BASEBALL ANALYST

JOURNAL OF SABERMETRICS

December, 1988 Vol. #39

Hi. My name is Rob Neyer (rhymes with liar), and I'm Bill's new assistant. As such, one of my duties is the editorship of the Analyst. Not to worry -- Bill will still be reading your articles and writing his own. I just get to do stuff like type the page numbers and lick hundreds of stamps.

Another change -- Bill is renting an office again (I think to ensure that I won't be at his home every day), the address of which appears at the bottom of this page. Please mail all correspondence to the new office.

Now, a note from a regular contributor:

Dear Analyst Readers,

The two articles I wrote on computer simulations are unfortunately flawed. When writing the program, I inadvertently added IBB to TBB, with the result being that the seeded walk totals with which I tested the program were seriously inflated. I was using a GiDP-adjusted stolen base version of runs created which, first of all, worked quite well on my calculator only because of the inflated walk totals, and secondly, which was mis-typed into the program code anyway, so that all of the results looked approximately three to four percent more accurate than they really were. The only conclusion which is unaffected by this is the assertion that baserunning -- taking extra bases on hits and outs -- ultimately accounts for about 50% of all runs. Please accept my apologies,

Gary Fletcher.

Thanks, Gary. We have five articles this issue.

Editor and Publisher: Bill James Address: Box 327, Oskaloosa, KS 66066

AN ADDENDUM TO BRUCE GARLAND'S PRELIMINARY REVIEW OF OUTSIDE

INFLUENCES ON RICH ASHBURN'S FIELDING STATISTICS

BY: Bill Carr

I read Bruce's fine article on Richie Ashburn with great interest. Because of some research I'm currently doing, I happen to have the Sporting News box scores for 1950 lying around. The article, therefore, led me to run a quick analysis of Ashburn's fielding stats for the 1950 season. Please note that I had to estimate innings for four games in which Ashburn only played part of the game in center field (best guess estimates based on box score info, primarily at bats).

HOME VS. ROAD

	INNINGS	PO	A	E	DP	Putouts/9 Innings
HOME	676.0	211	3	1	1	2.81
ROAD	618.0	195	5	4	1	2.84

DAY VS. NIGHT

	INNINGS	PO	A	E	DP	Putouts/9 Innings
DAY	805.0	254	6	3	1	2.84
NIGHT	489.0	152	2	2	1	2.80

BY STARTING PITCHER

	INNINGS	PO	A	E	DP	Putouts/9 Innings
CHURCH	152.0	54	0	1	0	3.20
DONNELLY	8.0	2	0	0	0	2.25
HEINTZELMAN	119.0	47	1	0	0	3.55
JOHNSON	73.0	23	0	0	0	2.84
MEYER	201.0	61	3	1	0	2.73
MILLER	179.0	55	0	1	0	2.77
ROBERTS	335.0	93	4	2	2	2.50
SIMMONS	227.0	71	0	0	0	2.81

BY MONTH

	INNINGS	PO	A	E	DP	Putouts/9 Innings
APRIL	58.0	8	2	0	0	1.24
MAY	224.2	71	0	1	0	2.84
JUNE	188.0	53	0	2	0	2.53
JULY	298.0	97	2	0	0	2.93
AUGUST	262.0	84	0	1	0	2.89
SEP/OCT	263.1	93	4	1	2	3.18

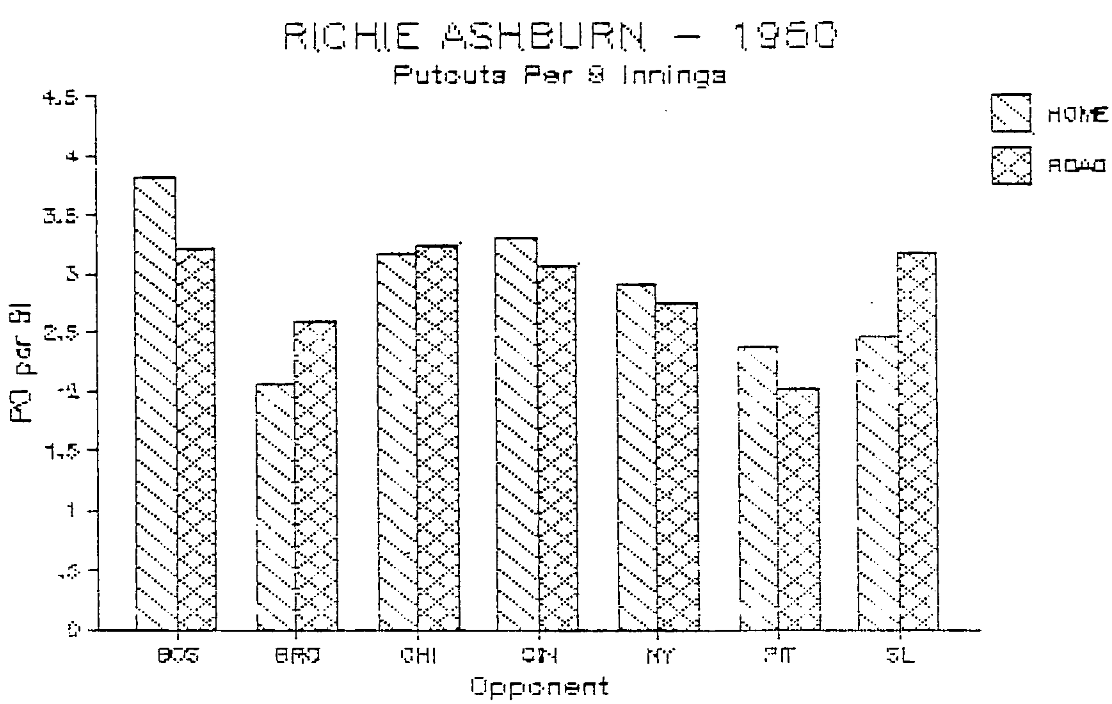
BY OPPONENT

Home	Inn	PO	A	E	DP	PO/9I	Road	Inn	PO	A	E	DP	PO/9I
BOSTON	59.0	25	0	0	0	3.81		101.0	36	2	0	0	3.21
BROOKLYN	109.0	25	0	0	0	2.06		79.2	23	2	0	0	2.60
CHICAGO	99.0	35	3	0	1	3.18		78.0	28	0	0	0	3.23
CINCINNATI	109.0	40	0	1	0	3.30		97.0	33	0	0	0	3.06
NEW YORK	102.0	33	0	0	0	2.91		88.1	27	1	2	1	2.75
PITTSBURGH	99.0	26	0	0	0	2.36		103.0	23	0	2	0	2.01
ST. LOUIS	99.0	27	0	0	0	2.45		71.0	25	0	0	0	3.17

In looking at this data, I find a couple of surprises. There was virtually no difference, at least for 1950, in Ashburn's range at home and on the road (the day/night split is, I guess, not too surprising). He had the <u>lowest</u> (except for Donnelly who started one game) putout ratio in Roberts' starts (Curt Simmons had the same number of strikeouts, 146, as Roberts in two thirds (215 vs. 304) the innings). That tells me that, at least in 1950, Roberts was a ground ball pitcher.

