BASEBALL ANALYST

JOURNAL OF SABERMETRICS

August, 1987 Vol. #31

Before we get to this month's fine collection of articles, I wanted to thank you all for writing. As long-time Analyst readers know, Jim Baker and I have spent a good part of the Editors Page over the years pleading with people to send material, for the cupboard was nearly bare. I am pleased to say now that the cupboard is no longer bare; we have good articles that I can't quite find room for in making up an issue. This does not mean we don't still need articles: we do, very much. But we need them now for a different reason, that being to create options and hopefully drive the quality up. While the quality of articles still is not what it might be--I know you people are capable of doing better work than I have yet been able to get out of you--certainly in terms of quality we have come a long way from where we started five years ago.

I also wanted to apologize here for not doing a better job of corresponding with people who send articles. Whenever I get an article in, I always think, "I've got to write to this person and thank him for the article. . .I also want to tell him this and that and the other about what I like and don't like about his work." I always think it, but I'm bad about getting it done. I mean to write to Dick O'Brien and Russ Eagle and Sandy Sillman and I don't know who all and thank them for the fine articles they have submitted, but I just don't get around to it. I apologize for that, and I hope you all will understand that I'm not defending my negligence, but I do appreciate your work, and keep it coming if you can.

Five articles this time:

Editor and Publisher: Bill James Address: Box 171, Winchester, KS 66097

Similarity Scores among the All-Time Greats

Robert O. Wood & Robert K. McCleery

I. Introduction

In a recent article in this journal entitled "Hall of Fame and a New Historical Ranking of the All-Time Greats", we described our methodology of evaluating the greatest ballplayers of all time. Basically it involves considering a wide array of statistics and combining them sensibly into a value formula. We took special care to take into account a player's era and home ballpark, as well as his defensive contribution. The earlier article showed that in general our results agree with the consensus among knowledgeable baseball people. We then pointed out various Hall of Fame (HOF) imposters, along with players that we think deserve to be enshrined who are not given their due respect, based upon their career values.

In this article we wish to address many of the same issues, but now focusing upon comparative rather than absolute values, namely similarity scores, a term coined by Bill James. We believe that our absolute formula is the best way to evaluate players, and that HOF membership is earned by those players with career values above some arbitrary cutoff. But we also realize that it is very difficult to convince people of our analysis's preeminence, especially in such a short article, and that relativistic arguments are easier to make and paradoxically often are more persuasive.

We have constructed the similarity score for each pair among the 850 hitters we have ranked. In truth this secondary analysis is a type of insurance for the reader. It may be that in our value formula we have given too much weight to power over speed, or too little to defense, etc., so that looking at similar players should obviate any such biases, and serves as a test of our methodology.

II. Similarity scores

The eight variables that make up our career value formula are: career batting average and slugging percentage relative to league average; home runs and runs batted in relative to league leader; defense; total bases including walks, steals, caught stealings, ground into double plays, and hits by pitch; runs scored; and post-season performance. (See the earlier article for a more detailed description of the variables and how we abstracted from a player's era and home ballpark. We also credited years lost to military service or to the color barrier.) To review our career value scale: a player with a value of 300 is probably a first-ballot HOF member, such as Yogi Berra, Willie Stargell, Ernie Banks, Paul Waner or Al Simmons; a player with a value of 265 has about a 50/50 chance of making HOF, like Nellie Fox, Bill Terry, Joe Gordon, Edd Roush, Ernie Lombardi, or Bobby Bonds; a player with a value of 230 should have virtually no chance of entering HOF, like Harvey Kuenn, Bobby Murcer, Jim Fregosi, Heinie Groh, Bill Mazeroski, or Ken Keltner. According to our analysis, there are currently 37 players with values exceeding 300, 80 exceeding 265, and 180 exceeding 230. After we (conservatively) estimate the careers of all active stars, the numbers become 38, 88, and 192, respectively.

