baseball ANALYST

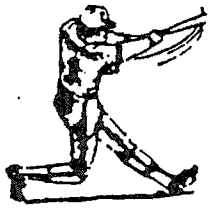
June 1984 The Journal of Sabermetrics Issue #12

BACK TO DRAWING BOARD FOR YOUNG HOPEFUL

RE-PRINT

ALTHOUGH HE LIKED HER AS A PERSON AND THOUGHT SHE HAD NICE COMPLEXION; AND WHILE HARBORING SYMPATHY FOR HER DESIRE TO BREAK INTO BASEBALL, THE KINDLY PATRON CONCLUDED THAT HER CROCHET GEOLOGICAL SURVEY OF THE NEW BALLPARK SITE WAS OVER-WORKED, AND THAT ITS 'RELEVANCE VALUE' FACTOR WAS WORTH NIL, TO THE POWER OF THE ABSOLUTE........

RAY CHEGLY IV Says:



QUIT WORK AND SLEEP ALL DAY

Live a life of alarming Luxury!

Listen, baseball is a stupid kid's game, and if you can't win millions betting on it, there's something very, very wrong with you! Why isn't everybody rich like me? Because they're all a pack of stupid jerks for not subscribing to my Super Secret Scientific Baseball Betz one page newsletter. I have mastered the baseball wager! BOOKIES DO NOT ANSWER MY CALLS--because they know that every time they talk to me they lose BIG PILES OF REAL CASH!!! I live high on a hill in Eldorado in a huge, beautiful home filled with many modern appliances. I have so many cars that they have to put innocent folks in prison to keep up with the demands for my license plates. And I have no formal education. NO COLLEGE DEGREE--NO HIGH SCHOOL DIPLOMA!!! I don't even own a calculator. BOWIE KUHN HATES ME because I have made more money at his sport than he could ever hope to.

Mr. Ray Chegly IV!! Although he doesn't look rich, he is. He is so rich he does not have to impress anyone by dressing fancy.

Why am I sharing my super secret scientific method with the general public?--Because I have TOO MUCH MONEY and want to be remembered forever as a philanthropic friend to all of mankind. What do I, a grilionare several times over, have to gain from a few thousand twice monthly checks of $20.00 (the fee for my betting service)--ABSOLUTELY NOTHING. Afterall, how many priceless classic cars can one man own before they all begin looking the same? How many seaside mansions does one really need?

Why do Nevada border lawmen KEEP ME OUT OF THEIR STATE UNDER THREAT OF ARREST AND EXECUTION? Because they know that I can bankrupt any betting parlor in the time it takes to play an Indians-Mariners doubleheader. Often I am asked, 'if you're so rich, why don't you just buy the National League?' MAYBE I ALREADY HAVE!!!(?)

Using my methods you will become so rich you will have to invent names to identify your financial staus. Think what it will feel like to have congressmen and senators over to your palace JUST TO CLEAN UP THE FLOORS!!!

Yes, more than anything on Earth I want to be a Godzillionare like Mr. Ray Chegly IV. Please rush me the first issue of his 'Baseball Betz' one page betting tipz newsletter at $20.00. If I am not completely satisfied, I can move out of my current address and change my name. Enclosed is my check and money order. Please send in all possible haste to:

NAME_____
ADDRESS_____
CITY_____ STATE____ ZIP_____
CLOSEST LIVING RELATIVE_____

945 Kentucky Street
Lawrence, KS 66044

$12.00 per annum

All submissions ready for reproduction, please

BASEBALL ANALYST

June 1984

Founder/Publisher: Bill James
Editor: Jim Baker
Business Manager: Susie McCarthy

Issue no. 12

IN THIS ISSUE:

ABOUT THE COVER: The cover this issue is another in a series of baseball works of Mike Ross of London, England.

ATTENTION CONTRIBUTORS: Please send your originals for publication as photocopies do not photocopy well. The darker your original, the better. I'll gladly return your originals if you so desire. As always, submissions are strongly encouraged.

PROJECT SCORESHEET: The project is running very well. We still need scorers for the Dodgers, Angels and Expos, as well as Pirates and Braves. Please contact me if you can help out with these teams (especially) or any other team that you are able to follow on a regular basis.

THE METRODOME AND HOMERUNS

Terry Bohn

In 1982, the Minnesota Twins moved downtown to the new Metrodome. The new stadium has been a source of controversy and criticism from a lot of people. One criticism leveled against the Metrodome is that it is a "homerun" park. Those that have witnessed games there have cited the large number of homeruns that fly out of the stadium--well, over the fences anyway. More homers have been hit in the Metrodome (338) than in any other ballpark in the American League, except for Tiger Stadium in Detroit (378).

In looking at the number of homeruns hit, it would appear that the Metrodome is a homerun park. Is it the ballpark?--Or are other factors involved? The Twins hit their share of homers and give up a passel no matter where they are playing. Over the past two seasons, since the Metrodome opened, the Twins are in the middle of the AL in homeruns hit (6th) with 289. In the same period, the Twins weak pitching has given up a league high of 371 roundtrippers. Is it the Metrodome that is the reason for the high homerun totals, or is it that the Twins happen to hit and pitch there half the time?

The method I used to determine whether or not a particular American League park enhanced or restricted the homerun was to add up the total number of homers hit in each teams games, by both teams, for the past two years. The Twins hit 289 and their opponents 271 for a grand total of 660. I divided into this total the number of these hit in the team's home park, again for both the team and their opposition. The Twins hit 137 in the Dome and their opponents 201 for a total of 338. Just over half of the total were hit in the Metrodome, .512. The Metrodome is a homerun park, but only slightly. There are five other American League parks that yield more homeruns.

RANK	PARK	TEA	TOTAL HOMER	HOME HOMER	% AT HOME
1-	Kingdome	Sea	559	326	.583
2-	Exhibition	Tor	565	317	.561
3-	Tiger Stad.	Det	675	378	.560
4-	Aneheim	Cal	594	321	.540
5-	Municipal	Cle	437	232	.531
6-	Metrodome	Min	660	338	.512
7-	Memorial	Bal	635	319	.502
8-	Fenway	Bos	591	290	.491
9-	Comisky	Chi	520	242	.465
10-	Alameda	Oak	582	268	.4605
11-	Yankee	NY	543	250	.4604
12-	Royals	KC	537	241	.449
13-	County	Mil	651	274	.421
14-	Arlington	Tex	446	187	.419

ROBERT STEWART SMITH asks:

Is Artificial Turf More Offensive?

Plastic "grass" stretched over a hard, asphalt base now forms the playing surface in six of the twelve National League parks and four of fourteen in the American League. While the best criterion for choosing a surface may indeed be Richie Allen's gastronomic one ("If a horse can't eat it, I won't play on it"), connoisseurs of quantification may legitimately wonder what the effects of artificial surfaces are on the game of baseball.

A perusal of comments by a variety of observers suggests that the bluish-green carpet makes both the view from the stands and the game itself decidedly more offensive. During the last World Series, for example, Earl Weaver and Tony Kubek agreed that "guys get more hits" on artificial turf. Columnist George Will's mighty pen even described the national pastime on polyturf as "pinball." Surely, then, the issue is not whether plastic-coated pavement converts Baltimore chops to prime rbi, but by what order of magnitude offensive statistics are inflated by inedible playing surfaces.