The home/road splits by opponent are presented in the accompanying bar-graphs and show an interesting consistency in fly ball tendencies by opponent regardless of location.

There's a lot of interesting information available in this kind of breakdown, and I'm going to keep plugging away on the '50 AL and NL seasons. I hope to be back with a comprehensive look in about a year. Meanwhile, I think the data presented above casts some doubt, even though it only covers one season, on the effect of the ballpark and of Robin Roberts on Richie Ashburn's defensive stats. I think he belongs in The Hall.



BREAKING IN A CATCHER

by Tom Hanrahan

INTRODUCTION

Frequently you will hear T.V. announcers referring to baseball players "adjusting" to new situations: a different park, a new team, a spot in the batting order, a new position, etc. I have done some research in an attempt to answer what might be some of the more important of these: Do pitching staffs go through an adjustment period when breaking in a new catcher? Does it apply to catchers with new teams, or just a change between leagues? Do veteran catchers handle pitching staffs better than rookie receivers?

THE DATA USED

Bill James' 87 Abstract presented records from 1982 to 1986 for every catcher active at the end of the 86 season. The Abstract listed games started at catcher, team ERA when the catcher was in the lineup, and overall team ERA for the year. I used the team's ERA as an indicator of how the pitching staff felt at home with the catcher; of course it included the effects of his overall defensive abilities. Specifically, I compared team ERA when the man started the game behind the plate with the overall team ERA, weighted by games caught. I also made comparison to the previous year's team ERA, as a way of measuring further the established performance level of the pitching staff. I used every catcher who started at least 10 games during any year. Ten games was also the minimum used to determine whether a catcher was in his first (second, third, etc.) year with a team. I know 10 games is a small amount, but I figured that any catcher who started 10 games probably hung around the team enough to get to know the pitchers. And, my data base would shrink even smaller if I raised the minimum.

GROUPING THE DATA

	Title
Catchers in their 3rd or 4th year with a team =	VETERANS
Catchers with no prior experience in the bigs =	ROOKIES
Rookies are subdivided into	
--- catchers who caught at least 50 games in their careers after the rookie season=	SURVIVORS
--- those who never caught 50 more games =	WASHOUTS
Catchers, other than rookies, in their first year with a new team ---	
--- in the same league =	NEW TEAMERS
--- after switching leagues =	NEW LEAGUERS

RUNNING THE NUMBERS

The table on the next page shows how catchers in the various categories fared. The titles over each column are explained this way:

- # of catchers - number of catcher seasons in the sample.
- # of games - how many games were started behind the plate by all of the catchers in the sample.
- ERA vs team - the weighted sum of the difference in ERA of the pitching staff between games started by these catchers, and the overall team ERA for the year.
- team vs last year - the difference between this year's and last year's overall team ERA.
- ERA vs last year - the sum of the previous two columns.

BREAKING IN A CATCHER

TABLE - CATCHER TYPE vs PITCHING STAFF EFFECTIVENESS

category	# of catchers	# of games	ERA vs team	team vs last year	ERA vs last year
NEW LEAGUERS	28	1685	.03	-.04	-.01
NEW TEAMERS	14	754	.00	.02	.02
VETERANS	50	3942	-.03	.09	.06
ROOKIES	42	1734	.09	.06	.15
- SURVIVORS	33	1414	.10	.06	.16
- WASHOUTS	9	320	.05	.05	.10

NOTES

1. Negative numbers mean better ERA.
2. You might note that the numbers are generally positive; that's part because in the years studied, offense was usually increasing, so ERAs were often higher than the year before. The important point is to compare the numbers to each other, not to a zero sum.
3. For these sample sizes, an ERA difference of .10 is fairly significant. If you assume a standard deviation (SDEV) of 3 runs per game for a normal team, that makes the SDEV for a 900 game sample to be equal to 3 / (900)^.5 = .10. Trust me, or go buy a statistics book.

OBSERVATIONS AND DISCUSSION

NEW TEAMERS vs VETERANS : Changing teams within a league apparently didn't hurt.
NEW LEAGUERS vs VETERANS : Changing leagues wasn't real traumatic, either. While the teams with VETERANS had a lower (-.03) ERA when their VETERAN catchers were in the game, they had poor performance compared to the previous year, while the NEW LEAGUERS actually had their staffs perform slightly better (-.01) than before they had arrived. The data show about the same results for both the NEW LEAGUERS and the NEW TEAMERS.
ROOKIES vs VETERANS : VETERAN-caught staffs were .03 runs per game better when the VETERAN receiver started, than with the reserve or other catchers. On the other hand, teams with ROOKIES were .09 runs per game better when "other" (presumably veteran) catchers started. These differences would lend credence to many broadcasters' oft-spoken comments regarding the advantage of having a Bob Boone or a Carlton Fisk to handle the staff. The difference between ROOKIES and VETERANS was .09 (.15 minus .06) if you compare performance with the previous season.
SURVIVORS vs WASHOUTS : The higher ERA when using ROOKIES as compared to VETERANS does not appear to be a quality leakage; that is, the staffs with rookies who hung on to play in the majors fared no better than staffs with the WASHOUTS who didn't stick. To further convince myself of this, I isolated the 10 catchers who appeared in both the ROOKIES and VETERANS categories in the study.

[For your info, the catchers and years used are presented in the form]
[Name Year (with R for rookie and V for veteran):]

[Bailey 84R,86V	Bilardello 83R,85V	Bando 82R,84V,85V	Engle 83R,85V]
[Kearney 82R,86V	Lake 83R,85V	Schroeder 83R,85V,86V	]
[Slaught 82R,84V	Tettleton 84R,86V	Virgil 82R,84V,85V	]

BREAKING IN A CATCHER

The VETERAN-led pitchers were .32 runs per game better than the ROOKIES group, when compared to expected performance. (If that were the real difference in having two different men behind the plate, it could mean .32 * 162 = 52 runs over the course of a season!) This was a small sample (1251 total # of games), but it confirmed that the difference between ROOKIES and VETERANS was not distorted by the presence of a few lousy ROOKIE catchers skewing the data.

