE24 totals, Votto is now in the lead, with Harper, Goldschmidt and Rizzo virtually tied for second. It's not a runaway race anymore; it's a dead heat.

Of course, you probably would want to plug these numbers into WAR, where Harper's (and McCutchen's) positions would factor in. So, too, would Goldschmidt's fielding numbers. You probably also should include the fact that Harper has accrued fewer plate appearances. There is much still to play with, but the basic concept is fully formed.

In 1979, **Don Baylor** won the American League MVP Award, largely thanks to an impressive RBI total of 139. Back then, we didn't have any advanced stats or breakouts to make sense of those numbers, so we grabbed onto some evidence of situational hitting wherever we found it.

Here's what we now know. Baylor came to bat with runners in scoring position 258 times in 1979, 22 more times than the major league runner-up (**Darrell Porter**). His OPS in those situations was .981, which is obviously very good but was still just 15th in the majors. His RBI count was driven more by opportunity than performance.

If we had been able to see his RE24 at the time, we would have seen a total of 37.9, **eighth in the AL** behind **Fred Lynn's** 60.6. In fact, Lynn led the majors in OPS with runners in scoring position at 1.188. Remarkably, his RE24 was much higher than Baylor's despite coming to bat only 166 times with runners in scoring position.

This debate of "What does valuable mean?" has gone on long enough. Many people have given up on the topic, ceding the floor to simple counting stats because, you know, who really knows? But we do know. Value = Winning. Winning is a function of hitting, hitting in the right situations and in the closest games. If you want to honor the best hitter, vote for the Silver Slugger. But if you really want to honor the Most Valuable Player, take a good look at his Game-Adjusted RE24. That is where value lies.

References & Resources (and Caveats)

- As always, thanks to Retrosheet, FanGraphs and Baseball-Reference for

their wonderful work and tools. Particular thanks to Tom Ruane for his research. Tangotiger has been my guide in much of my WPA research over the years.

- I used the 2007 Margin Factors from my original articles. Ideally, I'd update those for the current run environment. Also, the RE24 stats at FanGraphs and Baseball Reference are based on evolving run expectancy tables and won't be finalized until the season is over. I'm taking it on faith that these RE24 numbers are pretty close to the final deal.
- Tom Ruane, Retrosheet, **"The Value Added Approach to Evaluating Performance"**
- Dave Studeman, The Hardball Times, **"Long Live Baseball Analysis"**

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Season leverage index

by Dave Studeman

December 11, 2008

Well, I don't know about you, but I found the National League MVP voting truly enlightening, in a negative-BBWAA-sort-of-way. **Like a lot of sabermetric types**, I didn't think that **Ryan Howard** was the most valuable player on the Phillies, let alone in the league. I didn't even think he was one of the 10 most valuable players in the league.

Think about that. The guy led the league in home runs and RBIs by a good margin, yet a doofus like me thinks he wasn't one of the 10 most valuable players in the league? As **Craig Wright has written**, it really is unbelievable. This is a list of things that Craig figured a player would have to do to achieve such a feat:
{exp:list_maker}Play a big offense position like first base and play it poorly,
Have the third worst on-base percentage at his position,
Lead his position in outs made, and
Do it in a season with a lot of strong individual performances.
{/exp:list_maker}Yeah, Ryan Howard did all of those things. But that didn't matter to the BBWAA. Howard finished second in MVP voting, with 12 voters placing him first (the winner, **Albert Pujols**, had 18 first place votes).

One of the reasons Howard received so many votes was his September performance, when he batted .352/.422/.852 while the Phillies overtook first the Brewers and then the Mets to qualify for the postseason. The MVP voting is really all about the drama, and Howard provided game thrills that those poor bored sportswriters seemed to enjoy.

Now, we can complain all we want that April wins count as much as September wins, but that's not a very interesting way to look at things. It doesn't thrill your typical sportswriter and, honestly, it probably doesn't thrill your average fan. I have a feeling Howard might have done just as well if the MVP voting was conducted the same way

MLB conducts All-Star game voting: mass voting by casual and intense fans alike. Everyone likes a good stretch run.

Well, why not count late games in pennant races more than earlier games? After all, part of the fun of being a fan is seeing your favorite player or team come through "when it counts." And the heart of a fan is much more likely to be crushed in September than April. As my barber said to me the other day, "Maybe it would have been better if the Cubs had never competed at all. Why can't we just have fun again?" Yes, they're still trying to recapture their mojo on the North Side.

The entire issue got me to thinking: if we accept that September games provide more drama and have more impact than April games, can we quantify the difference? (You're probably not surprised I thought about the numbers angle.) For in-game situations, we have **Tangotiger's Leverage Index**, in which 1.0 is an average situation. When the Leverage Index is higher than one, it's more critical (2.0 or 3.0 are typically good cutoff points for really critical situations—about 3% of all plays had a Leverage Index of 3.0 or more last year). Can we derive the same sort of thing for in-season games?

Of course we can. I've put together a system that I think does the trick pretty well and I'm going to offer it here as a way of stimulating discussion and debate. Once we refine the method, we can apply it to Howard, Pujols and all the other MVP candidates to rank their "real-time" impact on the pennant race. Why? Because we baseball nerds like to really spoil the BBWAA party by quantifying everything they can think of.