Method of Analysis

One problem in deciphering the effects of artificial turf is that the method chosen must take account of the widely-believed notion that teams are tailored to the park in which they most often play. If teams with artifical turf on their home fields are built around players of type X (who may be fast, say, but not exhibit much power), and teams with grass in their home stadiums are built around type Y players, comparing statistics for games played in "grass parks" with those for "turf parks" will yield biased estimates of the independent effects of the surface itself. An example will help explain the problem.

Consider a league in which half of all parks have artificial turf and the other half grass. The games in a given "turf" park will match two teams with type X players a bit less than half the time (remember, the home team cannot play itself), while slightly more than half of the games at the park will pit type X against type Y players. Thus, about three-fourths of the at-bats, say, in "turf" parks will be by X-type players. Similar reasoning implies that about three-fourths of at-bats in grass parks would be by type Y players. Because our object here is to estimate the direct effects of the playing surface on a given group of players, simple comparisons by park will clearly confound this "playing surface" effect with the indirect effects of team selection.

The methodology chosen for this study involved tracking pairs of "grass" and "turf" teams through the 1983 season, totalling and comparing their statistics on each surface. Thus, we can compare how the same two teams performed on grass and artifical surfaces.

Three teams with artifical turf on their home fields -- Kansas City in the American League and Cincinnati and St. Louis in the National League -- were selected; none played in domed stadiums (I did not wish to contend with both polyturf and air conditioning at the same time), and each played in a home park essentially rated "neutral" by Bill James in his 1982 Abstract. The three teams were followed throughout the 1983 season, and data were collected whenever they played a grass-field opponent. At the end of the season, the data for each season series were totalled; one set of totals pertained to those games between the two teams played on grass and the other related to their games on artificial turf.

The data collected were the normal box-score items: singles, doubles,

triples, home-runs, runs scored, errors, stolen bases, sacrifice bunts, double plays, and half-innings played. The data were compiled on a per-game basis (totalling the hits, runs, etc. for each of the two teams). What emerged was a description of the average totals per full 9-inning game between the same two teams when they played on grass and when they played on artificial turf.

The Results

Table I contains simple counts of how many of the teams in the 22 season series in my sample had larger numbers of hits, runs, stolen bases, and so on (per 9-inning game) when they played on artificial turf. While there is some tendency for offensive statistics (except for home runs) to be greater on artificial turf, the most striking difference is with stolen bases. In 19 of the 22 series, stolen bases were greater on turf than on grass. The only other differences where random factors inherent in any sample can be plausibly ruled out as an explanation are the greater incidence of doubles on turf and -- probably because of more doubles and stolen bases -- the reduced incidence of double plays. Interestingly, however, total hits, total bases, and runs scored are dominant on turf only about half the time (which is what one would expect if the surfaces were equal with regard to these factors).

Table II, which contains data on the season series averages (by league) for both types of playing surfaces, helps illuminate the reasons underlying the apparent lack of differences. In Table II, the only statistics for which there are significant differences (ones that cannot be plausibly attributed to chance) in both leagues are stolen bases and triples. The

Table I

Number of Season Series
in Which Statistics Listed
Were Greater on Artificial
Turf Than Grass

(Total Season Series = 22)

Statistic (Per Full 9-Inning Game)	Of 22 Series, Statistic for Artificial Surface Greater in the Following Number of Cases
Singles	12
Doubles	15*
Triples	13
Home Runs	8
Total Hits	14
Total Bases	12
Runs	12
Stolen Bases	19**
Sacrifice Bunts	11
Errors	11
Double Plays	7*

** A difference this large would occur randomly between surfaces of equal properties less than 1% of the time.

* A difference this large would occur randomly between surfaces of equal properties 10% of the time or less.

Table II

Season Series Averages Per
Full 9-Inning Game, By
Type of Playing Surface

	Grass	Artificial Turf	Difference (% Difference) Turf-Grass
American League (10 Series)			
Singles	13.5	13.5	0
Doubles	3.0	3.3	+.3(+10%)
Triples	.5 (b)	.8	+.3(+60%)
Home Runs	1.5	1.4	-.1(-7%)
Total Hits	18.5	19.0	+.5(+3%)
Total Bases	27.0	28.1	+1.1(+4%)
Steals	1.4 (a)	2.1	+.7(+50%)
Sacrifice Bunts	.6	.4	-.2(-33%)
Errors	1.5 (b)	1.9	+.4(+27%)
Double Plays	2.2	1.9	-.3(-14%)
Runs	8.7	9.2	+.5(+6%)
National League (12 Series)			
Singles	13.2 (a)	11.9	-1.3(-10%)
Doubles	2.5 (a)	3.3	+.8(+32%)
Triples	.4 (a)	.8	+.4(+100%)
Home Runs	1.6 (b)	1.4	-.2(-13%)
Total Hits	17.7	17.4	-.3(-2%)
Total Bases	25.8	26.5	+.7(+3%)
Steals	1.7 (a)	2.3	+.6(+35%)
Sacrifice Bunts	.9	1.0	+.1(+11%)
Errors	1.7	1.5	-.2(-12%)
Double Plays	1.9	1.7	-.2(-11%)
Runs	8.7	9.0	+.3(+3%)

(a) This difference would occur randomly between equals 5% of the time or less.

(b) This difference would occur randomly between equals 10% of the time or less.

average number of doubles per game is greater in both leagues on artificial turf, while the average number of home runs is less; however, only in the National League sample were these differences statistically significant. Runs and total bases tend to be greater on artificial surfaces, but only by 3-5%. Interestingly, the number of singles is, if anything, less on turf than on grass. Consistent with Table I, there are fewer double plays on turf, but the difference is not statistically significant in either league.

What emerges from both tables is the conclusion that doubles, triples and stolen bases are increased by artificial turf at the expense of singles and home runs -- with the net result that hits, total bases, and runs are virtually the same on grass and turf. These results suggest several hypotheses. First, from the strong evidence concerning stolen bases one can infer that players are able to run faster on artificial surfaces. The hard asphalt base, which players claim wear down their legs by season's end, has less "give" than the dirt running lanes of natural surfaces. The enhancement of footspeed is probably also a factor contributing to the increased number of doubles and triples on "turf."

Second, the increased number of doubles and triples on artifical surfaces could also be caused by line drives getting between the outfielders because of the faster bounce afforded by the harder surface. This factor might convert some singles into doubles, which could account for the finding that the number of singles on artificial turf is no greater -- and perhaps less -- than the number on grass.

The advantages of line drives on turf may induce hitters to alter their swings and cause the third phenomenon to be explained: the reduced number of home runs in artificial surface parks. Given that lofted flies to the outfield have a greater probability of being caught in "turf" parks (remember: the outfielders can also run faster on the hard surface) and that line drives are more likely to get through the gaps, one suspects

that those hitters who have the ability to go for either home runs or line drives tend more often to go for the latter when playing in parks with artificial surfaces. (I don't think any other factor can account for the reduced numbers of home runs. There were no systematic grass/turf differences in playing field dimensions -- and while all of my artificial turf parks had walls over 8 feet high, the differences in home runs persisted when my sample of "grass" teams was restricted to those with home-park walls over eight feet high in all fields.)