All variables except post-season performance were used to construct the similarity scores. Two perfectly matched players have a score of 1000. We use sample standard deviations of the other seven variables as the penalties, and if one player is exactly one standard deviation away in each of the categories from another player, their score would be 850.

Let us call the similarity score of the most similar player to a given player as his uniqueness rating. Uniqueness ranges from 991 to 855, and it should be no surprise that the 855 belongs to Babe Ruth. As a guide to the similarity scores, note that roughly 60 of the 850 players have uniqueness ratings below 950, 325 between 950-970, 375 between 970-980, and 90 with a uniqueness rating above 980. This last group of 90 consists of the most "common" players in history, whereas the first group of 60 consists of the most unique players, and almost all are superstars. There is a definite inverse relationship between values and uniqueness ratings.

Table 1

	Value		Value	Similarity
Jim Landis	141.95	Mickey Stanley	140.24	988.80
Jeff Heath	229.04	Hal Trosky	232.09	988.48
Curt Blefary	124.33	Ken Harrelson	127.64	987.78
Kirk Gibson	198.48	Tony Conigliaro	194.82	987.59
Al Bumbry	163.80	Ron Leflore	157.99	987.02
Phil Caverretta	203.40	Buddy Lewis	206.23	986.49
Bob Horner	209.88	Roger Maris	218.13	985.83
Craig Reynolds	149.28	Gene Alley	147.58	985.01
Ron Cey	245.97	Gil Hodges	232.90	984.64
Horace Clarke	154.14	Bobby Richardson	163.54	984.45
Rocky Colavito	234.60	Boog Powell	235.38	983.58
Alan Trammell	236.89	Lou Whitaker	234.59	983.25
Willie Horton	209.05	Greg Luzinski	213.74	983.18
Johnny Callison	216.39	Rick Monday	214.96	982.76
George Scott	209.09	Frank Schulte	202.86	982.19
Harold Baines	221.16	Kent Hrbek	219.09	981.66
Rich Gedman	194.99	Doug Rader	190.23	980.78
Dwayne Murphy	182.68	Rico Petrocelli	192.40	980.40
Jose Cruz	249.66	Minnie Minoso	254.88	979.99
Bob Boone	202.70	Jim Sundberg	209.50	979.93
Fred Lynn	252.11	Larry Doby	256.33	978.72
Leon Durham	197.11	Ben Oglivie	196.86	978.65
Pedro Guerrero	244.86	Babe Herman	256.81	974.81
Rickey Henderson	263.54	Vada Pinson	260.00	973.97
Al Oliver	256.87	Zack Wheat	270.40	973.93
Dwight Evans	230.18	Jimmy Wynn	236.43	973.56
Jimmy Piersall	171.67	Bill Virdon	156.37	973.43
Jack Clark	230.65	Norm Cash	244.62	972.64
Steve Garvey	273.02	Sherry Magee	262.81	969.91
Paul Molitor	222.02	Gil McDougald	205.86	967.90
Willie Davis	271.13	Vada Pinson	260.00	967.61
Cesar Cedeno	258.75	Kiki Cuyler	258.36	966.00
Carney Lansford	226.95	Carl Furillo	221.46	965.38
Jim Rice	277.28	Billy Williams	274.13	946.88

Table 1 presents some of the most interesting similarity pairs we have found. Included are very similar pairs from retired players, along with the most similar players to some active or recently retired players. Listed along with the names are the players' career values and the similarity score. Note that the career values (and associated seven variables) are estimated for active players and are listed in the table.

Since the Hall of Fame is the focus of this article, we present the table as background, to give the reader an idea of the similarity concept as applied to our analysis, and move on to HOF issues.

III. Hall of Fame Imposters

In our earlier article we presented the career values of all current Hall of Fame members. There are several with values under 230, our arbitrary cutoff below which a player deserves no chance of making HOF. Recall that there are 180 players with values above 230, so we consider the cutoff as generous. In this section we provide further evidence of these players' imposter status via a similarity analysis. The player's value is given in parentheses, and note that the values may differ slightly from those of the earlier article since we have revised our park effects.