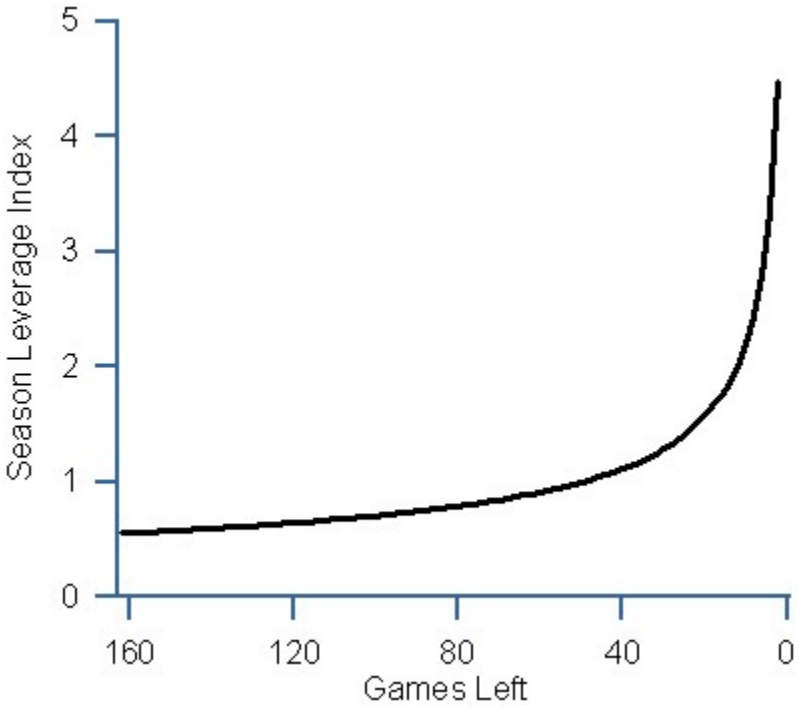
I'm going to list all the gory mathematical details at the end of the article. Suffice to say that I used a binomial distribution of how often a team is likely to gain X wins in the next Y games (assuming that all teams are naturally .500 teams) over 162 games. I then compared the difference in the binomial distribution between an incremental win and a loss at all points of the season, and indexed that to the 112th game (which, it turns out, represents the average difference).

Like I said, I'll go into more detail in a second—let me first show you the results. For this index, I assumed that a .500 team would need to play .500 ball for the rest of the season to reach the postseason (I don't think it really matters exactly what the goal is because we're indexing the results). Here are the month-by-month results:

Month	LI

April	0.6
May	0.6
Jun	0.7
Jul	0.9
Aug	1.1
Sept	2.2

So, according to this methodology, September games are almost four times more critical than April games. In fact, they're twice as critical as August games. That's a pretty steep curve, which is even more apparent when you graph the index of every other game as the season progresses:



This curve may seem too steep to you, but it makes a lot of sense. Of course, when there is only one game left in the season, the potential outcome is going to have a lot more impact than an April game, when there are 150 to 160 more games to go. The question I have is: is this how steep the BBWAA (and many fans) think of it? In fact, don't you sometimes get the feeling that MVP voters only consider second-half performances and **nothing** else? Why would **Manny Ramirez** finish fourth in MVP voting otherwise? Why would **Chase Utley** finish 15th?

So, let's go back to our original inspiration. What does this crude tool do for Ryan Howard? To get at that question, I guesstimated monthly Runs Created for Pujols and Howard by calculating each player's GPA for each month and applying **this simple math**. (By the way, GPA worked pretty well as a Runs Created estimator for

these two players. It gave Ryan Howard 108 RC vs. "actual" RC count of 113 (per Baseball Reference) and Pujols wound up with 150 RC vs. BRef's 160). I then applied the monthly Leverage Index for both players, like so:

	Original Runs Created			Leveraged Runs Created	
	Howard	Pujols	LI	Howard	Pujols
April/March	11	32	0.6	6	19
May	20	29	0.6	13	19
June	12	11	0.7	9	8
July	20	23	0.9	17	20
August	16	32	1.1	19	36
Sept/Oct	28	24	2.2	60	52
	108	150		125	153

Howard closes the gap, but Pujols performed pretty well in September, too, and when you factor in fielding prowess, Pujols gains another 20 runs on Howard. It's still a slam dunk decision.

But the case isn't closed. This is a very crude approach—it doesn't include the specific leverage of each game played by these two players. The Cardinals played fewer critical September games than the Phillies did, and the LI multiplier ought to reflect that.

Good news: I'm almost finished building that spreadsheet. But I first wanted to get some feedback about this general approach. Let me hear you.

Gory mathematical details

I'm not really a math geek; I just play one on the Internet. So take this system with a grain of salt. But to derive in-season game criticality, I used the ol' **binomial distribution** or **Bernoulli distribution**. A binomial distribution (which is a simple function in Excel called BINOMDIST) is a distribution of possible outcomes, given a certain number of trials and an underlying assumed probability. It works for any "binomial" outcome, such as wins/losses, heads/tails, yes/no, up/down, on/off.

More from The Hardball Times

A Hardball Times Update
by RJ McDaniel

Goodbye for now.

For instance, if you have four games remaining and a 50% chance to win each one, the binomial distribution predicts this range of outcomes:
{exp:list_maker}0 wins: 6.25% of the time
1 win: 25% of the time
2 wins: 37.5% of the time
3 wins: 25% of the time
4 wins: 6.25% of the time {/exp:list_maker}Nice symmetrical output, isn't it? If you multiply the probabilities times the number of wins, you get an average of two wins—as you should. Note that if you add these probabilities, you find that a team will win at least two games 68.75% of the time. In other words, if it has to play at least .500 ball the last four games to make the postseason, there is nearly a 70% chance that it will do so. This may run counter to your "natural" assumption that this team has only a 50% chance of making it in that situation.* That's the power of the distribution output.