It would appear, then, that the modern innovation of playing Abner Doubleday's game on a plastic carpet has not produced a noticeable offensive bias. Ironically, it has probably helped make the game a bit more old-fashioned. Consider, for example, the element of baseball most clearly affected by artificial turf: stolen bases. In the years prior to 1914, stolen bases rarely averaged less than 1.7 per game in either league, but after that the stolen base declined in frequency -- so that from the 1930's on stolen bases in both leagues averaged less than one per game (usually they were in the .5 to .7 range). By the late 1970's however, stolen bases per game were again in the pre-1920 range: 1.6 per game in the National League and 1.3 in the American. Given the 35-50% increases in steals on artificial turf, and given that more than half of all parks still have grass surfaces, not all of this doubled incidence of stolen bases can be attributed to the direct effects of polyturf. However, it is tantalizing to consider the great likelihood that plastic turf has nudged the game -- at least in some ways -- toward the style common in the days of Ty Cobb. But somehow, to call a latter-day Cobb the "Polyurethane Peach" just doesn't sound right!

THE PYTHAGORIAN THEOREM AND TWENTY-TWO RECENT MANAGERS

By David F. Hoppes

The baseball-version of the Pythagorian Theorem, as described in Bill James' 1982 Baseball Abstract, is an excellent tool for correlating a team's winning rate to its total runs scored (R) and its opponents' total runs scored (OR):

$$\text{Fraction of games won} = \frac{R^2}{R^2 + OR^2}$$

The correlation is slightly better when the exponent 1.83 is used vice 2, and this 1.83 value was used in the results reported here.

I have calculated that for all teams from 1920 through 1983, the average difference between the wins calculated by the Pythagorian Theorem and the team's actual wins is only 3.24 games per season. Surely no other correlation is both so simple and yet so accurate.

This report shows how the teams of twenty-two recent managers did when their team's winsare compared the wins calculated by the Theorem. The best that any of these prominent managers could do was to "beat" the Theorem by an average of 2.22 games a year; and this honor belongs to Earl Weaver. At the other end of the list, six managers won less games than would have been predicted by the Theorem, with Tommy LaSorda's teams falling almost one game a season behind the predicted level.

How important are these results? They do not address the question of whether a team's performance level is better or worse than should be expected from the team's collective talents. Rather, the Theorem indicates whether a team made the most of the runs it scored and its run defense. A positive difference against the Theorem would probably indicate that a team won more close games than it lost, but it could also indicate that the team lost more one-sided games than it won.

The summary results for the Twenty-two managers are listed below. These results include all seasons for which the manager directed the team for at least half of its games.

SUMMARY RESULTS

MANAGER	NO. SEASONS	AVG. GAMES PER YEAR ABOVE OR BELOW THEOREM PREDICTION
Earl Weaver	15	+ 2.22
John McNamara	9	+ 2.09
Billy Martin	14	+ 1.78
Don Zimmer	7	+ 1.76
Sparky Anderson	14	+ 1.71
Walt Alston	23	+ 1.37
Chuck Tanner	13	+ 1.19
Ralph Houk	19	+ 1.13
Whitey Herzog	8	+ 0.88
Gil Hodges	9	+ 0.74
Al Lopez	18	+ 0.60
Dave Bristol	9	+ 0.40
Casey Stengel	25	+ 0.17
Dick Williams	16	+ 0.15
Danny Murtaugh	12	+ 0.05
Red Schoendienst	12	+ 0.03
Bill Virdon	12	- 0.24
Lou Boudreau	16	- 0.35
Al Dark	13	- 0.35
Gene Mauch	22	- 0.57
Leo Durocher	24	- 0.67
Tommy LaSorda	7	- 0.95

Listed below are the seasons in which these twenty two managers had the best records and the poorest records vis-a-vis the Theorem.

MOST WINS OVER THAT PREDICTED BY THEOREM

Manager	Year	Team	Actual Wins	Predicted Wins	Difference
Alston	1954	BKN	92	80.53	+11.47
Anderson	1970	CIN	102	90.54	+11.46
Stengel	1943	BOS (NL)	68	57.7	+10.3
Boudreau	1955	KC	63	62.8	+10.2
Weaver	1977	BAL	97	87.6	+ 9.4
McNamara	1981	CIN	66	56.6	+ 9.4

MOST WINS UNDER THAT PREDICTED BY THEOREM

Manager	Year	Team	Actual Wins	Predicted Wins	Difference
Weaver	1972	BAL	80	90.12	-10.12
Stengel	1962	NY (NL)	40	50.09	-10.09
Durocher	1970	CHI (NL)	84	93.6	- 9.6
Houk	1966	NY (AL)	70	79.4	- 9.4
Durocher	1953	NY (NL)	70	79.0	- 9.0
Mauch	1961	PHI	47	55.7	- 8.7

By far the most suprising single season is Baltimore - 1972. Earl Weaver, the modern time champion at outperforming the Theorem, came in 10 games below the predicted 1972 win total. If this one season were deleted from Earl's overall record, he would have averaged over three games a season ahead of the Theorem predition, a full one game a season ahead of the next highest manager, John McNamara. Earl's 1972 Orioles had a mediocre-to-poor offense; they scored 519 runs versus a league average of 537, and batted .229 versus a league average of .239. But the Oriole pitching and defense were outstanding, holding opponents to 430 runs.

In the strike-shortened 1981 season two teams performed outstandingly vis-a vis the Theorem. Weaver's Orioles were 7.5 games ahead of the predicted level while McNamara's Reds were 9.4 games ahead. If these results were extrapolated to 162 game seasons the Orioles would have been +11.6 games and the Reds would have been +14.1 games.

More on this subject, including results of earlier eras, is included in Pete Palmer's excellent article in the 1982 edition of The National Pastime.

FOUR-DECADE CANDIDATES FOR 1990

by Daniel Greenia

Active players born after 7/1/47 who played in the 60's:

	Birthdate	First Team	Current Team
Oscar Gamble	12/20/49	Cubs	Yankees
Bill Buckner	12/14/49	Dodgers	Cubs
Rick Dempsey	9/13/49	Twins	Orioles
Ted Simmons	8/9/49	Cardinals	Brewers
Jerry Reuss	6/19/49	Cardinals	Dodgers
Dusty Baker	6/15/49	Braves	Giants
Steve Garvey	12/22/48	Dodgers	Padres
George Foster	12/1/48	Giants	Mets
Toby Harrah	10/26/48	Senators	Yankees
Bill Russell	10/21/48	Dodgers	Dodgers
Mike Jorgensen	8/16/48	Mets	Braves
Carlton Fisk	12/26/47	Red Sox	White Sox
Richie Hebner	11/26/47	Pirates	Cubs
Gene Garber	11/13/47	Pirates	Braves
Darrell Evans	5/26/47	Braves	Tigers
Amos Otis	4/26/47	Mets	Pirates
Nolan Ryan	1/31/47	Mets	Astros
Al Oliver	10/14/46	Pirates	Giants
Mike Torrez	8/28/46	Cardinals	Mets
Rollie Fingers	8/25/46	A's	Brewers
Hal McRae	7/10/46	Reds	Royals

When games end and what that tells us, or
Empirical evidence about common assumptions in baseball models
by Clem Comly

Introduction

Sabermetrics provides facts to help answer old questions and encourage new ones. One popular type of question is what kind of team would a specific group of players make. An important tool in working on this type of question is the mathematical model. A model is only as useful as its assumptions are accurate. All of the 1982 season box scores were reviewed to compare three assumptions with actual data. This paper will review what a model is, out line a common baseball model, present the data from the box scores, and evaluate the three assumptions.