CONCLUSIONS

(1) Staffs with veteran catchers perform better than those with rookie catchers. This could be due to either a) improved defensive skills as a catcher matures, or b) better handling of major league pitchers. It does not appear to be c) familiarity with a particular staff of hurlers, nor d) that the veteran catchers were just better catchers than the rookies in this sample. To determine whether a) or b) is the correct interpretation, we might need to study some OLDER rookie catchers, whose defensive skills would not be improving as they became vets.
There's some recent data we can add here. In 1987, there were 4 good, young rookie catchers in the majors: Steinbach, Surhoff, Nokes, and Santiago. In 1988, the second year for each with their clubs, their teams' ERAs improved by an average of (minus) .22 runs per game relative to the league (A's -.87; Brewers -.56; Tigers +.30; Padres -.36). Did their defensive skills improve, or did they handle pitchers better? Either way, I'll conclude that:

(2) A team with a rookie catcher last year could expect to allow fewer runs this coming year - by somewhere between, say, .05 and .20 runs per game, all other things being equal.

(3) A general manager, in a move to try to win next year's pennant, should not have the fear of breaking in a new receiver to the pitching staff dissuade him from importing a veteran catcher from another team in either league.

On the other hand,
(4) Planning to bring in a rookie catcher may not be a good move for a team that wants to win the title next year. Yes, the Tigers won in 87 with the rookie Nokes, but it was mostly the offense that did it. Although their pitching declined in 88, I think it's obvious that they had a lot of hurlers in the decline stage of their careers.

A SOMEWHAT UNRELATED BUT IMPORTANT COMMENT

If differences in team ERA of .10 can be attributed to using different types of catchers, then very possibly larger differences than that could be found between individual catchers' (in)abilities to call a game. And just maybe, this would mean that the defensive value of the man behind the plate is more important than that of middle infielders', and thus would be the most crucial of any everyday player on the field.

Tom Hanrahan
104 Endeavor St.
Lexington Park, MD
20653

PITCHER RUN AVERAGE

The most reliable of the pitcher statistics is the Earned Run Average (ERA), but it has three major flaws.

(1) ERA for a starter is deficient because it is contaminated by performance of his bullpen, a factor I will call the "bullpen effect". The reliever controls the fate of the inherited runners, yet they are the responsibility of the starter.

(2) ERA for a reliever is deficient because it doesn't reflect hi ability to handle inherited runners, perhaps his most critica responsibility.

(3) The ERA definition of "unearned" run too often produces resul that run contrary to common sense.

We can make two refinements to ERA that would remedy, to a large extent, these shortcomings. First, we can more sensibly allocate runs between a reliev<u>ee</u> and his reliev<u>er</u>. Second, we can more sensibly define what <u>are</u> "unearned" runs. For ease of expression I will refer to this improved ERA as Pitcher Run Average (PRA).

Central to the theory of PRA is our ability to quantify the run potential of an inning. Palmer has assigned historically derived averages of what might be called "situational worth" or "run values" or "potential runs" to each of the 24 runners/outs situations. I will use these values throughout this discussion.

<u>Run Values</u>	Number Of Outs		
Runners	0	1	2
none	.45	.25	.10
1st	.78	.48	.21
2nd	1.07	.70	.35
3rd	1.28	.90	.38
1st,2nd	1.38	.89	.46
1st,3rd	1.64	1.09	.49
2nd,3rd	1.95	1.37	.66
full	2.25	1.55	.80

How Does PRA Work?

It helps to understand PRA by working through an example. Let's imagine that a reliever inherits an inning which PRA values at 1.3 runs by the table of run values.

We'd explain PRA to the relievee in this way ... The situation you left for your reliever is valued at 1.37 runs. PRA will charge you run account with these runs. It doesn't matter how many of them yo reliever actually permits to score, you'll be charged with 1.37, a so always will every other pitcher being relieved in the same position. In this way we isolate the relievee's performance from that of his reliever, and eliminate the bullpen effect.

This is how we'd explain PRA to the reliever ... You've inherited a position worth 1.37 runs. As compensation for this initial disadvantage, PRA will credit your run account -- actually reduce it -- by 1.37 runs. However, you will be charged in full for any runs that do score. That is, if you permit 2 runs to score, you've done relatively poorly and PRA will charge you with (-1.37 + 2) or 0.63 runs. If you hold the runs to 1, you've done comparatively well, and PRA will credit your account with (-1.37 + 1) or -0.37 runs. A so-so reliever, over the long haul in many such circumstances, will allow to score an average of 1.37 runs, precisely offsetting his credits. A good reliever will permit to score an average of fewer than 1.37 runs. And Stan Clarke's record will go from bad to terrible.

A Game Scored Under PRA

Clancy tires in the 7th, leading 4-2, with runners at 2nd and 3rd, and one out. Henke relieves, allowing one run to score on a sacrifice fly before retiring the side.

The ERA run allocation is simplistic. Clancy is charged with 3 runs; the 2 runs he allowed earlier in the game, plus the runner that scored on Henke. Henke is not charged with any runs.

PRA rations the blame quite differently. At the point Clancy was relieved the inning was valued at 1.37 runs. Clancy is charged with 1.37 plus two earlier runs for a total of 3.37 runs. Henke overachieves somewhat. He is credited with 1.37 runs, less the one he allowed, for a net credit of (-1.37 + 1) or -0.37 runs.

The game continues. Henke leaves in the ninth with two out and a runner on 1st. Ward promptly serves up a 2-run dinger, then a walk followed by a single. He gets the hook, leaving men on 1st and 3rd with two out, a position valued at 0.49 runs.

ERA would crudely charge both Henke and Ward with 1 run. This comes down far too hard on Henke and not nearly hard enough on Ward.

An inning with a runner on 1st and two out is valued at 0.21 runs. PRA charges Henke with these runs, making his accumulated effort on the day worth (-0.37 + 0.21) or -.16 runs.

PRA initially credits Ward with 0.21 runs for the runner inheritance, but then charges him 2 runs for the dinger. In addition, PRA notes that the position he left for the next guy is worth 0.49 runs. Therefore PRA's net charge to Ward for his dismal outing is (-0.21 + 2 + 0.49) or 2.28 runs.

Next in relief is Eichhorn, who quickly induces a pop-up to end the game. PRA credits Eichhorn with 0.49 runs -- this was the run expectation of the position but none scored.

Jays win the game 6-5. ERA and PRA have quite differently apportioned accountability for the 5 earned runs, as shown in the following table. The PRA allocation is much closer to our intuitive impression of how well the pitchers did.