**I'm ignoring competition here. Once you factor in a competitor, the odds change.*

So, to derive the index, I created a binomial distribution table for 162 games and all possible outcomes (keeping the underlying probability at 50%). I then took each "cell" in the table and compared the difference between a loss and a win at that point. In other words, I looked at the binomial distribution in the next column and took the difference between the team being a win under .500 and a win over .500 (assuming that the goal is to finish the season at .500). As the season progresses, this difference increases from .0626 in the first game to .5000 in the next-to-last game.*

**I only used even-numbered games to derive the index, because there were no ".5" wins in the table.*

Then I divided those indices by the overall average, which is .1123. And that's how I calculated in-season game leverage.

When I apply this math to specific players, I will look at the postseason requirements of each specific team at the time of each game. I'll calculate how far behind or ahead

they are in the postseason race (all the wild card permutations make my head spin), and use that as the postseason goal instead of .500. To simplify the math, I'll assume that all teams are .500 teams. I know this isn't reality, but we're not trying to predict pennant races; we're just developing an index.

If you have questions or comments, please leave them at Ballhype. In the meantime, I'll get back to my spreadsheet.

References & Resources

Ken Ross' **A Mathematician at the Ballpark** is a great source of information about Bernoulli trials and binomial distributions. Of course, Tangotiger's Leverage Index concept (my favorite sabermetric invention of the last umpteen years) was the original inspiration for this approach.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Comments are closed.

The drama index

by Dave Studeman

December 29, 2008

The other day, I wrote about something I called the **Season Leverage Index**. The idea is simple: I want to quantify how much more important victories in September are compared to victories in April.

After the fact, of course, they're not more important at all. A win is a win, no matter when it occurs. But as the season is played, and the games dwindle down to a precious few, each game a contending team plays certainly feels more important. Books are written about the final contests of a season, not the first ones. Think we ever would have heard of **Merkle's Boner** if Fred Merkle hadn't touched second base in May? Or that we ever would have heard Bobby Thomson's **Shot Heard 'Round the World** if it had been smacked in the spring? It would have been more like a shot heard 'round the borough.

Yet I'm not really satisfied with this nebulous feeling that late-season games are more important than early-season games. As I said, I don't think they're really more "important." I think they're more "dramatic." And I like drama as much as the next theater critic. There's nothing wrong with enjoying (and writing about) those end-of-season nail-biting battles.

Secondly, being an unmitigated baseball nerd, I want to know "how much more dramatic." Seriously. If September games are more dramatic than April games—so much that we even vote for our MVPs based on how well they did in the more dramatic contests—how much more dramatic are they? Or, what's the potential for them to be dramatic? (In the end, players provide the drama; the timing of those games just provides the context.)

That's what I started to do in my last article, with my notion of season leverage index. But I've had second thoughts about the name. We're not really talking about

leverage here, we're talking about drama. So I've got a new name for my concept: the drama index. There is no guarantee I won't change the name again next week.

The idea is simple, but the spreadsheet is ginormous. To calculate the drama index, I assume that each team has a natural distribution of games it is likely to win in the remaining season. I do this by calculating a binomial distribution (BINOMDIST in your Excel spreadsheet) for each possible situation during the season. For instance, the binomial distribution of a .500 team with ten games to go will project the following number of wins this percent of the time:

Games	Percent
0	0%
1	1%
2	4%
3	12%
4	21%
5	25%
6	21%
7	12%
8	4%
9	1%
10	0%

So if this team has to win seven or more of its ten remaining games to make the postseason, there is a 17% chance they will do so (count from the bottom up: 1% plus 4% plus 12%). That's called cumulative probability.

Next, I look at what would happen to the binomial distribution if a team were to win or lose that day's game. For instance, if our example team were to win its game, it would have to win six of its next nine games to qualify for the postseason (momentarily setting aside its competitors' outcomes). The cumulative probability of winning six of nine is 25%. But if it were to lose that day's game, it would have to win 7 of its next nine games and the cumulative probability of that is 9%. So, the drama index is 16 points (25% minus 9%). That's the potential impact of that day's game on the team's chances of making the postseason.

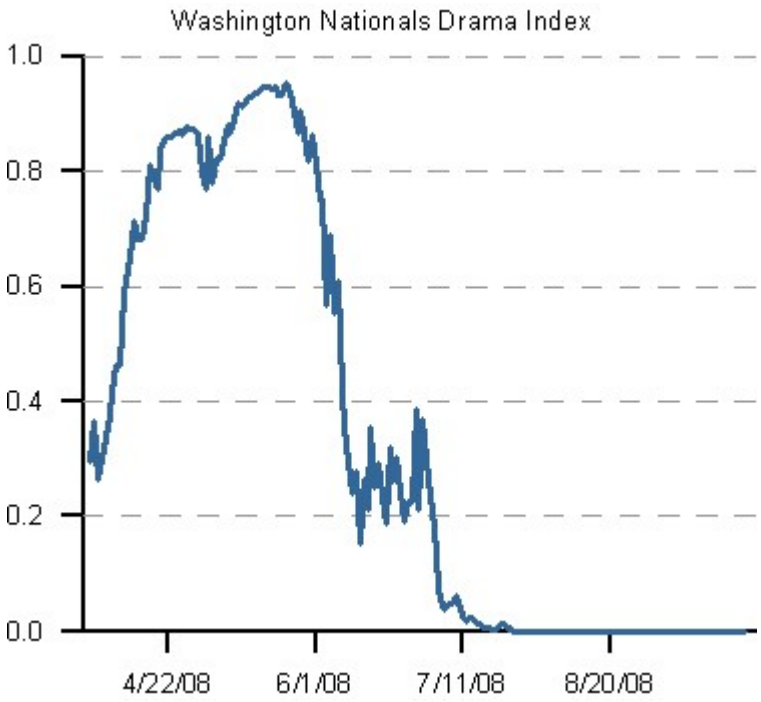
My recipe for the whole shebang: look at each day of the season and calculate the number of wins each team is behind the nearest postseason position (in other words, its division leader or the wild card leader, if closer) or, for division and wild card

leaders, how much they are ahead; then use the binomial distribution to calculate the impact of a win or loss on its chances of reaching the postseason (assuming the competition plays .500 ball the rest of the season. Since the index is updated daily, .500 isn't a prediction. It's just a guideline for the index.)