Models in general

A model is a group of mathematical relationships (equations) that will give you a result if you give it first a specific set of facts. A model is useful if it can supply you with an answer as good as you could come up with without it and if it is easier to gather the specific facts it needs than to get the answer directly. Another neat thing about models is that you can put them together. The answer you get from model A can be one of the specific facts you put into model B to get the answer you really want.

For example, how many more wins would the Phils have had in 1983 with Matuszek at 1B all year instead of Rose? We could just think and think and make our best guess, or we can make a model.

Answer=Wins with Matuszek (M.) - 1983 Phillies wins
Wins with M.=162x winning percentage with M.

We know how many games the '83 Phils won. Now all we need to guess is the winning % with M, still not to easy. Let's add a form of the Bill James equation (see Baseball Abstract):

winning %=runs squared divided by the sum of runs squared and opposition runs squared
$=R^2/(R^2+OR^2)$

So now we can use the model if we supply the R and OR for the Phils with Matuszek. If we call Rose and Matuszek equal deffensively, then OR with M. should be the same as the Phils' '83 OR. R with M. should equal '83 Phils' runs +runs M. is responsible for - runs Rose was responsible for. Now, all we have to do is estimate the runs Rose and M. are responsible for. Let's estimate runs responsible for with Bill James runs created formula RC=f(H,BB,SB,AB,TB) (read this as runs created is a function of hits, walks,steals,at bats, and total bases- while the equation is not spelled out in this form the model contains a concrete equation that requires a number for each item in the parentheses). Now the model is:

winning $\%=R^2/(R^2+OR^2)$
R='83 Phils runs +RC for M. for all season - RC for Rose
RC for M.=f(Matuszek's stats for whole season)
RC forRose = f(Rose's stats for 1983)
Wins with M.=162 x winning % with M.
Answer= wins with M. - '83 Phils' wins

This model is easier if it is easier for us to provide an estimate of Matuszek's statistics if he had played for all of 1983 than to pull how many wins difference there would have been directly out of the air. But keep in mind the model will provide answers only as accurate as the facts that are provided and the truth of the relations in the model.

Baseball Model Outline

winning $\%=R^2/(R^2+OR^2)$
R=f(team hits,at bats, walks, steals,total bases,etc.)
team stat= sum of the products of [illegible]er's plate app. and stat per P
OR=f(ER,errors)
ER=sum of 1/9 of product of pitchers' IP and ERA

team IP = sum of pitchers' IP
outs by offense = outs by defense = 3 x team IP
team plate appearances (PA)=f(outs by offense, team offensive stats)
hitter PA = team PA/9 + 90 - 18 x # of hitter's batting order slot

This model will give you a winning percentage for a collection of hitters and pitchers. The three functions not explicitly given in the model are not directly involved in the rest of the article, but interested readers can find several equations to use for these functions in other sabermetric literature or they can contact me.

To get an answer from the model, the specific facts it needs are:

the ERA of each pitcher,
the IP each pitcher will pitch (sum must = team IP),
the batting order slot for each of nine hitters, and
the offensive statistics per plate appearance for each hitter.

Let me answer three questions you might be thinking now. First, why worry about the batting order slot of the hitter? A higher batting order spot gives a hitter more plate appearances. Putting better hitters towards the top of the lineup should improve run production by shifting plate appearances from poorer hitters to the better hitters. If this model is to represent what would happen when these players were brought together, this shift of plate appearances is realistic, but how do you guess what the shift should be? The last equation in the model makes this shift of plate appearances much easier by replacing the question of how many more PAs for this hitter than the next one with put the nine hitters into a batting order.

Second, why can't I have more than nine hitters when I can have as many pitchers as I want? The answer is because the model uses a lineup format for the hitters, and a lineup only has nine slots, there are only nine hitters. But you can get around this limit. If you assume more than one player shares a spot in the batting order on a percentage basis, you can put all these players in the model by creating the stats for the model's "hitter" in that slot from the stats of all the sharing players. For each stat, calculate a weighted average from the players'stats and their percentage of the playing time.

Third, don't the equations in the model run in a circle and never give an answer? Yes, it appears the following part of the model would have you go round and round:

team PA=f(outs by offense, team offensive stats)
player PA =f(team PA)
team stat=f(players PA)

But this type of model, for reasons I won't go into, can be solved with a bit of sweat or a handy computer. You guess the team PA and solve the equations for players PA and team stats. Then use the team stats to get a new number for the team PA. If the new number is the same as the old one (or very close), then all's well. If not, then repeat the process, but use the new team PA instead of your guess. Each successive new team PA will come closer to the final answer. When the most recent team PA is basically the same as the previous one, you are done.

Discussion

One question is how many more plate appearances will a player get by moving up one spot in the batting order. In Percentage Baseball (1964), Earnshaw Cook quotes Bobby Bragan that one spot higher was worth 17 more plate appearances in a 154-game season. Cook notes that 17=154/9 and theorizes that the final batter's batting order spot is random. Thus, for 162 games, each spot would

come to the plate 18 times more than the following spot. I worried about this because the last batter isn't truly random, but a form of distribution related to binomial that may need more than 20-odd plate appearances to smooth out enough to appear random. Also, even if the chance to be the last batter was random, the hitters with the highest on-base percentage should dodge the bullet more often and push the last plate appearance to the poorer hitters.

The data was converted to team PA game by game and divided by9. The remainder of the quotient is the slot in the batting order of the last batter (except if the remainder was 0, the batter was batting 9th). The distribution by league was:

	1	2	3	4	5	6	7	8	9	Ave.	
National	209	221	201	205	207	208	218	223	252	216	
American	245	255	220	237	248	256	235	269	305	252.2	(note that the 2
NL-Ave.	-7	5	-15	-11	-9	-8	2	7	36	0	extra gameswera
AL-Ave.	-7	3	-32	-15	-4	4	-17	17	53	0.2	a Balt.-Mil. tie)

There is a trend towards the on-base % theory, but on the league basis it doesn't seem very large. The largest difference is the # 9 slot in the AL and that's only about 4 plate appearances for each team. Below are the teams with the highest and lowest number of game-ending plate appearances in the first four batting order slots.

Excess plate appearances by slot for season vs.#9 slot

	1	2	3	4	5	6	7	8
Atlanta	139	118	93	75	56	49	35	22
Phila.	138	122	110	92	81	63	42	22
Assumption	144	126	108	90	72	54	36	18
Atl.-Assm.	-5	-8	-15	-15	-16	-5	-1	4
Phi.-Assm.	-6	-4	2	2	9	9	6	4

How many runs difference does this make? Let's look at Atlanta. The next table will have total plate appearances by batting order in 1982 and the predicted PAs for the same team PA total. Then we'll take the difference and using runs created/PA for the Atlanta

CHART ONE: 1982 American League distribution of sum of runs plus left on base for 27- and 24- outs--in that order.