	Run Allocation ERA	PRA
Clancy	3.00	3.37
Henke	1.00	-0.16
Ward	1.00	2.28
Eichhorn	0.00	-0.49
	----	----
Total	5.00	5.00

Unearned Runs

Certainly few would disagree that the ERA treatment of unearned ru often produces silly results. Consider the case where after pitchi a perfect game for 8 2/3 innings, a pitcher blows his cool and goe on to allow 7 runs after a teammate commits an error. The ERA "solution" to this event is to define all 7 runs as unearned, on t theory that the error should have been the 3rd out. Humbug. The co of the error was actually very small and quantifiable, although ER is not sophisticated enough to measure it. PRA, however, can.

First PRA determines, what the cost of the error has been in terms of runs scored and changes in situational value. PRA compares what did happen with what should have happened. The difference in run valuation is the damage caused by the error, the "unearned" runs.

Then PRA reimburses the pitcher for the damage by crediting him wit the unearned runs. Gives 'em back. Leaves him in a situation valued as if the error had not been made, although he now faces a differen runner/out position. The offset to this credit is the fact that any runs that do score during the commission of the error or subsequent to it are charged as earned runs. Over time, of course, an average pitcher will eventually allow to score the number of runs with whic he was credited, so errors in the long term won't impact his PRA.

Let's examine this concept in more detail with a couple of examples Take the Stieb blow-up. What did happen resulted in two out with a man on 1st, a situation valued at 0.21 runs. What should have happened was three out, zero runs scored. So the error cost Stieb 0.21 runs. Big deal. Allow him credit for them, but charge the 7 ru that did score, for a net cost of (-0.21 + 7) or 6.79 runs.

Another example. The bases are loaded with one out. Moseby boots a routine fly ball and when the dust settles, two runners have scored and 2nd and 3rd bases are occupied. How would PRA handle this?

First, PRA would appraise the cost of the error, the unearned runs. What did happen caused 2 runs to score, and left an inning with a run value of 1.37 --- the expected runs to be scored from an inning with men on 2nd and 3rd with one out. Total situational cost subsequent to the error, 3.37 runs.

What should have happened is bases loaded with two out, a position which is valued at 0.80 runs. (Note that here there may be room for judgement by the official scorer. He might deem that the ball was hi deeply enough that the runner on 3rd would have tagged and scored anyway, in which case the scorer would deal with a somewhat different "should have happened")

So, what have we? The error cost the pitcher (3.37-0.80) or 2.57 unearned runs. PRA allows him a credit for this amount but charges him with any runs that really do score -- in this case with the two that did score and any subsequent ones.

Observe that this process has left the pitcher with a situation valued as if the play had been made. With proper execution he would have been faced with an inning valued at 0.80 runs -- bases loaded, two out. As a consequence of the error PRA credits him with 2.57 unearned runs, charges him for the 2 runs that did score, and leaves him in charge of an inning valued at 1.37 runs, for a net positional value of (-2.57 + 2.00 + 1.37) or 0.80 runs. Bingo, no change.

Again, I'll repeat, over the long term a pitcher with average skills will neither benefit nor lose from this process, because the number of runs he will allow to score will precisely offset his credits. However this technique <u>will</u> benefit the records of pitchers who can "bare down in the face of adversity" and shrug off an error (Doyle Alexander? Jimmy Key? old pitchers?) and it will detract from the record of pitchers who lack self-control and self-destruct at times after an error is committed (Dave Stieb? young pitchers?).

Were PRA adopted by the baseball world, what would be its impact on pitcher statistics?

* PRA for a league or a team would be maybe 5 to 10% higher than ERA. Due to its more sensible definition of unearned runs, PRA considers relatively more runs to be earned.

* PRA includes the measurement of pitcher composure. In past years, Dave Stieb's apparent inability to accept errors as part of the game would be reflected in a higher PRA.

* For short relievers, PRAs would be lower than ERAs. Really good short guys might have PRAs less than 1.00, perhaps even negative, but this is rank speculation.

* PRA accurately measures reliever success with inherited runners, and it should become <u>the</u> measure of reliever performance.

* PRA comparisons between pitchers on different teams would be much more meaningful than ERA comparisons, because the bullpen effect will have been eliminated. Of course, there's still the park effect and the turf effect and the weather effect and ...

Doug Bennion
39 Playter Blvd
Toronto, Canada

RACISM IN BASEBALL-OFFENSIVE PERFORMANCE

James R. McNesby

INTRODUCTION.

Charges of racism in baseball have been made in recent times and to deny all such charges is not reasonable in view of the historical record. It has been reported [Scully, 1974] that in the period 1953-1965 black major leaguers had batting averages twenty points higher than did Whites and that Blacks had distinctly better won-lost averages among the pitchers. During this period black outfielders batted .012 higher than whites, although there was no significant difference between the two groups for second basemen and shortstops. First basemen and third basemen were reported to have particularly large batting average differentials favoring blacks. Scully infers from these observations that "stacking by position" occurred, which "stacked" blacks in non-central positions. It has been suggested [Eitzen, 1978] that the "central" positions are pitcher and catcher and that blacks have been excluded from these positions by racist management policies. This point has been made also in considerable detail by Edwards [Edwards, 1973] for a variety of sports. Richard Lapchick [Lapchick, 1984] reinforces the charge of "stacking" mainly by echoing the charges made by Edwards and by citing some facts: in 1983 of the 58 catchers in the major leagues none were black Americans although seven percent were "Latins" and only 6.6% of pitchers were "black" with another 7.7% being "Latins". There seems to be no distinction made by Lapchick between black Latins and white Latins although he does make a distinction between black Americans and white Americans. Bill James [Bill James Baseball Abstract, 1987, Ballantine Books, N. Y., 1987] compares records of pairs of black and white rookies who had similar rookie records and finds that the black member of the pair out-performs the white member by a "huge margin" over their respective careers. He is unable to explain the very distinct differences favoring black players.

I will try to show (a) that some residual bias still exists against selection of black players but that it is much smaller than in the past, (b) that player batting averages follow a normal distribution, (c) that black players have demonstrably greater foot speed on the average than white players and (d) that "stacking by position" that occurs is consistent with the hypothesis that batters are "stacked" not on the basis of a blind bias but on the basis of their particular athletic gifts. Racism as motivation for white management's failure to place Blacks in certain leadership positions, e.g., catcher does not seem to me to be a reasonable explanation of the "stacking" phenomenon. I address myself in what follows to racial makeup among major league players active in 1987 and to a statistical comparison of the performances of the Blacks and Whites for the 1987 season. Two categories only are considered, Blacks and Whites, although some information on Latins may be deduced from the data.