In the above example, the highest drama index goes to the team that has to win five or six of its remaining ten games, because the impact of a win or loss in those two situations gives them the same result (25 points). I'm not going to show you the math. Trust me.

For technical reasons I'll describe later, I used a .550 winning percentage in my binomial calculations instead of .500. But .500 is most useful for the explaining. Oh, and one last note: I divided each team's drama index by the average drama index of a .500 team throughout the year. So a drama index of 1.0 is a useful benchmark.

Enough with the details. I'll give more details below, and I'm happy to answer any questions in the comments. But let's look at some particular teams. Here is a graph of the day-by-day drama index of the Washington Nationals:



In late May, the Nats were "only" eight games out of first, and there was a little sizzle going on. But overall there was very little drama to the Nationals' season, and none at all after mid-July. In fact, every single one of their regular-season games had a drama index under one.

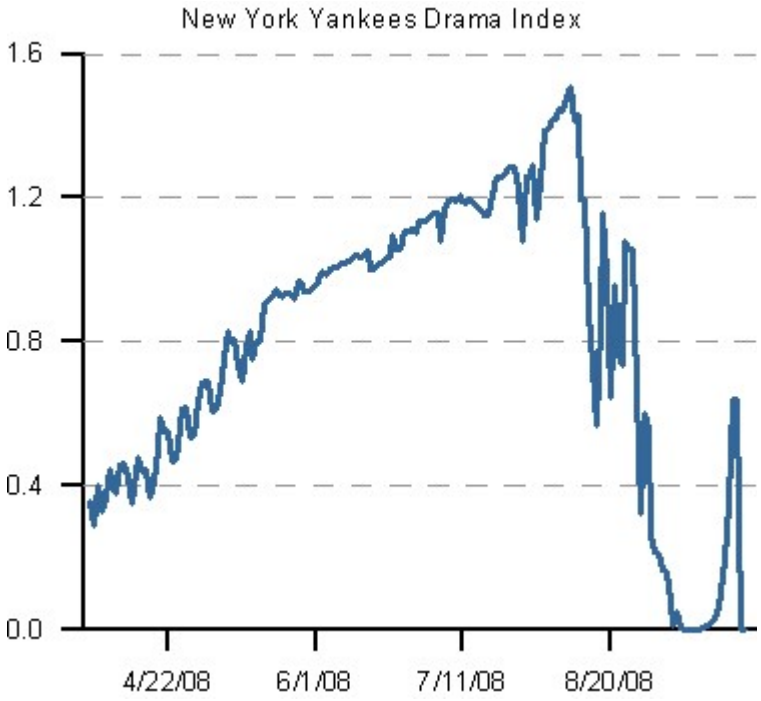
There were six teams that didn't have a single game with a drama index of at least one. They were:
{exp:list_maker}Chicago Cubs
Kansas City Royals
Los Angeles Angels of Anaheim
Seattle Mariners
Tampa Bay Rays
Washington Nationals{/exp:list_maker}See how it works? The Nationals had no drama cause they were never really competitive. The Angels had no drama cause they were too good and the competition was too weak. The Rays didn't have any drama in this system because, even though they battled the Red Sox for first place in the AL East, they had the wild card qualification to fall back on. Their postseason appearance was never seriously at stake.

By the way, several people mentioned playoff probabilities (such as those at **Baseball Prospectus**) as a way of measuring game-to-game drama. However, playoff probability is different. If two teams are tied for the division lead late in the season, they're playing very dramatic games—but if they both win, their playoff probabilities will hardly change. The analogy is imperfect, but playoff probabilities are like **WPA**; they reflect the outcome. The drama index is similar to **Leverage Index**; it reflects the importance of the game, regardless of the outcome.

Back to the matter at hand. The following teams had the most games with a drama index greater than 1.0:
{exp:list_maker}Minnesota Twins: 68
Los Angeles Dodgers: 66
New York Yankees: 64
New York Mets: 59{/exp:list_maker}You're probably not surprised by this list, though I was a bit surprised that the Yankees were in the top four. Here is a graph of their day-to-day drama index:

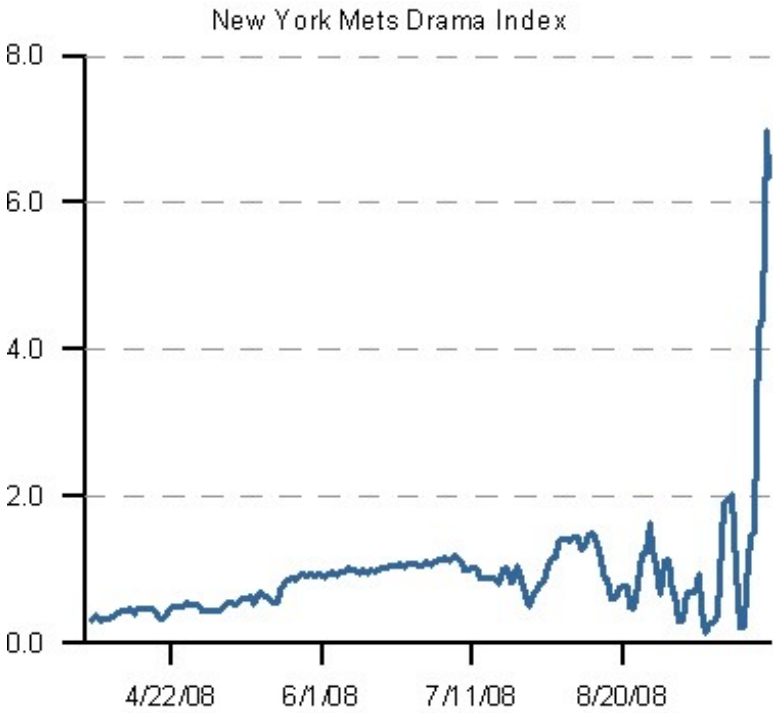
More from The Hardball Times

A Hardball Times Update
by RJ McDaniel
Goodbye for now.



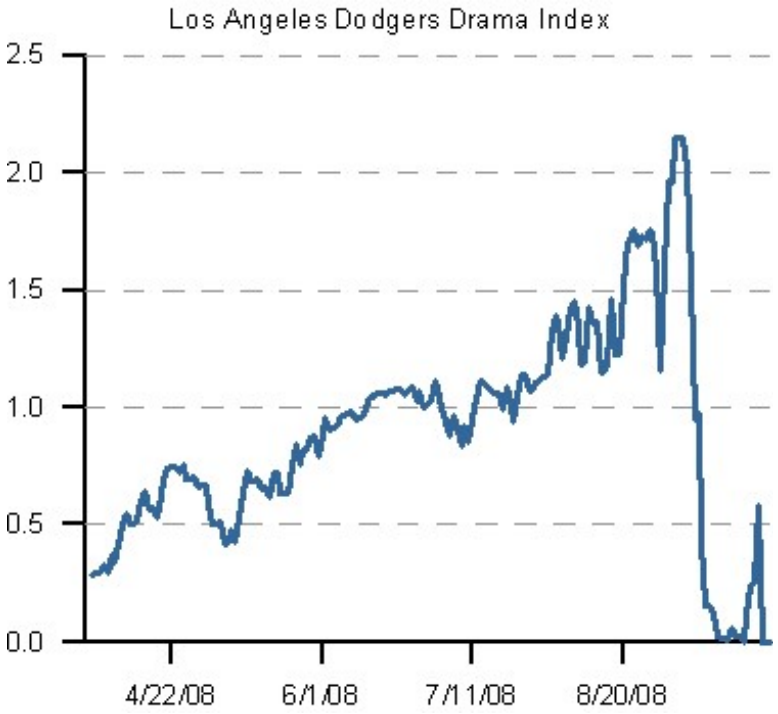
The Bronx Bombers were fighting in June and July, and in late July they had almost tied the Red Sox for second place (and the wild card lead). The potential for theatrics was high. But Boston surged ahead and the Yankees couldn't maintain the pace. By early September, their DI was pretty much zero, though there was a little sizzle around September 22 when they hadn't been eliminated yet. Once they were, their index dropped to absolute zero.

The Mets, on the other hand, kept the drama going until it spiked at the very end:



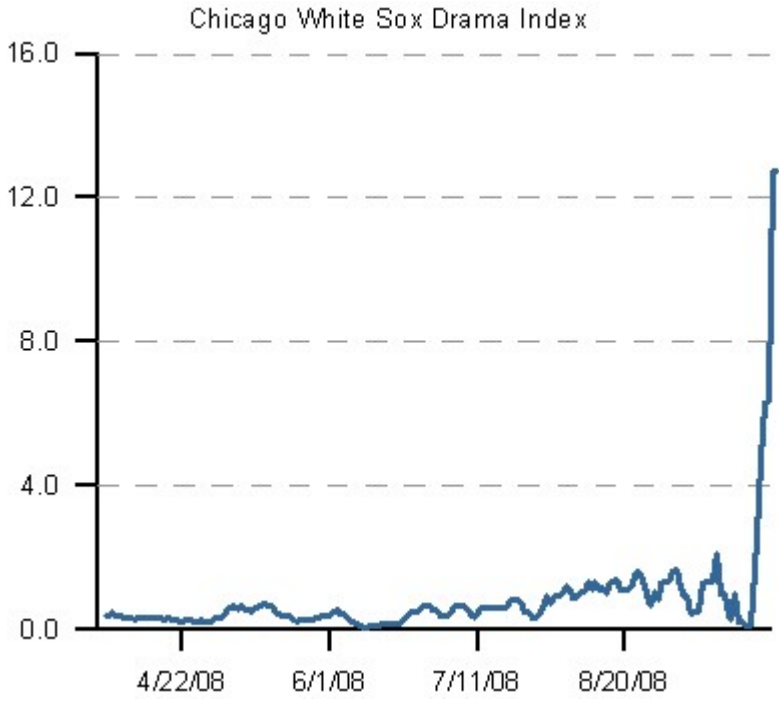
There were some wild swings to the Mets' late season, as they battled first the surging Phillies for the NL East lead, then the surging Brewers for the wild card lead. Their last two games were particularly dramatic; the next-to-last game actually had a slightly higher index than the last game. When they won and the Brewers lost, the DI for the final game eased a bit.

The Dodgers also had a lot of ups and downs to their season, though not as dramatic as the Mets':



August was the month of drama at Chavez Ravine, when the Dodgers surged to catch the Diamondbacks. Their index spiked over 2.0 just before they finally caught Arizona in early September, and it fell quickly as they pulled away. There was a little surge of drama at the end, when the Diamondbacks played a little better and the Dodgers played .500 ball. But once they clinched the title, the index fell to zero.