Chick Hafey (228.72). Very similar to Al Rosen. Most similar recent player is Bill Skowron.

Earl Combs (228.30). Somewhat similar to Ginger Beaumont and Lloyd Waner. Most similar recent player is Mickey Rivers.

Dave Bancroft (225.38). Most similar players are Dick Bartell and Phil Rizzuto. Similar recent players are Dave Concepcion and Willie Randolph.

Joe Tinker (222.15). Most similar player is Frank White. Also similar to Sherm Lollar and Travis Jackson.

Travis Jackson (218.53). Most similar to Jimmy Williams. Also similar to Tony Cuccinello, Joe Tinker, Bill Freehan and Frank White.

Lloyd Waner (210.40). Similar to Pete Runnels, Johnny Pesky, Roy Thomas, and Willie Wilson.

Johnny Evers (209.68). Similar to Bill Russell, Dick Bartell, Phil Rizzuto, and Miller Huggins.

Roger Bresnahan (201.52). Very similar to Gil McDougald. Also similar to Bill Bradley and Hans Lobert. Most similar recent players are Tim McCarver and Garry Maddox.

Fred Lindstrom (200.32). Very similar to Vic Power and Baby Doll Jacobson. Most similar recent players are Terry Puhl and Garry Maddox.

Frank Chance (194.70). Very similar to Chick Stahl. Also similar to Wally Moon, Tony Gonzales, and Bing Miller. Most similar recent players are Terry Puhl and Jerry Mumphrey.

Rick Ferrell (189.10). Similar to Al Lopez, Willie Kamm, and Marty Marion. Most similar recent players are Rick Burleson and Bill Russell.

George Kelly (188.94). Very similar to Joe Rudi. Also similar to Bobby Thomson, Frank Thomas, and Jeff Burroughs.

Ray Schalk (164.60). Similar to Everett Scott, Al Lopez, Mickey Doolan, and Rollie Hemsley. Most similar recent players are Ed Brinkman and Alfredo Griffin. Note that Schalk is somewhat similar to Bob Boone and Jim Sundberg, but is forty points lower in career value, due to both inferior offense and defense.

It is clear that while the similar players to the above HOF members are good players, we do not think of Cooperstown as their final resting place. Although it is more obvious from the complete list of similar players, there is a great deal of in-breeding within the HOF. On the list of the ten most similar players to these HOF imposters, we often see other imposters. This leads us to believe that some sort of "least qualified similar member", i.e. lowest common denominator, strategy is used by HOF voters. It is our hope that our absolute methodology when combined with the relativistics of similarity scores can shake some sense into responsible, knowledgeable baseball people, and perhaps even into HOF voters.

IV. Deserving Future Hall of Famers

If we were given ballots for the Hall of Fame, in the absence of any way to register the degree of intensity of preference and only allowed to say in or out, we would probably vote for a player if and only if his career value exceeds 280. We realize however that these standards may be a trifle stringent, so we will consider a player as "deserving" if he has a value over 265, with better than a 50/50 chance of making HOF.

In our earlier article we listed the current HOF members and their, in several cases, woeful career values. This fact does not directly imply that HOF voters have lower standards than we do (although we are sure they do), since we also demonstrated that the voters suffer greatly from an era illusion that made them induct all those 1920-1940 hitters with the superficially gaudy statistics.

The obsession with high batting averages of that era has in turn worked against the players of the modern (post-1960) era. This section will discuss the most similar players to eight stars since 1960 who we feel merit HOF consideration, yet are often overlooked. The player's career value is listed in parentheses, and if active, both his (estimated) future and current values are given.

Dave Winfield (305.33, 291.84). A truly great player whose home ballparks have helped to obscure this fact. Has won six gold gloves. Of his ten most similar players, seven are current HOF members along with Eddie Murray, Jim Rice and Tony Perez. Most similar are Rice, Goose Goslin and Duke Snider.