EVALUATION OF PLAYER'S RACE.

The race of a player is obtained by my own evaluation based on a single criterion, the photographs in the 1988 "Who's Who in Baseball". Although such evaluations are subject to error in a few cases such errors are relatively rare. No player's statistics are used if his photograph was omitted in the 1988 edition. The color evaluation having been made, black players were classified as Black Latins if they were born in a Spanish-speaking country. White Latins were classified as White along with white Americans. Of the non-pitchers active in 1987, 29% were black Americans, 9% were black Latins and 62% were white (Americans plus Latins).

THE DATA.

Table 1 summarizes career information for players active through the year 1987. No attempt has been made at more sophisticated evaluations of offensive performance such as those which take account of sacrifice flies, stolen bases and being hit by pitches as contributing to run production[James, Bill, The Great American Baseball Stat Book, Villard, N. Y., 1988]. Only batting averages (BA) and slugging averages (SA) were used. In the tables, BlA and BlL are used to signify Black American and Black Latin. All averages are for the major league careers of the players. The tables are by position so that "stacking" may be assessed. When only one black American or black Latin exists at a particular position it is so noted. Batting averages by position and by race are computed by dividing the total hits by the total at bats.

TABLE 1

CAREER OFFENSIVE PERFORMANCE BY POSITION AND BY RACE

POSITION	RACE	NUMBER	BA	SA	%BLACKS
CATCHER	1BlA+BlL	4	.277	.394	7.6
	W	49	.253	.388	
1 BASE	BlA	9	.290	.462	25.0
	BlA+BlL	12	.290	.456	
	W	36	.280	.443	
2BASE	BlA	9	.269	.373	38.9
	BlA+BlL	14	.269	.378	
	W	22	.266	.370	
3BASE	BlA+1BlL	9	.263	.369	21.4
	W	33	.270	.428	
SS	BlA	7	.261	.351	44.2
	BlA+BlL	19	.259	.344	
	W	24	.261	.363	
OF	BlA	69	.276	.431	58.9
	BlA+BlL	78	.277	.434	
	W	55	.268	.419	

TABLE 2

CAREER OFFENSIVE PERFORMANCE, ALL NON-PITCHERS

RACE	BA	SA	SB/AB
BlA+BlL	.275	.417	.037
W	.268	.411	.017

For all non-pitcher positions, other than catcher, the proportion of black Americans exceeds that of the general population(the population of Blacks in the United States is approximately 13 %). Outfield positions show the greatest proportion of black Americans of any position and catcher the least. According to table 2 the single category in which Blacks are overwhelmingly superior to Whites is stolen bases per at-bat where Black/White ratios are approximately equal to two. As Table 2 shows, Blacks' performance is better than Whites only by some .007 for both batting average and slugging average for all hitters, a difference that suggests racism in player selection overall has greatly diminished since the sixties but is still present. The superiority of Blacks in stolen bases per at-bat may be viewed as an indicator of superior speed which may, in turn, help explain the domination by Blacks at outfield positions which also demand great foot speed. Table 1 provides no evidence that racism exists at second base, third base and shortstop for the offensive part of the game. The key question is why the outfield is so skewed toward black Americans and why Blacks are not represented more at catcher. A hypothesis not often made is that perhaps Black Americans have a peculiar talent suited to the demands of the outfield and that they have a talent only ordinary for the catcher position.

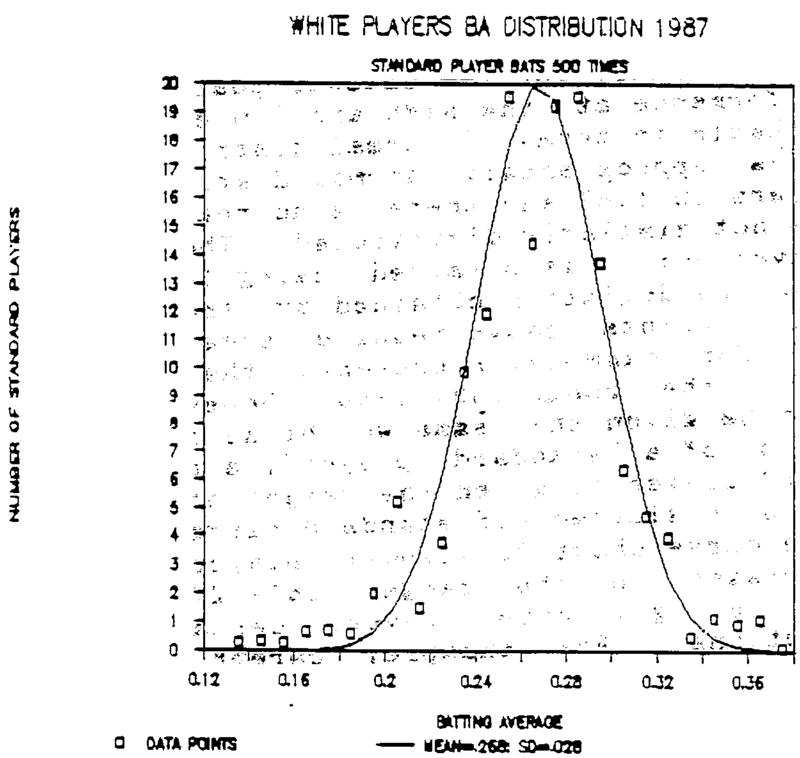
Although performance at the high and low extremes may not be accurately describable in terms of normal distributions, the figure shown demonstrates approximately normal distribution for white major league players in 1987 and there is no reason to believe that black players are not similarly distributed. The figure contains a line (normal curve) which is obtained using two parameters: the mean and the standard deviation obtained by the usual statistical methods. The data points shown require some explanation. The point at BA=.355, for example, represents the number of players whose average is in the range .351-.360. However, a player with 600 at bats cannot be given the same weight as one with 6 at bats. Hence the invention of a "standard player", a player with 500 at bats. The standard player is given the weight of 1.00 and a player with 250 at bats is considered 0.5 standard players, and so on. At the maximum of the curve about 20 standard players (10,000 at bats) have batting averages in the range .261-.270. Reference to statistical texts (e. g., Youden) reveal certain characteristics of the distributions. For example, career batting average distributions show that while 50 % of black players exceed the black average of .275, a requirement in a normal distribution, only 40 percent of the white players do. Similarly, only 46 percent of the white players exceed the mean black players' slugging average.