Runs+LOB	Bal	Bos	Cle	Det	Mil	NY	Tor	Cal	CWS	KC	Min	Oak	Sea	Tex
1	-	-	-	-	-	-	-	-	-	-	1-	-	-	-
2	1-	1-	1-	-	-	2-	-	-	-	-		-	1-	2-
3	1-	3-	1-	-	1-	1-	2-	1-	-	-	2-	1-	0-1	1-
4	1-1	2-	2-	4-	3-	3-	6-	-1	4-	2-	1-	7-	1-	6-
5	2-1	3-	5-	4-2	2-	3-	3-1	3-1	5-	5-	6-2	5-1	11-	7-1
6	7-	2-	7-	6-	5-	9- 1	9-2	6-	3-	10-	6-	10-1	8-	9-1
7	2-1	4- 3	6-2	10-	8-3	5-1	8-4	5-2	13-1	9-1	6-2	9-4	7-3	9-3
8	10-3	12- 2	11-2	6-3	9-4	10-2	9-1	10-3	4-4	6-4	11-4	7-1	14-1	9-1
9	9-3	15- 1	9-5	14-6	11-5	15-0	10-4	8-4	9-2	18-4	16-4	15-3	10-2	13-2
10	7-4	16- 3	7-	14-4	4-1	11-8	12-1	8-2	10-5	10-2	17-2	8-5	7-5	13-4
11	12-9	7- 3	8-1	9-3	8-3	6-5	15-3	7-2	9-2	12-3	12-2	11-5	8-3	12-3
12	6-5	11- 2	10-6	9- 1	14-2	7-2	8-6	10-8	12-1	5-6	10-3	12-2	11-6	7-3
13	3-3	7- 5	5-2	6-5	6-5	7-5	11-5	13-2	10-6	3-2	5-4	8-1	9-3	6-
14	12-2	6- 1	8-3	9-6	5-2	3-3	8-2	6-5	7-8	4-7	6-3	6-1	4-3	5-2
15	12-3	6- 4	7- 4	3-2	1-2	5-1	3-1	4-4	2-4	4-3	5-1	4-1	2	6-4
16	5-3	3- 3	8-1	3-3	4-5	4-2	1-1	5-4	5-2	4-4	2-2	3-	2-1	1-2
17	4-6	1- 1	2-1	3-2	5-3	5-	1-1	3-5	5-4	2-3	1-2	3-1	3-1	4-
18	5-1	1- 4	4-1	3-2	-4	3-3	3-1	1-	4-	2	2-	3-1	2-1	3-1
19	--	2- 4	-	3-1	2-2	1-1	-	1-2		3-1	2-2	1-3	1-	
20	1-	-2	3-1	-	4-5	2-	1-	1-	2-2	2	-1	1-	1-	1-

1982 AL games that don't fit grid above, in form opposition's team innings pitched x 3-sum of R+L

Bal	Bos	Cle	Det	Mil	NY	Tor	Cal	CWS	KC	Min	Oak	Sea	Tex	Additional g	
39-11	30-16	29-17	30-9	30-7	28-10	25-11	36-6	29-15	31-18	30-5	18-5	26-6	25-10	Sea	NY
29-18	33-11	39-8	30-10	30-14	25-18	34-14	25-11	38-17	26-14	39-13	29-17	25-12	30-10	33-9	36-16
39-10	33-12	33-14	52-13	30-12	21-1	30-8	25-14	36-13	30-17	30-10	39-21	33-9	36-12	42-12	36-12
30-14	30-10	30-8	33-17	33-12	36-16	29-12	33-11	33-11	36-8	36-14	47-19	30-12	29-8	31-15	25-13
30-13	36-8	21-14	32-16	32-10	33-11	29-13	48-19	26-16	31-18	30-12	36-14	30-14	32-14	26-6	36-8
33-20	33-16	54-17	28-14	30-8	36-10	26-8	58-23	20-7	30-15	48-22	48-14	26-20	26-11	26-10	33-8
48-10	36-13	25-10	33-7	30-10	21-7	36-18	28-13	42-11	34-9	33-10	48-18	33-14	36-20	36-17	Mil
33-13	35-16	30-19	33-14	30-20	39-11	33-12	29-10	33-22	30-15	33-9	32-11	36-17	30-12	33-9	36-13
36-18	35-17	33-17	25-13	30-3	30-20	30-8	21-6	30-15	36-10	30-13	36-8	30-13	25-13	36-12	30-12
25-14	15-15	42-13	36-11	30-8	39-20	39-20	39-15	30-12	33-15	30-16	30-12	25-11	25-5	Cal	33-12
25-18	33-23	36-14	36-11	33-11	30-8	25-11	35-17	26-10	25-10	33-17	33-11	30-7	26-11	27-21	33-11
30-17	18-11	25-13	30-10	30-14	18-12	36-14	28-13	36-21	28-17	30-10	25-13	60-17	33-10	27-21	Clev
30-14	42-12	21-7	30-11	33-9	46-18	33-11	33-11	26-12	26-12	30-11	31-8	48-24	42-9	27-21	26-10
36-11	31-13	42-21	35-15	33-10	33-12	29-16	30-11	33-10	26-6	29-15	29-10	26-19	36-13	27-21	26-12
29-15	33-16	40-22	24-22	36-16	49-16	34-13	34-15	27-21	27-25	27-24	39-16	36-13	31-9	Clev	31-1
24-21	36-9	27-22	27-26	27-21	27-21	24-21	27-24	27-24	27-25	24-21	27-22	42-17	27-21	36-17	30-1
24-22	29-16	27-24		27-22	27-24	29-10	27-25	24-22	27-27	24-21	27-23	33-14	24-21	30-14	32-12
24-23	27-22	24-21		27-23	27-24	30-7	38-20		24-22		39-9	30-15	25-7	26-12	
	24-22	24-22		27-24	24-21	32-17	30-21		24-23			25-11	25-15		
	24-22	24-22		27-27	24-21		27-21		24-24			33-15	30-14		
	24-23	51-13		24-23	36-18		27-21		24-27			36-10	30-12		
	24-24														