Ron Santo (286.05). Was a whopping 232 votes shy in the last HOF balloting. Most similar player is Bobby Doerr, but Santo had more power. Both were excellent glove men in hitters' parks. We believe both deserve to be in HOF.

Gary Carter (285.52, 253.65). The three most similar players are Bench, Santo and Doerr. Due to Mets' success, he will probably make HOF. If he does not suffer a career-influencing injury and achieves the career we have estimated for him, Carter will deserve to make HOF.

Ken Boyer (278.24). Was 220 votes shy in the last HOF balloting. Most similar players include Cal Ripken, Joe Cronin, Doerr and Santo. Boyer would probably already be in HOF if he had not been a 24 year old rookie.

Orlando Cepeda (276.35). "Only" 131 votes shy in the last HOF balloting. Most similar players are Murray, Dave Parker, Duke Snider, Al Simmons, Reggie Smith, and Richie Allen. Missed his prime (age 27) season of 1965 due to injury. While we have credited some career-influencing injuries in our value formula, we do not give Cepeda anything for this. The fact that his career batting average dipped below .300 will work against him longer than his post-career drug involvement.

Richie Allen (276.23). A staggering 255 votes shy in the last HOF balloting. A misunderstood and underrated player. Sabermetricians seem to be the only ones to respect his career. His value listed above does not include any "credit" for numerous injuries during his career. Most similar players are Cepeda, Frank Howard, R. Smith, Snider, Kiner, Bob Johnson, Chuck Klein, Murray, and D. Parker.

Buddy Bell (272.05, 256.93). Most similar players are Frankie Frisch, Robin Yount, Doerr, Santo, and Boyer. Will probably retire as having played in the third most games at third base in history. Has won six gold gloves. If he can play productively through the 1988 season (as we have predicted), he may well merit HOF selection.

Willie Davis (271.13). Most similar players are Vada Pinson and Robin Yount, and Pinson was 262 votes shy in the last HOF balloting. Also similar to Cesar Cedeno, Bobby Bonds, Reggie Smith, Larry Doby and Fred Lynn, none of whom have received their due HOF consideration. Indeed Bonds only garnered 24 votes in last year's voting (on a weak ballot).

Many of the players that our analysis claims to be under-valued, including most of the above, are well above average in virtually all phases of the game — defense, average, power, longevity. However, these players have no single facet in which they were historically outstanding. Thus, their names do not come up often in either media or fan discussions since they lack a convenient "handle". Active players who may fit this mold include Buddy Bell, Fred Lynn, Robin Yount, Keith Hernandez, and Ryne Sandberg.

V. Conclusions

In order to rank the greatest players in the history of baseball, we devised a career value formula which reflects our personal beliefs about what characteristics should be considered, taking into account era and park effects. We constructed the value formula without knowing about similarity scores, and yet we now believe that the variables which make up value are precisely those which capture similarity as well. In other words, if we had sought the best way to construct similarity scores, we would have derived a value formula first, as that is the proper way to address the similarity concept. We needed to consider a mixture of "quality" statistics (e.g. batting average) and "quantity" statistics (e.g. total bases). We wanted to take into account both the player's career statistics in seasonal notation (abstracting from era and park effects), as well as the length of his career, making sure to include defense. All three factors must be considered in the determination of similar players, as well as career value.

In this article we have employed the new similarity score concept in addressing various Hall of Fame issues. We demonstrated that several HOF members do not belong in Cooperstown by looking at who their contemporary counterparts are. Also we showed that several modern (post-1960) stars have not yet received their due HOF consideration.

We hope that we have given the reader a general feel for our analysis of the all-time greats in these two articles. We have focused upon the Hall of Fame since it is one of our favorite topics, besides being a subject of debate among all baseball fans. It is our further hope that our work in the rankings and similarity scores among the all-time greats has caused some revisions in people's opinions on some specific players, but more importantly has spurred additional research into such matters.