Another property of normal distributions is that as the distribution moves towards the tail, e. g., towards high performance, the distribution with the higher mean (Blacks) becomes increasingly populated relative to the distribution with the lower mean(Whites). Thus it is not surprising that six of the top ten in batting average and five of the top ten in slugging average are black even though only 38 percent of the players are black. Table 3 shows the top ten in each category.

TABLE 3

TOP TEN BATTERS (CAREER) BY RACE
(AT LEAST 1000 AB)

BATTING AVERAGE		SLUGGING AVERAGE	
BOGGS(W)	.354	MATTINGLY(W)	.543
MATTINGLY(W)	.331	SCHMIDT(W)	.540
BRETT(W)	.312	TARTABULL(B1L)	.520
PUCKETT(BA)	.311	GUERRERO(B1L)	.518
GUERRERO(B1L)	.310	BELL(B1L)	.514
RAINES(BA)	.308	RICE(BA)	.513
MADLOCK(BA)	.305	PHELPS(W)	.508
ROSE(W)	.303	SHEETS(W)	.508
FERNANDEZ(B1L)	.302	HORNER(W)	.508
RICE(BA)	.302	MURRAY(BA)	.502



CONCLUSIONS

The data indicate that some lingering racism exists in player selection but that it is not nearly as pervasive as it has been in the past. Stacking by position certainly exists. However, the charge that is often made that the stacking is due to racial bias on the part of those responsible for selection of players seems to me to be without foundation. There are other explanations for stacking, one of the more attractive of which is that management being profit oriented, i. e., win oriented, selects the players at all positions whose abilities are most likely to promote its ends. Without being too specific about the reasons, black athletes appear to have special abilities in particular aspects of baseball. A hypothesis that Blacks have superior short range foot-speed as part of their natural gift on a statistical basis predicts that color-blind player selection should result in disproportionately black outfielders and only proportional representation at positions not requiring great foot speed, e.g., pitcher and catcher. In view of the outstanding success of black Americans at outfield positions and in other sports-cornerbacks, running backs and wide receivers in football, all positions in basketball, sprinters and longjumpers in track- it is strongly indicated that superior speed may be the overriding determinant in positional "stacking".

REFERENCES

Bill James Baseball Abstract, 1987, Ballantine Books, N. Y., 1987

Eitzen, D. S. and G. H. Sage, Sociology of American Sport (2nd Edition), Wm. C. Brown, Dubuque, Iowa, 1978

Edwards, H., Sociology of Sport, Dorsey Press, Homewood, Ill., 1973

James, Bill, The Great American Baseball Stat Book, Villard, N. Y., 1988

Lapchick, R., Broken Promises, St. Martin's/Marek, N. Y., 1984

Scully, G. W., Discrimination: The Case of Baseball, Government and the Sports Business, R. G. Noll (ed.), Washington D. C. : The Brookings Institution, 1974

Youden, W. J., Experimentation and Measurement, National Science Teachers Association, Washington D. C., 1966

PARITY: A NEW LOOK

Robert O. Wood & Robert K. McCleery

In virtually all sports it seems as though championship teams are no longer able to repeat. Baseball is perhaps the prime example. Over the past ten years, no team has won consecutive pennants and, moreover, there have been 10 different World Series winners. Additionally, over the past seven years, in which 28 division titles have been decided, 21 different major league teams have won their division. In this article, we will attempt to address whether not being able to repeat is a new phenomenon, supply various explanations for it, and present a new way to look at issues involving parity. We will purposefully avoid the normative issue of whether parity is a desirable goal.

As a first step, we will consider the effects of introducing divisional play in 1969. 50 pennant winners were able to repeat out of 134 total (37%) before 1969, and 8 out of 38 (21%) since 1969. Perhaps of more relevance, since 1969 there have been 21 out of 72 (29%) repeating *divisional* champions. Although the percentage differences may appear to be large, from a formal statistical standpoint, they are only marginally significant (even after taking into account the changing number of teams in a league or division). The relatively short time span of the divisional era is certainly the cause. There is no doubt that should the record of the past ten years be approached in the next ten, the differences would then be highly statistically significant.

One possible explanation of parity is the inherent randomness associated with baseball. Many games over the course of a season are won or lost due to "luck"; a potential double-play groundball is just out of the reach of a fielder, a blooper falls in for a key hit, a line drive is foul by inches, a pitch is not quite in the location that the pitcher had intended, a hitter guesses a curveball correctly, etc. Therefore in evaluating a team (or an individual pitcher), one should not place too much confidence in a single season's win/loss record.

As a result of this randomness, it is typical for a team's record to vary substantially from one season to the next. Indeed, the absolute change has averaged 9.1 games over the years. To avoid spurious movements, we have omitted the war years 1942-1945 (since changes in team fortunes often were a result of inscription into military service) and the first year of each expansion. In total the data set consists of 153 different seasons covering play in both leagues since 1900. The large degree of variability alone does not explain the non-repeating phenomenon described above. However, when we look more closely at win differentials, a striking pattern emerges.

Table I exhibits the pronounced inverse relationship between what place a team finished and how many more games it wins the next year. A first place team *declines* 8.80 games, on average, the next year. On the other hand, a last place team *improves* 9.28 games, on average, the next year. Between these extremes, a strict monotonic relationship exists. Table I also details that only 30 of the 153 first place teams actually *won more* games the next year, whereas only 28 of the 153 last place teams *lost more* games the next year. [Bill James discovered this phenomenon (his "Law of Competitive Balance") in a 1976 study. He organized the data by win totals rather than by position in the league standings. Our results in this area are supplementary to his original study.]

We refer to the pattern of Table I, and exhibited in many other baseball statistics, as the "Up-Down" phenomenon. Teams that are up tend to go down, and vice versa. There are a variety of reasons why we should expect to see such a pattern. The previous year's last place teams try harder, and attempt not to duplicate the same mistakes, often by not fielding the same players, and may improve their off-season work habits. On the other hand, first place teams can become complacent and forget how much hard work it took to make it to the top, not to mention the ubiquitous banquet circuit for champions, as well as other off-season distractions.

TABLE I

Differential of team wins (last year vs. this year) compared to last year's place in league standings

	First Place	Second Place	Third Place	Fourth Place	Fourth from Last Place	Third from Last Place	Second from Last Place	Last Place
Mean	-8.80	-4.51	-2.26	-1.37	+0.78	+2.41	+3.97	+9.28
Standard deviation	9.85	9.98	9.73	10.27	11.71	10.64	10.90	10.69
Won more	30	52	67	73	83	94	90	125
Won less	123	101	86	80	70	59	63	28

A statistician will go even further to claim that the predominant factor explaining the Up-Down phenomenon may be that good luck (as well as bad) cannot persist. A team that wins a lot of games is either a very good team which received overall balanced luck, or merely a good team which received very good luck. Since there are many more good teams than very good teams, it is likely that any team which wins a lot of games does so partly as a result of good luck -- and this luck is likely to disappear the next season. The same argument applies in reverse to predict that teams with losing records likely received overall bad luck, and as their luck evens out they will likely rebound the next season.