1982 National League distribution of sum of runs plus left on base by team for 27- and 24-out games in that order.

Runs + LOB	Chi	Mon	NY	Phi	Pit	StL	Atl	Cin	Hou	LA	SD	SF
1				1-				1-	1-			1-
2	2-		1-	1-		1-	1-	2-	2-	1-	1-	
3	1-	1-	1-	6-	-1	1-	1-	2-	1-		2-	3-1
4	1-	7-	5-	1-	4-	3-	7-	4-1	3-	4-	8-1	4-
5	9-1	2-2	3-	5-	6-	5-	2-	7-	2-	4-	5-	2-1
6	8-2	9-1	7-	5-2	7-	5-	2-1	9-	8-2	4-1	6-1	5-1
7	6-1	11-	10-1	9-3	10-1	10-4	5-2	11-1	8-3	11-2	10-3	11-0
8	12-5	6-2	14-2	10-4	9-1	6-3	13-	10-1	13-4	9-4	8-2	12-4
9	10-3	7-2	16-3	12-5	14-2	9-2	8-2	10-4	12-5	6-3	7-4	9-2
10	13-1	10-4	7-3	5-4	9-	10-6	11-1	14-1	15-4	8-2	12-3	10-5
11	15-4	13-1	13-2	6-7	12-3	10-2	16-5	11-1	10-7	9-	14-2	10-4
12	8-3	13-6	10-3	6-7	11-10	11-3	10-4	10-10	10-3	7-6	6-7	7-3
13	9-4	4-3	10-4	4-5	4-1	8-4	5-5	6-1	7-3	12-3	5-3	10-2
14	7-2	7-1	5-2	6-3	7-5	4-7	4-2	7-3	3-2	7-6	8-	3-3
15	10-	2-5	5-2	5-	1-2	5-2	5-2	4-3	5-1	3-1	6-4	2-5
16	2-2	4-	4-2	5-1	2-1	2-1	2-1	1-1	4-1	6-1	4-1	5-2
17	-1	1-3	5-	3-1	1-3	3-1	3-	2-1	2-	3-	1-2	5-
18	-1	5-1	1-4	3-	2-2	-1	4-1	2-	1-	1-1	1-1	1-
19	1-2	2-1		2-	2-	-2	3-2	-1	-1	3-	1-	2-
20	1-1			-1	1-1		1-1			2-3		2-3

1982 NL games that don't fit grid above, in form opposition's team innings pitched times 3-sum of runs + LOB

Chi	Mon	NY	Phi	Pit	StL	Atl	Cin	Hou	LA	SD	SF
36-11	25-15	39-12	39-20	35-20	26-19	30-15	30-16	32-14	26-14	36-15	30-15
26-9	31-16	45-13	33-16	30-13	31-20	26-13	30-10	30-11	33-13	45-16	28-14
30-11	30-10	36-18	30-6	45-16	30-12	36-7	30-10	36-8	36-12	30-10	39-22
63-16	30-11	26-10	26-13	37-20	30-19	29-14	45-11	36-13	30-15	33-13	33-12
29-16	36-14	26-10	28-19	26-19	30-18	30-15	45-13	30-2	39-9	28-13	30-12
29-13	29-13	36-19	36-15	28-14	29-15	26-8	21-9	26-15	33-7	30-13	45-18
39-15	33-14	36-9	44-15	25-7	33-10	30-16	33-8	30-15	33-13	33-15	33-16
42-20	33-8	25-9	21-7	36-15	36-14	30-11	39-18	30-7	36-14	39-19	31-19
30-19	30-15	30-14	33-15	33-11	31-15	26-16	30-9	25-13	30-18	45-16	26-18
38-18	30-17	30-5	33-9	30-16	30-10	36-18	30-8	30-14	30-5	28-12	35-20
30-11	30-16	28-9	26-7	33-12	30-10	26-11	33-12	33-8	29-11	28-9	45-16
31-22	33-8	30-4	26-12	30-13	33-11	42-9	42-11	28-8	47-21	33-11	30-14
27-21	39-9	33-12	30-7	33-12	30-6	31-16	36-14	30-15	30-13	48-20	33-15
24-23	29-13	30-5	25-12	25-15	30-15	25-11	30-12	30-9	33-12	30-15	26-10
	31-16	30-9	33-9	30-12	30-9	30-12	39-17	32-24	25-16	45-15	39-16
	26-19	42-16	33-11	30-12	33-12	28-10	39-6	33-6	33-9	26-10	30-11
	30-9	30-7	33-19	51-17	51-26	30-6	32-14	45-17	29-12	28-10	25-19
	30-12		45-15	15-6	14-11	30-10	40-19	31-16	29-17	30-13	34-15
	42-10		33-15	40-18	25-14	33-14	42-12	29-16	32-18	33-14	36-12
	26-6		36-21	32-18	35-18	36-12	24-21		36-13	33-17	28-14
	28-14		26-5	25-14	39-18	30-10			63-23	37-9	26-14
	33-19		24-21	30-12	30-5	30-8			42-18	24-21	27-22
	27-21		27-24	26-6	26-6	29-11			39-14	24-21	
	27-22		27-25	27-21	42-17	25-8			28-8		
	27-23			27-22	27-21	36-29			30-17		
	27-30			27-22	27-23	28-20			27-21		
				27-23		27-0			24-21		
						24-21			27-22		
						27-22			27-22		
						24-24					

regulars in a common batting order, we'll get an idea of the error caused by applying this assumption to what is perhaps the team furthest from it.

	1	2	3	4	5	6	7	8	9	Total
Real PA	766	745	720	702	683	676	662	649	627	6230
Assump PA	764	746	728	710	692	674	656	638	620	6228
Diff in PA	2	-1	-8	-8	-9	2	6	11	7	2
Est. RC/PA	.12	.13	.16	.14	.13	.11	.10	.08	.06	
Est .Runs	.24	-.13	-1.3	-1.1	-1.2	.22	.6	.88	.42	-1.6

It appears that the assumption is close to reality, and that the error it introduces is minutein even the worst cases.

Pitcher-Pinch hitter mix

The second question is how often is the ninth spot in the batting order filled with a pinch hitter in stead of a pitcher? I found how often for each NL team the # 9 slot bats by taking the sum of the excess plate appearances for slots 1-8 and subtracting that from the team PA, then dividing the difference by 9.(Or using the 18 at-bats/slot, #9 PA=(team PA-648)/9) I also compiled by team the plate appearances for all pitchers. Voila:

	Chi	Mon	NY	Phi	Pit	StL	Atl	Cin	Hou	LA	SD	SF
#9 PA	609	623	605	604	626	615	627	613	595	634	614	616
Pitcher PA	380	454	406	432	452	438	428	408	433	437	446	429
Difference	229	169	199	172	174	177	199	205	162	197	168	187
Pinch AtBat	269	198	243	189	231	202	220	251	179	270	182	227

Pitchers batted an average 70% ofthe time in the # 9 slot. Chicago had the lowest pitcher % in the # 9 slot, 62%, and was the only team under 67%. Montreal, San Diego, and Houston were high at 73%. Obviously, a significant portion of pinch hitters don't bat for pitchers, but without pinch plate appearances instead of pinch at bats we can't be very exact. If all pinch PAs end in at bats, then 77% of all pinch PAs are in the #9 slot. If pinch PAs are as likely to be at bats as the average hitter, then 69% are in the #9 slot.

Better means less

A Phillie in the early'60s once said he enjoyed being on the road because the Phils always got to bat one more time than the home team. Normally in a model you simplify real life by saying the pitchers have to get as many outs as the hitters make. But as the Phillie knew, poor teams get to make more outs than their pitching has to get. I can see how much the influence is now because I have innings pitched against (IPA) for each team as well as the normal IP. Using the 26 points of data to get a line that best fit IPA-IP versus wins (in 162 games), the relationship was (with a least squares r^2=.94):

IPA-IP=78-.96 x Wins

How big is this? A 100-game winner 's pitchers will pitch 18 more innings than its batters will hit. A 60-game winner will bat in 20 more innings. How big a difference is 18-20 innings? That's eight to 10 runs, one more victory or loss.