ON BATTING ORDER

By Doug Bennion

The media and many fans seem to dwell inordinately on a team's batting order. Why do they do this? As long as the order isn't grossly stupid, I don't think it is a significant factor on run production, although it may impact psychologically on some players. I would speculate that even a grotesque lineup shuffle would have a smaller impact on a team's won/loss record than you might think. Let me see if I can convince myself, and you, that this is in fact so.

I'll deal with a truly dumb hypothetical move. In 1986 the Toronto Blue Jays sometimes featured Jesse Barfield in clean-up and Manny Lee in the 9-spot. Jesse led the majors in dingers, with 40, batted .289, and slugged .555 -- a very good 4- or 5-spot hitter. Manny had only 78 at-bats, but extrapolating from his Syracuse years he might, for a full year in the majors, have hit a couple of home runs, batted .240 and slugged .290 -- fully deserving to be entrenched in the 9-spot.

I'm going to try to prove my supposition is correct by examining the impact of a Jesse/Manny switch in the order. Beyond idiotic. If this switch makes little difference, none will. How many games would this have cost the Jays?

I know of no way to measure the difference directly in wins or losses, but we can measure roughly the difference in overall runs that might be scored. Then we'll address the issue of translating run difference into win difference.

We'll tackle the problem in three ways; a common sense assessment; a comparison using a "run projection" formula; a computer simulation. If the results from all three analyses are reasonably consistent, I'll have some confidence in their relevance.

Common Sense Assessment

Let's skim through the early part of a game after the switch has been made. Inning one. There is a decent chance that there will be a runner on base when Manny strides to the plate. Sure enough, Moseby has doubled to the right field corner, setting up just the situation we've been dreading. Yup, as expected, Manny pops up to end the inning. Hiss boo. Jesse might have driven the runner home.

Nothing much happens in the second inning. In the third, Jesse smokes one up the power alley to drive Whitt home, who had drawn a walk. Hmmmm, wait a minute, not too likely that Manny would have ripped one like that.

Do you see what the major impact of the switch is? It hasn't so much reduced run production as it has deferred it. Jesse will still get his licks in, but an inning or two later. Manny will get to hack away, but earlier.

There are several implications to consider; two which are significant and a few of lesser import. I will first address those considerations which I think should be mentioned but will have little effect on run production.

The first is the fact that I will not specifically try to account for variations in runs scored by the two once they're on base. Barfield in the 4-spot would be on base more often than Lee in the 4-spot. With the 5-, 6-, and 7-hitters due up and raring to drive them home, wouldn't this diffence in on-base frequency be a factor? No, I don't think so. It's true that these guys would have fewer opportunities to drive in Lee, but it's equally true that the 1-, 2-, and 3-hitters would have more opportunities to drive in Barfield who would be on base more frequently than Lee in the 9-spot. Largely offsetting factors, I would think.

Secondly, I cannot precisely quantify the fact that with Lee in the 4-spot, the 3-hitter can be pitched around, detracting from his production. However, this adverse effect would in part be offset by the fact that the 8-hitter, who was being pitched around somewhat to get to Lee, will now see better pitches. The opponents revert to pitching around your 3-hitter instead of your 8-hitter. A factor, certainly, but with not great overall impact, especially on the Jays. I'd guess that this consideration might cost the team 3 or 4 runs over the year. I will more or less verify this conclusion later in the piece.

Third, before the switch, Barfield in the 4-spot would be followed by Bell in the order, so Barfield would see some good pitches. After the switch, in the 9-spot followed by Fernandez, pitchers might be a little more careful pitching to Barfield, but Tony is certainly no pussycat at the plate. I'd say that Barfield would see very little difference in the way he was pitched to.

In two major respects, however, the realignment will cost the team some runs.