But, of course, nobody could beat the American League Central for drama, where the White Sox and Twins extended the season for two extra days of excitement. Let's talk about the Sox:



That is one wacky graph. The Sox only had one game with a drama index above 1.7 before the final Wednesday of the season, then they played nothing but thrillers. It was mostly their own fault, as they lost five straight games before winning their final three.

See that little extension to the right at the very tippy top of the graph? That's because the Sox had two do-or-die games in a row (against the Tigers and Twins), with the maximum drama impact of 12.7. That's the way the system works. Games at the end of the pennant race are nearly thirteen times as dramatic as the benchmark, and about 40 times more dramatic than games at the beginning of the season.

As sort of a postscript, here is a little information for you to nosh on: the record of all teams in games with a drama index greater than one. For each team, you've got the number of games of high drama (over 1.0), the team's overall average drama index, and their record in games over 1.0. The best teams when the games "counted" were the Brewers, with a winning percentage of .708, and the Phillies, at .700.

Team	G>1	Avg.	W	L	Win%
MIN	68	1.109	41	27	.603
LAN	66	0.867	37	29	.561
NYA	64	0.763	36	28	.563
NYN	59	0.916	35	24	.593
FLA	40	0.632	21	19	.525
PHI	40	0.778	28	12	.700

DET	39	0.631	21	18	.538
TEX	39	0.641	18	21	.462
CHA	35	0.831	22	13	.629
STL	33	0.594	16	17	.485
ATL	31	0.538	10	21	.323
BAL	29	0.474	15	14	.517
OAK	28	0.486	13	15	.464
MIL	24	0.819	17	7	.708
SF	22	0.554	8	14	.364
ARI	16	0.490	7	9	.438
HOU	13	0.592	4	9	.308
CLE	9	0.389	3	6	.333
TOR	9	0.537	3	6	.333
SD	6	0.419	1	5	.167
PIT	5	0.491	0	5	.000
CIN	4	0.490	0	4	.000
COL	4	0.579	2	2	.500
BOS	1	0.323	1	0	1.000
CHN	0	0.169	0	0	.000
KC	0	0.358	0	0	.000
LAA	0	0.150	0	0	.000
SEA	0	0.245	0	0	.000
TB	0	0.239	0	0	.000
WAS	0	0.334	0	0	.000

Think those records contributed to the fine MVP showings of the Ryan's? (Howard and Braun). That will be the subject of my next article.

References & Resources

Here's a little information about the team-by-team (and game-by-game) drama index in 2008:

{exp:list_maker}Last year, the average drama index for each team was .546. About 6% of all team games had a drama index of absolute zero (By that, I mean that one of the two teams in the game had a DI of zero.).

When a team has to win its one remaining game to make the playoffs, the drama index is 12.73. That applied to only a few team games last year: the White Sox/Tigers and the White Sox/Twins games, after the regularly scheduled season had ended. Games in which there is a tie for first going into the last game of the season, such as those played by the Mets and Brewers last year, have a drama index of 6.36. The

index is lower than that of the one-game playoff because there is a 50% chance that the other team will lose its game.

Games in which the team has to win two of its last two games, such as the Mets' next-to-last game, have a DI of 7.0. At that point, the Mets were a game behind the Brewers, and the index assumes Milwaukee would win one of its two remaining games. Therefore, the Mets would have to win two to force a playoff.

{/exp:list_maker}The drama index is very steep. As the season closes in late September, most competing teams will have a game DI of 2.0 to 4.0; only in the last four or five games is the index likely to exceed 4.0.

Why did I use binomial distribution tables for a probability of .550 instead of .500? Consider the following cumulative probability table for a .500 team:

	Games Left		
Wins Needed	0	1	2
0	1.00	1.00	1.00
1		0.50	0.75
2			0.25

Now, consider the team that has to win two of its next two games (which is 0.25 on the table). The drama index of that situation is .5 (if you move to the left on the table, you can hopefully see why; it's .5 minus 0). Now look at the team that has to win one of its next two games (0.75 on the table). Move to the left, and the drama index is also .5 (1.0 minus .5).

I don't know about you, but this table doesn't pass my smell test. I think that having to win two of the next two games is more dramatic than having to win one of the next two. I think the problem is related to the way the table deals with situations at the "ends" of the distribution.

To fix it, I changed the underlying probability to .550, which is the winning percentage of a 90-win team (which usually qualifies for the postseason). The table now reads:

	Games Left		
Wins Needed	0	1	2
0	1.0000	1.0000	1.0000
1		0.5500	0.7975
2			0.3025

Now, the drama index for a team that has to win two of its next two games is .55. For the team that has to win one of its next two, it's .45. I have no idea if that's "right," but at least it passes my smell test.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Comments are closed.

Postseason probability added

by Dave Studeman

January 15, 2009

So now we know two things about the 2008 season: how "important" specific games were in each pennant race **(the drama index, or DI)**, and how much players helped their teams win specific games **(Win Probability Added, or WPA)**. Like peanut butter and chocolate, love and marriage, a horse and carriage, these two things go together. Let's do it. Let's create a new stat.

Why? Because with these two tools, we can derive a unique look at baseball players in 2008, one that's in tune with how many fans and writers think of "value." In fact, our original question concerned last year's National League MVP voting, in which **Ryan Howard** finished a strong second to **Albert Pujols**. Howard had strong superficial stats (home runs and RBI's) but didn't come close to Pujols' real overall value. Yet many people thought Howard should have won the MVP anyway because of his strong September performance.