Note that throughout the article, "luck" does not refer to what is commonly meant by the term. We are referring to the statistical property of realized small-sample probabilities converging to their underlying probabilities as more samples are drawn. For example, when a batter gets a two-out game winning hit in the bottom of the ninth, the hit is "lucky" even if it is a home run to the upper deck since before he stepped into the batter's box, he was perhaps only 3 in 10 likely to get a hit. If the situation could be replicated 1000 times, then the batter would likely get very close to 300 hits, and make the last out of the game 700 times. One season is often not sufficiently long for actual success ratios to lie near their underlying rates, in this case 3 in 10. Our term "luck" is concerned with probabilities not evening out in the short term, rather than with a team that has a win/loss record different than it "deserves" in consideration of how well it has played.

Injuries are often blamed when a team has a bad season, and especially when a championship team fails to repeat. [How many times did we hear that the 1987 Mets entire starting rotation was at one time or another on the disabled list?] Every team has its share of injuries, and bemoaning this fact is often the way a manager attempts to keep his job. In this article, we treat the timing and extent of injuries as part of the randomness which impacts every team. While it is undoubtedly true that first place teams suffer injuries, lax off-season conditioning (or other factors applicable mainly to first place teams) plays only a very minor role in becoming injured. Thus, as an explanation of parity, we lump injuries in with the other variables we have collectively labelled "luck". First place teams seem to get more injuries the next year simply because they had *fewer* than normal the year before, and this was one reason why they won the pennant to begin with.

The outlawing of the reserve clause which brought about free agency is certainly the most discussed explanation of parity. We caution the reader that sabermetric research has found that the effect of one player (even a superstar) on his team's record is grossly exaggerated by fans, general managers and owners. Perhaps of more import are the advent of the amateur and minor league drafts, huge television revenues which have effectively "levelled the playing field" for owners, extensive scouting combines, and the dilution of talent occasioned by successive expansions.

Another possible explanation of parity is the existence of a division illusion that deludes a team which wins a weak division with a mediocre record into believing that it is actually better than mediocre. If the illusion exists, it is likely that other teams in the division, not blinded (as much) by their standing, will make needed changes in the off-season and pick up games on the first-place team, thereby leading to greater parity. We test whether a winner of a poor division performs the next season similar to the truly best team in the league, as well as whether a team which finishes last in a strong division performs the next season similar to the truly worst team in the league.

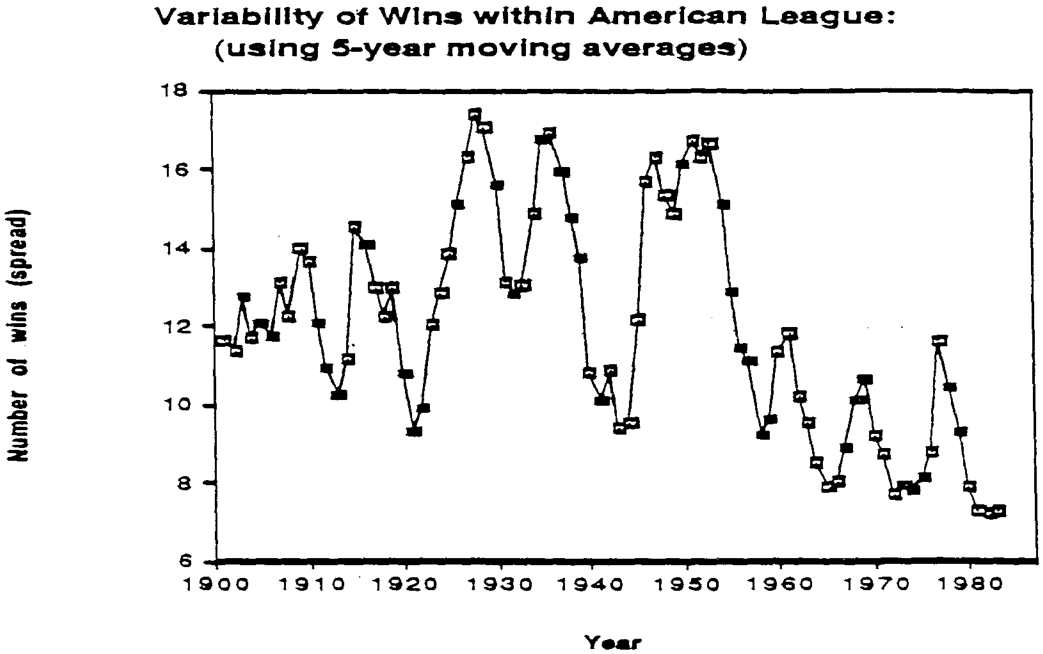
The effect that we have found pertains mainly to *winning games*, and has little to do with finishing in first place, per se. The better first place teams decline on average 11.81 games as compared to only 7.61 games for the worse first place teams. At the other extreme, the better last place teams improve on average only 2.10 games as compared to 10.35 games for the worse last place teams. Both differences are statistically significant. Thus there is scant evidence of an illusion that would cause teams to put too much stock in how they placed in their division.

A striking result emerges from the above analysis. The overall amount of luck in baseball has not changed much over the years, and if anything, has decreased. The absolute change in season-to-season wins of all teams averaged 9.28 games before 1969 and 8.56 games since the advent of divisional play. Using the introduction of expansion in 1961-1962 as the break-point, the average absolute change has diminished from 9.44 games to 8.39 games. One of the implications of parity is that over a relatively short time frame, virtually any team can win a pennant. The above result disproves the argument that today's parity is due to greater overall team movements. In fact, teams are at least as consistent as they ever have been, and are probably even moreso.

In order to investigate the hypothesis that parity is a result of greater dispersion of talent within a league as opposed to increased year-to-year randomness based upon "luck", injuries, etc., we need a measure of the true ability of a team. To this end, we calculate 5-year moving averages of team wins. In this manner, the year-to-year variability discussed above is "washed out". We convert all records to 162-game seasons for the sake of comparability over time. If parity is greater today, we would expect a smaller spread among teams' 5-year moving averages than in earlier times. Parity implies that virtually all teams average roughly 75-85 games per year, whereas years ago the Yankees could average 100 victories per year and the Senators only 60, even over a long time period. Figure I depicts the spread (technically, standard deviation) of 5-year moving averages of American League teams' win totals. For example, the spread among the 8 A.L. teams' 1901-1905 average win totals is 11.66 games, graphed as the first point (1901,11.66). The graph is similar for the National League as well.