The first consideration is the fact that because the hitters in the one to three spots have higher on-base-percentages, they will get on base a little more often than the hitters in the six to eight spots, so the 4-hitter will have more opportunities than the 9-hitter in which to drive in runs. Manny in the 4-spot won't drive in as many as Jesse, not for any "clutch" reasons, but because Manny is just much less effective overall at the plate. The second problem is that the number 4-spot enjoys more at-bats, because games will often end with the 4- to 8-spots at bat. When Jesse is in the 9-spot we'll lose some of his production.

I'm going to attempt to quantify these differences. Its not possible to achieve absolute accuracy here; rather my objective is a general impression that's not unreasonable. My assumptions are as follows:

* The 4-spot will get 680 at bats, the 9-spot, 600. I'm ignoring walks.

* When at bat in the 4-spot, there is a 25% chance that there will be a runner(s) in scoring position. In the 9-spot, that chance is somewhat less, say 22.5%.

* Manny will hit for a .240 average, Jesse for .300.

* To reflect the differences in power, when Manny gets a hit with runner(s) in scoring position, he will drive in 1.25 runs whereas Jesse will drive in 2 runs.

Trust me. These are reasonable, conservative assumptions that probably exaggerate the production differences between the two hitters. I won't bore you with the arithmetic, but given these parameters when Jesse bats fourth he drives in 102 runs and when Manny bats ninth he drives in 41 runs, for a total of 143 runs. (These are reasonable proxies for their real-life production. Jesse had 108 RBIs and Manny would have had about 45 given a full year. So our assumptions aren't totally stupid.) When the hitters switch position, Jesse drives in 81 runs and Manny drives in 51 runs, for a total of 132 runs. The original alignment is better by about 11 runs. If you use different sets of reasonable assumptions the differential will still be of the same order.

If we add a few runs for the three lesser factors described previously, the total difference is, say, 15 runs.

Run Projection Analysis

For this purpose I have chosen to use a run projection formula that I developed called Run Average. Run Average is unerringly accurate. Think of it as something like Runs Created per at-bat. In 1986, Jesse's Run Average was .171 which means that his run contribution was .171 runs for each at-bat. Manny didn't have enough at-bats for a meaningful calculation, but I will estimate it conservatively at .090, which would put him in the league basement.

When Jesse is in the 4-spot with 680 total at-bats, and Manny in the 9-spot with 600 total at-bats, the total run contribution by the two is 170 runs (Not to be confused with RBIs). When the players switch positions, the total contribution is 164 runs. The original alignment is better by about 6 runs.

Here we can try to quantify the difference due to the fact that after the switch, the opposition pitches around our 3-hitter rather than our 8-hitter. Say the 3-hitter has a Run Average of 0.140 and suffers a 10% reduction in production (a lot) due to the fact he isn't seeing as many good pitches. In 700 at bats, this would cost the team about 9 runs. Now, assume the 8-hitter has a Run Average of about .120 and benefits from a 10% increase due to the fact he sees better pitches. In 620 at-bats this would <u>gain</u> 7 runs. Total net cost, 10-7 or 3 runs, verification of our gross assumption in the previous section.

The total cost is, say, 9 runs, a conclusion of the same general magnitude as our common sense assessment.

Computer Simulation

We can use a computer to play a realllly fast baseball game. The technique is called a Monte Carlo simulation, but all we need to know right now is that the simulation is based solely on percentages and it replicates a season's play with surprising accuracy. The simulation uses individual player offensive statistics, with generalized play percentages (Quick, what's the chances of advancing a runner from first to third on a single? With a runner on third and fewer than two out, what percentage of runners will score on a ball hit to the outfield in the air? What percentage of runners will score from third when a ground ball is hit? And so on, and so on). If given the opportunity I will describe it in more detail in a future piece.