That's been our quest: to measure the impact of Howard's timely performance. We've answered the question of how "important" September games truly are (I prefer the word "dramatic") with our drama index. Now, we're going to combine the drama index with WPA to develop PPA (oh no! another acronym!), or Postseason Probability Added. Quite simply, we're going to multiply each player's WPA in each game times that game's drama index. And that's all the explanation you're going to get out of me, because it's that simple.

So how do Pujols and Howard compare?

Player	Team	WPA	DI	PPA
Howard, R	PHI	2.37	0.75	4.48
Pujols, A	STL	6.39	0.63	4.39

Well, look at that: a virtual PPA tie. Even though Howard's overall Drama Index was only slightly higher than Pujols', Howard produced in games with the highest index (in other words, "when it counted") and he completely closed the four-win gap in WPA. Pujols' production was more even during the year, and the ratio between his WPA and PPA reflects that.

Remember, however, that we're not including fielding prowess here, and Pujols was about two wins better than Howard last year. Pujols was still the more valuable player, by two wins. Two wins is a lot. Howard shouldn't have received a single vote ahead of Pujols on any ballot. But we know that many MVP voters don't value fielding, and our PPA result seems to be a good reflection of the actual MVP voting.

But that's not the end of the story.

It turns out that neither Pujols nor Howard finished first in NL PPA last year. A certain New York Met actually contributed more than any other player to his team's pennant drive, and second place wasn't even close. Here are the top ten NL PPA leaders among all position players:

Player	Team	WPA	DI	PPA
Beltran, C	NYN	5.02	0.97	8.30
Braun, R	MIL	3.68	0.87	4.63
Fielder, P	MIL	3.20	0.84	4.63
Howard, R	PHI	2.37	0.75	4.48
Pujols, A	STL	6.39	0.63	4.39
Ethier, A	LAN	3.78	0.90	4.38
Ramirez, H	FLA	4.74	0.68	4.31
Ramirez, M	LAN	3.51	1.07	4.13
Berkman, L	HOU	6.71	0.59	3.45
Wright, D	NYN	4.18	0.98	3.39

In the last four games of the season, the Mets had drama indices of 4.25, 4.44, 7.00 and 6.36. In those four games, Beltran batted .429/.529/.643. Before those four games, his PPA was 4.44, in line with the other leaders. But that stretch of games boosted him far beyond the rest of the pack.

When you add in the extras—Beltran plays a premium position, plays it as well as anyone in the game and is among the very best baserunners in the game—you have a very, very credible case for **Carlos Beltran** as the 2008 National League MVP. Alas,

he finished 21st in the voting.

You probably noticed that **Manny Ramirez** is eighth on the list, even though he played less than half of the year in LA. It has to be said that the scale of PPA is an issue: if we were to compare Manny to replacement level instead of average, he would rank lower, maybe much lower. When you add the fact that Manny plays a "secondary" position—and not very well—you have to conclude that his fourth-place finish in the MVP voting wasn't justified.

Overall, however, this list is a pretty good reflection of actual MVP voting. On the surface, it seems to do a better job than WPA or any other general stat of matching **the actual MVP results**. I have a feeling that, if we were to add a "kicker" for each team that actually makes the postseason, we'd have an even better match.

How about the American League PPA? How well does it match the MVP voting? Hold onto your hats:

Player	Team	WPA	DI	PPA
Span, D	MIN	1.98	1.56	5.16
Mauer, J	MIN	4.88	1.20	3.71
Dye, J	CHA	-0.26	0.88	3.48
Hamilton, J	TEX	2.80	0.66	2.99
Ramirez, A	CHA	0.31	0.97	2.54
Bradley, M	TEX	2.09	0.65	2.26
Cabrera, M	DET	2.95	0.64	2.24
Giambi, J	NYA	1.96	0.77	2.23
Quentin, C	CHA	3.81	0.57	2.18
Morneau, J	MIN	3.87	1.16	1.88

Yes, that **Denard Span**. On September 25th, Span hit a run-scoring triple in the bottom of the eighth to tie a key game with the White Sox (drama index of 4.7, WPA of .425). Multiply those out, and you get a PPA of more than two for just one play. Span also played a lot of critical games for the Twins—I thought that his average Drama Index of 1.56 was a typo until I realized that Span didn't play at all in May and June. He played "when it counted." To his credit, Span didn't have any big negative WPA days when it counted; most players do.

No wonder the BBWAA had such a hard time choosing an MVP. Alexei Ramirez fifth in PPA? The actual MVP winner, **Dustin Pedroia**, finished 27th in our rankings.

My own MVP choice, **Joe Mauer**, would be the logical PPA MVP pick, particularly when you factor in the fact that he is a fine catcher, to boot.

Here's another strange one: **Jermaine Dye**. Dye's overall WPA was actually below average, but he really delivered when it counted. Here's a breakout of his average WPA per game, grouped by how dramatic each game was:

More from The Hardball Times

A Hardball Times Update
by RJ McDaniel
Goodbye for now.

DI	G	WPA
0-1	120	-0.002
1-2	25	-0.011
2-3	2	-0.049
4-5	1	0.049
6-7	2	0.059
12-13	2	0.114

You're probably noticing that a few games can make a big difference. That is the nature of this particular beast. If you're going to insist that **when** players perform is as important as **how** they perform, a few games will count a whole lot more than the others. I'm not doing the insisting, by the way. I'm just doing the math.

Still, we've got two lists here. One appears to be a very good reflection of how some people interpret their MVP ballots. The other is a bit of a mess.