Please send comments for publication to the Baseball Analyst, other comments to Clem Comly

308 Colonial Dr.
Wallingford, PA 19086

DAN RAPPOPORT

Criteria for Hall of Fame Selection*

To be elected into the Baseball Hall of Fame, a player has to meet certain criteria. He must have participated in the big leagues for ten full seasons. Beyond that, there are no objective criteria for selection. The remaining basis for selection is subjective. Each sportswriter involved is free to produce his own method of evaluating the baseball veterans. This frequently produces arguements among the fans. "Why did he get elected - all he could do was hit homers?" "Why hasn't he been elected - he led the league in this, that, these, and those?"

There needs to be some objective method for comparing player performance over a number of years. Any method put forth will not be fool-proof. There are qualities about ballplayers that are not quantifiable and must be considered in voting for the Hall of Fame, i. e., team leadership. But by using the right objective method, there would be no possible way for that method alone to show any bias among the eligible sportswriters voting.

The methodology to be used here only treats one objective indicator of ballplayer performance, RBI. This is done for ease of explanation. Any other objective criterion could be examined by the same method, whether a batter or pitcher is being judged. Even RBI is not entirely objective. It is meant to be a measurement of clutch hitting but some hitters with less RBI could have higher percentages for advancing runners and/or batting in runners from third base. It is safe to assume that if the sample of RBI data is large enough, then the more RBI a player has in any season, the better a clutch hitter he is. The best clutch hitters belong in the Hall of Fame, other things being equal.

* * *

We are trying to compare RBI performances over time but qualitative changes have been made, meaning a given RBI total does not necessarily mean the same thing today as it did 35 years ago. This study will only concern itself with the post-1920 era, for before Babe Ruth's time, the ball was considered "dead". In today's ball games, artificial grass is used. This makes it easier to get RBIs since the ball bounces more erratically. Also much more traveling takes place today and this has an effect on performance by regulars. These are all qualitative factors and there is no way they can be taken into account in evaluating performances from different eras.

Another problem in comparing RBI performance is that a player may lead the league but if the league was a pitchers' league, his total would be lower, on an absolute basis. His performance might be relatively higher, when compared to other players in the same season, than a player who had 15-30 more RBIs and finished fifth or lower in the league RBI race. Many players would have more RBIs in the later season, than the player who led the league in the former season with a not so high total.

The method proposed here to make any statistic comparable from more than one season is similar to that used by John Warner Davenport in Baseball Graphics for determining the triple crown index of either batting or pitching: percentiles. Percentile says that a person's performance is better than a particular percent (computed) of those in the sample. Percentiles enable us to compare ballplayers whose RBI totals (one year) come from different RBI distributions (different years). It was felt that in order to compute these percentiles, ten seasons of RBI data (both leages combined) would be adequate. A player's RBI total was used if he had 300 or more AT BATS. The source for the data was the Encyclopedia of Baseball (1969), Who's Who in Baseball (1983), and Street & Smith's Yearbook (1978). Due to the Encyclopedia's listing annual statistics by team, a player had to get the required number of at bats on one team, in most cases. Therefore, some RBI totals that should have been included were not or a man was short-changed of some RBIs due him. This occured when a player was on more than one team in a season.

The ten seasons used to create an overall RBI distribution were not chosen randomly. They were all chosen after the end of the dead ball era. The seasons chosen were selected such that the extreme RBI totals would be adequately represented in the sample. Of 34 150+ RBI seasons after 1920, 11 are represented in the sample. In fact, for each year used to create the sample, the first two digits in the RBI total leading the majors in a given year is never identical. Thus, Hack Wilson set the all time record with 190 in 1930, Lou Gehrig had 184 in 1931, Ruth had 177 in 1921, right on down to 1968 when Ken Harrelson led the majors with but 109. The other years chosen were 1933, 1943, 1955, 1962, 1964, and 1977. These ten years are not claimed to represent RBI performance perfectly, but they represent great years for batters in general, pitchers in general, War time, expansion time, and the designated hitter. Also, each decade since the end of the dead ball era is included. Note that neither Ruth's nor Roger Maris' big HR season, 1927 or 1961, is included.

For each season chosen, a frequency distribution of RBI is created, using the same intervals for all seasons. The intervals are of 10 RBIs: beginning 10-19 RBIs, 20-29 RBIs,...,190-199 RBIs. As was mentioned, each RBI performance during these ten years in which the batter had 300 or more AT BATS is included.

When all ten frequency distributions are created, the frequencies for each season for the same interval are added up to arrive at one distribution for the ten years. We now have one distribution of RBI totals for single seasons in which to compare any forthcoming performance and that distribution consists of ten years of data.

Some observations on this distribution are stated. There is a sample size of 1372 performances. The years after expansion account for a larger share of the sample since there are more teams. 1977 had 213 players and 1955 had but 105. In 1930, 32 players had 100+ RBIs compared to 3 in 1968. In 7 out of the 10 seasons, the interval with the highest frequency was 40-49 RBIs. There were 238 players in this interval. Of the 12 cases in which a player had 10-19 RBI; 6 occured in 1968. In 11 cases a player had 150 or more.

*This article was written in 1978. Since that time several of the players discussed have made the HOF; Klein, Mize and Wilson.

TABLE I
RBI Frequency Distribution: Ten Selected Seasons

RBIs	# Players	% of Total	Cumulative # Players	Cumulative % of Total
190-199	1	0.1	1372	100.
180-189	1	0.1	1371	100.
170-179	3	0.2	1370	100.
160-169	3	0.2	1367	99.8
150-159	3	0.2	1364	99.6
140-149	3	0.2	1361	99.4
130-139	10	0.7	1358	99.2
120-129	20	1.5	1348	98.5
110-119	43	3.1	1328	97.0
100-109	67	4.9	1285	93.9
90- 99	68	5.0	1218	89.0
80- 89	105	7.7	1150	84.0
70- 79	149	10.9	1045	76.3
60- 69	182	13.3	896	65.4
50- 59	215	15.7	714	52.1
40- 49	238	17.3	499	36.4
30- 39	160	11.7	261	19.1
20- 29	89	6.5	101	7.4
10- 19	12	0.9	12	0.9

When this total frequency distribution is completed, we can calculate percentiles for RBIs using the standard formula in an introductory statistics textbook. The following table summarizes the comparison of 1372 players chosen from ten seasons.

TABLE II
RBIs Associated with Various Percentiles - Post 1920

PERCENTILE	# RBIs*
99th	141
95th	114
90th	103
75th	79
50th	59
25th	43
10th	32
5th	26
1st	20

* Rounded off to the nearest RBI.

If a ballplayer were to bat in 59 runs in the upcoming season, he would be in the 50th percentile. In other words, his performance would be better than 50% of the ballplayers in the sample. Hopefully, the results can be projected to include the entire population of baseball players who played since 1920; so that 59 RBIs is better than 50% of all seasonal RBI marks since 1920. That is the intent of this article. The 50th percentile is called the median.

The difference between the 25th and 75th percentiles is 36 RBIs. This is known as the interquartile range. This is the range over which half of the sample falls. Despite the overall range being 172 RBIs, 50% of the players sampled are within an interval of 43 to 79 RBIs.