Trust me again. First I played ten <u>seasons</u> of games ... to minimize chance fluctuations ... on my simulator with Jesse in the 4-spot and Manny in the 9-spot. These electronic Jays scored an average of 805 runs per season (the flesh and blood Jays scored 809 runs). I then played a further ten seasons with a Jesse/Manny switch, producing an average of 791 runs per season, a difference of 14 runs per season.

Well I'll be darned ... our common sense assessment suggests a 15 run difference, our run projection method a 9 run difference, and our Monte Carlo simulation a 14 run difference. Wow, reasonable accord.

Have I convinced you? You can jiggle the assumptions around a little to come up with perhaps 5 runs or perhaps 20 runs, or you might look askance at this interpretation or that technique, but I'm convinced that the weight of evidence is persuasive that this maniacal switch would cost about a dozen runs, give or take a few. How do we translate this run deficiency into a number of losses?

It is commonly accepted -- easily verified -- that to turn a win into a loss, or vice versa, requires a run differential of about 10 runs. For example, from 1984 through 1986, the Jays have won 274 games and lost 211, and they have scored 2318 runs and allowed 2017. Subtract 31 wins to even their won/loss record. Now, using our 10:1 ratio, subtract 310 runs from the runs scored total. Jays runs scored would now be 2008, which combined with runs allowed of 2017 should produce an even won/loss record. Presto! Ten runs equals one loss.

The bottom line, then, is that the Jesse/Manny switch would cause the team to lose perhaps one additional game. Maybe zero, maybe one, maybe two, unlikely more. Just one loss for a real bonehead switch. Far subtler switches, which the media are always discussing, would have a considerably smaller impact on runs...small fractions of a win. In fact, our Monte Carlo simulation would illustrate that the difference between the optimum team lineup and the dumbest possible team lineup .. likely featuring Manny in the four spot and Jesse in the nine spot .. would be only forty/fifty runs or four/five games over a season.

I will conclude by suggesting that a manager shouldn't spend sleepless nights worrying about some subtle lineup shuffle .. the marginal gain in overall offence just isn't worth the effort. He would benefit the team and himself more by getting a good night's sleep.

Bias in Infielding Evaluation 2:
Pitching Handedness and Strikeout Tendencies

Charles Pavitt

In Analyst 24, I started to study the effect of bias, or systematic error, on the evaluation of fielding. At that time, I explored the effect of what I called a pitching staff's "handedness," the proportion of a staff's innings pitched by righthanders versus lefthanders, on infielders' assist totals. A mostly righthanded staff would, given platooning, face a preponderance of lefthanded batters. Given the tendency to pull the ball, the result would be a relative increase in balls hit to the first and second baseman, and thus an overestimate of their relative abilities as measured by assists per game. For the same reason, less balls would be hit to the shortstop and third baseman, and assists per game would underestimate their abilities. Analogously, a mostly lefthanded staff would mean an overestimate of the ability of the left side of the infield and an underestimate of the right side. Based on data from the 1980 and 1981 seasons, there are significant biases in the assist totals of second and third basemen and shortstops that are predictable from a pitching staff's handedness. In a secondary analysis, I showed that this bias could have a substantial effect on the relative standing of shortstops at either skill extreme (based on assists per game) playing for teams with pitching staffs with proportionately high or low amounts of righthandedness.

The goal of this paper is to extend the previous study in three directions. First, I will now include five seasons (1980 to 1984) in the analysis. Second, I will explore the effect of bias in both leagues separately. Third, I will estimate the impact of a second possible bias on infielding evaluation, a pitching staff's tendency to strike out opposing batters. A proportionately high total of strikeouts means a reduction in batted balls and a underestimate of the ability of all infielders as measured by assists per game.