Figure I offers a new look at the entire history of the American League. Variability of wins is a see-saw affair until a low point is reached around 1920 (corresponding to the time that the once-powerhouse Red Sox sold Babe Ruth to the second division Yankees), thus resulting in a short burst of parity. Variability reaches its all-time high (and thus parity its all-time low) in the late 1920's as Ruth and Gehrig lead the Yankees to several pennants. A degree of parity returns in the early 1930's as the Athletics and Tigers enjoy sudden success, and then just as quickly fall. The Yankees of the late-1930's retain league supremacy with the arrival of DiMaggio, and drive up the variability measure once again. World War II ushers in a new era of parity, albeit of an artificial variety, as the St. Louis Browns win their only pennant. Soon thereafter the Yankees dominate the 1950's with Mantle, Berra, and Ford. However, variability actually decreases as other teams take turns having good seasons. The pronounced downward trend in variability (and thus the upward trend in parity) from the late-1950's to the present is interrupted only briefly -- but discernibly -- by the expansions of 1961, 1969, and 1977. Parity has in the 1980's attained its all-time high point, and is likely to remain stable or even increase a bit in the future.

Figure I



Win differentials of first place teams can be used as another sign of parity. Using the 1961-1962 expansion as the dividing line between eras, first place teams declined on average 7.67 games before expansion, as compared to 11.61 games after expansion. This difference of almost 4 games is statistically significant. At the other extreme, last place teams improved on average 9.12 games before expansion, as compared to 9.68 games after expansion. Although this difference is not statistically significant, the effects are clear. The modern era has witnessed a dramatic increase in parity, which is exhibited by both first and last place teams moving back to the pack much more frequently and quickly than in earlier times.

We therefore see that expansion has conflicting effects on parity. In the short run (about 4-5 years), the ineptitude of the expansion teams increases the variability of team win totals in a league. But in the longer run, as the expansion teams are better able to compete, the effect of diluting the talent pool in the major leagues becomes predominant, and greater parity results. Of course the first factor can be considered an "investment" that must be undertaken before returns to parity are realized in the second period. It is the second factor that is most often cited when the greater degree of parity in the modern era is discussed.

In earlier times true ability in a league was more spread out across teams. Today, due to several factors, the distribution is much tighter. Teams' established win levels are now on the order of 87, 85, 83, 81, 79, etc., instead of the earlier 95, 90, 85, 80, 75, etc. Actual win totals in a given season can be viewed as a team's established win total plus a random term (positive or negative) which captures all the "luck" impacting the team over the season. Effectively, in earlier times, the leagues were divided into quality classes, with the Yankees for many years being the sole member of the top American League class. Only rarely and for a short time could a team move into an adjacent quality class. When first- and second-division clubs were discussed, everyone knew which teams were referred to. Today, since true ability across teams is so closely bunched, there is no firm quality class structure. The same degree of year-to-year randomness inherent in baseball is now sufficient to capriciously move teams up and down virtually the entire league standings each season. This phenomenon is the essence of parity.

Another possible implication of parity is its effect on the number of games required to win a pennant. It seems logical that greater parity would imply that as teams gravitate toward the pack, first place teams would be able to win fewer games than their earlier counterparts. Using win totals converted to a 162-game season for comparability over time, we test whether first place teams win the same number of games after the advent of expansion. In the division era, we call the team with the best regular season record the first place team in the league.

Let us remark at the outset that there are biases to our test. Projecting that a first place team will maintain its winning percentage if its season were extended to 162 games is tantamount to projecting that it will receive the same degree of luck in the extra games as well. From our earlier arguments, we know that winning teams have likely received good luck, and that good luck is not likely to persist. Thus we may be over-inflating win totals of pre-expansion first place teams. Seasons were 140 games from 1901-1903 and 154 games from 1904 until the first expansion in 1961-1962 after which teams played 162 games. Winning percentages would not likely be more than .050 lower in the extra games, so that this bias is at most an over-inflation of the pre-expansion average first place win total of 0.5 games.

Secondly, when we include the first few years of each expansion (when the expansion teams are inept), we may be inflating win totals of first place teams in our second sample. Another bias that is harder to get a handle upon may also reside in the post-expansion data. If one division in a league is much stronger than the other, an unbalanced schedule would cause team win totals to be more even across divisions than true ability. This effect would, in general, cause the data to erroneously exhibit a greater degree of parity, and, in particular, lower the first place team's win total. We have not attempted to quantify these biases, and leave it as a topic for future research. We take some solace in that they work in opposite directions, so hopefully cancel out to a large extent. It is unlikely that these two effects in combination bias the post-expansion average first place win total by more than 1 game in either direction.

The results are as expected. In the American League first place teams averaged 103.6 wins (on a 162-game season basis) before 1961, and only 100.0 wins after the advent of expansion. In the National League first place teams averaged 103.1 wins (on a 162-game season basis) before 1962, and only 98.6 wins after the advent of expansion. These differences are highly significant both in a formal statistical sense as well as in a baseball sense. Many pennant races are decided by fewer than the 4 games that first place teams can no longer win in the modern era. Note also that the 4-game differential easily exceeds the maximum amount of the biases built into the test discussed above.

In this article we have considered various explanations and implications of parity. In our review of its explanations, we began with win differentials from one season to the next based on what place a team finished. We presented the dramatic inverse relationship inherent in the data. First place teams almost always decline the next season, on average by nearly 9 games. Last place teams, on the other hand, almost always improve, on average by over 9 games. An explanation of this Up-Down phenomenon is based upon the statistical nature of the "luck" involved in winning any baseball game. Teams that have won a lot of games are likely to have received favorable luck over the course of a season, and since luck must even out over time, are likely to decline the next season. The argument works in reverse to predict that teams which have lost a lot of games are likely to rebound the next season.

The Up-Down phenomenon has been intensified by the greater degree of parity in today's game. Modern era first place teams decline much more in the following season than their earlier historical counterparts, and last place teams improve more quickly as well. Moreover first place teams are able to win fewer games today than in the past. However, we have also found that there is less variability in abilities of teams within today's leagues. It is this complex combination of circumstances that is at the heart of parity. We hope to have laid a proper foundation for, and at the same time prompted, further research into this relatively new area.