* * *

Now that we have seen the method recommended to use for selection of candidates into the Hall of Fame, we need to discuss criteria for selection. Obviously, RBI is not the only criterion that should be considered. It seems that for every statistic considered, a ballplayer should rank at a certain percentile or greater for a certain number of years. What percentile and number of years are chosen is arbitrary. Whatever is chosen should indicate stability and excellence in performance. The level of the criteria chosen will also reflect avoidance of injuries. A greater number of years chosen will indicate a greater degree of stability. A greater percentile chosen will indicate a greater degree of excellence. These criteria will answer the twin questions of how good was he, and for how long was he this good.

What proportion of ballplayers should be Hall of Famers? 10%? 5%? 1%? This is for the sportswriters to decide. This reporter would recommend that the Hall of Fame consist of players who in addition to being in the majors for ten full seasons were in the 95th percentile (top 5%) in any of three or more statistics for five or more years apiece, and urge that this is not a requirement, but a guideline. Qualitative factors may still be used as a basis for induction if the player in question does not meet the above criteria.

Are these criteria realistic? To find out, some percentile rankings for some ballplayers are computed for RBIs only. Ralph Kiner is chosen because his selection was disputed. Also included are Hack Wilson, Chuck Klein, and Johnny Mize because they are considered to be among the most overlooked. Vern Stephens is included because he was a rarity - a shortstop who hit with power. He is not enshrined. Also, after easy-to win election Hall of Famers Ted Williams and Jimmy Foxx; he had the third highest slugging percentage among former Red Sox: .492. Tony Perez and Billy Williams are used as contemporary examples. Perez had eleven consecutive 90+ RBI seasons. Williams has the most homers among retired players who are not in the Hall of Fame: 426. Carl Furillo is not in the Hall of Fame, but he played in the All Star Game and was in the same lineup as three Hall of Famers: Campanella, Snider, and Robinson. Two batters who are not noted for home run hitting are also included: Ernie Lombardi and Arky Vaughn. There are some observers who feel these two men have been overlooked by the Hall of Fame.

The data shows that both Kiner's and Wilson's RBI performances were in the 95th percentile or better for five years, four consecutive, and six years in the 90th percentile or better. Klein's record was a shade better. In his first five 300+ AB seasons,

Klein was in the 95th percentile or better. In a sixth season, he was above the 90th percentile. Johnny Mize's record is perhaps more outstanding than any of the three superstars just discussed. Although his RBI performance was in the 95th percentile or better for four years compared to five for each of the other three heroes, his RBI performance was in the 90th percentile or better for seven seasons. He had more RBI performances in the 90th percentile or better than Klein, Wilson, or Kiner. If these players fared this well in other measures of performance, they would deserve to be in the Hall of Fame.

Stephens did something that neither Kiner, Klein, Wilson, nor Mize did. For three straight seasons his RBI performance was in the 99th percentile. For three straight seasons he was better than virtually all 1372 players in the sample. This information would make it imperative that Stephens be enshrined. However, he only had one other year in which he was even in the 90th percentile or better. Therefore, unless he did better on some other statistic or there are some qualitative factors about Stephens' performance that make him deserving, he does not deserve to be in the Hall of Fame based on the criteria stated on page four.

Tony Perez' and Billy Williams' records may not look as brilliant as the sluggers previously mentioned. Those players mentioned previously rated well for the "how good" aspect in evaluation. Neither Perez nor Williams ever led their league in RBIs. They both have had only two years of being in the 95th percentile or better. But in eleven consecutive seasons, Perez was never below the 84th percentile. For thirteen consecutive years, Williams was never below the 79th percentile. Williams' percentiles are clearly higher than Perez' for the twelfth and thirteenth seasons of their careers, respectfully. These consistent performances are something that neither Klein, Wilson, Kiner, nor Mize can come close to bragging about. Both Perez and Williams should be considered for the Hall of Fame when they retire and become eligible in 1982, respectfully. This assumes they rate well in other areas. They both rate well for the "how long" phase of evaluation when RBI performance is considered.

Arky Vaughn exceeded the 80th percentile four times (three in succession) and Ernie Lombardi did it twice, in succession. This is not good enough to be labeled Hall of Fame material, but as has been emphasized throughout this paper, this is not meant to be the only statistic used in evaluation. In addition, qualitative factors must be considered.

The criteria set forth in this paper appear to be realistic for deciding who does and who does not belong in the Hall of Fame. Of course, only one statistic, RBI, was examined and the standards call for at least three statistics at which one performs at a high level for a specified period of time.

Finally, no one mentions Carl Furillo as Hall of Fame material. But in a ten year span, 1947-56, when the Dodgers won six pennants; he was in the 78th percentile or better for RBIs nine times. Now you baseball card buffs who remember the Miller Lite Beer commercial, would you trade twenty five Mary Thronberrys for one Carl Furillo?

TABLE III

Percentiles Associated With RBI Performance For Ten Hall of Fame Candidates

H. Wilson			C. Klein			A. Vaughn			Lombardi			J. Mize		
YR	RBI	Percentiles	YR	RBI	%	YR	RBI	%	YR	RBI	%	YR	RBI	%
24	57	47	29	145	99	32	61	53	32	68	63	36	93	85
26	109	93	30	170	100	33	97	87	33	47	31	37	113	95
27	129	98	31	121*	97	34	94	86	34	62	55	38	102	90
28	120	97	32	137	99	35	99	88	35	64	57	39	108	93
29	159*	99	33	120*	97	36	78	74	36	68	63	40	137*	99
30	190*	100	34	80	76	37	72	67	37	59	50	41	100	89
31	61	53	35	73	69	38	68	63	38	95	86	42	110*	94
32	123	97	36	104	91	39	62	55	39	85	80	46	70	65
33	54	43	37	57	47	40	95	86	40	74	70	47	138*	99
			38	61	53	41	38	17	41	60	52	48	125	98
			39	56	46	42	49	35	42	46	29	49	64	57
			40	37	16	43	66	60	44	58	49	51	49	35
									45	70	65			

R. Kiner			C. Furillo			V. Stephens			B. Williams			T. Perez		
YR	RBI	%	YR	RBI	%	YR	RBI	%	YR	RBI	%	YR	RBI	%
46	81	77	46	35	13	42	92	85	61	86	81	67	102	90
47	127	98	47	88	82	43	91	84	62	91	84	68	92	85
48	123	97	48	44	26	44	109*	93	63	95	86	69	122	97
49	127*	98	49	106	92	45	89	83	64	98	88	70	129	98
50	118	96	50	106	92	46	64	57	65	108	93	71	91	84
51	109	93	51	91	84	47	83	78	66	91	84	72	90	84
52	87	82	52	59	50	48	137	99	67	84	79	73	101	89
53	116	96	53	92	85	49	159*	99	68	98	88	74	101	89
54	73	69	54	96	87	50	144*	99	69	95	86	75	109	93
55	54	43	55	95	86	51	78	74	70	129	98	76	91	84
			56	83	78	54	46	29	71	93	85	77	91	84
			57	66	60				72	122	97	78	78	74
			58	83	78				73	86	81	79	73	69
									74	68	63	80	105	92
									75	81	77	81	39	19
									76	41	21			

* Led league