I've only looked at position players so far. Let's look at pitchers next. The National League:

Player	Team	WPA	DI	PPA
Santana, J	NYN	4.08	1.06	6.54
Sabathia, C	MIL	3.17	1.30	4.37
Lidge, B	PHI	5.37	0.85	4.31
Webb, B	ARI	3.43	0.49	2.85

Lincecum, T	SF	4.73	0.55	2.43
Smith, J	NYN	1.10	1.04	2.30
Hamels, C	PHI	2.51	0.79	2.15
Billingsley,	LAN	1.46	0.84	2.15
Qualls, C	ARI	0.42	0.49	2.13
Wilson, B	SF	1.77	0.55	2.02
Marte, D	PIT	2.41	0.78	1.99
Hudson, T	ATL	2.30	0.78	1.70
Myers, B	PHI	-1.46	0.81	1.67
McClung, S	MIL	-0.05	0.82	1.64
Wade, C	LAN	1.85	0.89	1.62

Good relievers rank particularly well in the WPA system (another reason to use replacement level as the benchmark; it would give a boost to starters who pitch more innings), so I expanded the list to fifteen to include more starting pitchers.

A few pitchers did appear on MVP ballots. Sabathia finished sixth, Lidge was eighth, Santana was 14th, Webb was 17th and Lincecum was 23rd. In other words, the top five finishers in pitcher PPA all made it onto MVP ballots. Once again, the National League standings make a lot of sense.

I want to point out one other pitcher on the National League list, Philadelphia's **Brett Myers**. As you may know, Myers had a terrible first half and was demoted to the minors in early July. When he returned to Philly, his reemergence helped fuel the Phillies' pennant drive, and he pitched very well in some of their most dramatic games. For that effort, he turned a negative WPA (-1.46) into a positive PPA (1.67).

Here are the American League pitchers:

Player	Team	WPA	DI	PPA
Danks, J	CHA	2.99	1.11	7.08
Baker, S	MIN	2.82	1.29	5.29
Jenks, B	CHA	3.47	1.02	4.13
Rivera, M	NYA	4.47	0.80	2.83
Nathan, J	MIN	3.26	1.31	2.83
Mijares, J	MIN	0.48	4.05	2.53
Buehrle, M	CHA	1.57	0.81	2.44
Mussina, M	NYA	2.20	0.76	2.38
Lee, C	CLE	5.96	0.39	2.37

Guardado, E	MIN	-0.33	2.08	-0.77
	TEX	2.85	0.87	3.08
Duchscherer, J	OAK	2.16	0.69	2.27
Ziegler, B	OAK	3.20	0.54	2.20
Downs, S	TOR	2.51	0.58	1.98
Chamberlain, J	NYA	2.28	0.68	1.80
Soria, J	KC	4.08	0.37	1.77

Because Chicago and Minnesota played the most dramatic games of the year, their pitchers rank highly here. **John Danks** is the best example of the extreme impact a game can have. He pitched a fantastic game in the final game of the season for the Sox, picking up a total of **7.5 PPA points in one game alone**! Take that one game away, and his PPA is below average.

We had two extremes in the American League last year that impacted the PPA list. You may recall, from **the drama index article**, that the White Sox had a "hockey stick" graph. Until their last six games, they didn't have a DI over 1.7. After that, they were drama queens, with an average DI of 7.6. At the other extreme, K-Rod (who finished sixth in MVP voting) doesn't make this list because the Angels had no dramatic games last year. They ran away with the division.

I think this explains the wacky results in the AL. PPA seems to work pretty well when the league is generally competitive, across divisions and over the full season. In the American League, the only real drama concerned the Central division, and that drama was extreme. So PPA doesn't match the MVP results (or any kind of "smell test" I can think of) as well.

Of course, we can play around with this system. I could set up the distribution tables so that final games aren't dramatically better than other games, but then I'd be muting the impact of those other games, too. I could set an arbitrary cutoff for the highest drama index at seven or whatever, but that's so … arbitrary.

I actually did play with a running five-game average of the drama index instead of setting the drama index specific to that game. I used a five-game average because most pitching rotations are five-man rotations, and DI seems to really take off in the last five games of the season. That did mute the impact somewhat. For instance, **Joe Mauer** leaped to the top of the AL list and **Denard Span** dropped a bit. But it didn't make a huge difference. I think the system that I've presented here does what it's supposed to do.

Every idea, carried to its logical extreme, becomes a caricature of itself. We've created a caricature here, but I think we've learned a few valuable lessons along the way. In general, sabermetrics helps processes like MVP voting by providing insight where generalities were previously used. For instance, RBI's were used (and, in many cases, still are used) as a proxy for batting prowess and clutch hitting. Sabermetrics has brought a microscope to RBI's, and provided better data such as Base Runs and true clutch hitting figures.

Here's another example: many MVP voters like to reward players who play for contending teams, and some will almost always vote just for players whose teams make the playoffs. But we don't have to rely on team winning statistics anymore. We've not only shown the relationship between certain skills and winning (the essence of **Win Shares**), we've even developed stats that document, on a play-by-play basis, how much each player contributed to his team's chance of winning (the essence of **WPA**). We don't have to rely on a team's winning record as proof of how much individual players contributed to winning.

We've now done the same thing with "performing when it counts during the season." Using an approach that makes intuitive sense, we've given more weight to performances in September games, and we've found that sometimes the results make a lot of "sense." And sometimes they don't.

Which is fine by me. Despite my penchant for math and big spreadsheets, I'm not really a reductionist. I believe that MVP voting shouldn't be totally objective. But I do like to shine a light on subjectivity when we can. Consider this exercise a flash of light.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Comments are closed.