The analysis was performed similarly to the earlier study. Data for each team for each of the five years was used, resulting in a sample of 60 for the National League and 70 for the American League. The data consisted of the proportion of the team's innings pitched by righthanders, the pitching staff's mean number of strikeouts per game, and the mean number of assists per game at each infield position. First, to determine whether the predicted biases occurred in fact, correlations were calculated between assists per game and, in turn, pitching handedness and strikeouts per game, for each league separately and combined. The results were as follows:

Handedness and Assists/Game

	A. L. (N = 70)		N. L. (N = 60)		Combined (N = 130)	
Position	Corr.	Prob.	Corr.	Prob.	Corr.	Prob.
First	.193	.110	.191	.145	.198	.024
Second	.228	.058	.433	.001	.326	.001
Short	-.379	.002	-.251	.054	-.283	.002
Third	-.490	.001	-.566	.001	-.541	.001

Strikeout/Game and Assists/Game

Position	A. L. (N = 70) Corr.	Prob.	N. L. (N = 60) Corr.	Prob.	Combined (N = 130) Corr.	Prob.
First	.269	.025	-.256	.049	.021	.814
Second	-.298	.013	-.363	.005	-.265	.003
Short	-.230	.056	-.438	.001	-.250	.005
Third	-.233	.052	-.232	.075	-.284	.002

As in the earlier study, the handedness bias was supported for three of the four positions. In each league, increases in righthanded innings was associated with more assists by second basemen and less assists by shortstops and third basemen at a significance level of less than .06 in all cases. While via eyeball the correlations seem to differ between leagues, via statistical analysis they do not. Thus, we can pool both leagues and use the correlations for the combined data, which due to increased sample size, are more stable. As earlier, the effect for first basemen is lower than expected, although statistically significant for the combined data. As I stated in the earlier paper, Pete Palmer has suggested that this result could be due to the fact that first basemen play close enough to the bag against lefty batters to make the putout on grounders, while against righties they are further from the base and need to flip the ball to the pitcher to make the play. This would imply that lefthanded pitchers, facing a preponderance of righthanded batters, would make more putouts than lefthanded pitchers, who would face more righties. However, Dallas Adams (Analyst 30) has presented evidence showing that lefthanded pitchers make less putouts than righties, a finding that serves as circumstantial evidence against Pete's suggestion. Thus, this finding is still in need of explanation.

As for strikeouts, we have general support for the claim that more strikeouts means less infielding plays. An exception is for American League first basemen, with a significant positive explanation. Perhaps A. L. first basemen get more dribbling grounders from overpowered righthanded batters when strikeout pitchers are on the mound. If so, why not in the N. L. too? The result is probably a fluke.

In order to determine the practical implications of all this for evaluating infielders, I did a series of multiple regression analyses (thanx to Dallas Adams for supplying me with a program for this). The regression model is

$$Z = a + b_1 X + b_2 Y$$

with Z = the expected number of assists by an infielder
X = the percentage of righty innings
Y = strikeouts per nine innings
b_1 and b_2 = the regression coefficients (weighting functions)
and a = the value of Z when X and Y both equal zero (intercept)

The resulting regression coefficients were:

Handedness and Assists/Game

Position	A. L. (N = 70) b	Prob.	N. L. (N = 60) b	Prob.	Combined (N = 130) b	Prob.
First	.002	.106	.002	.203	.002	.025
Second	.004	.044	.008	.001	.006	.001
Short	-.008	.001	-.008	.009	-.006	.002
Third	-.009	.001	-.012	.001	-.010	.001

Strikeouts/Game and Assists/Game

Position	A. L. (N = 70) b	Prob.	N. L. (N = 60) b	Prob.	Combined (N = 130) b	Prob.
First	.053	.024	-.052	.067	.002	.894
Second	-.130	.010	-.125	.007	-.108	.001
Short	-.121	.045	-.235	.001	-.115	.002
Third	-.104	.031	-.118	.005	-.105	.001

One can then use the results of the regression analysis to find out whether a particular team's fielders at a position performed well, given their pitching staff's handedness and number of strikeouts. As an example, I decided to evaluate the performance of American League shortstops during 1984. This was the season that Cal Ripken, playing every inning as always, broke the league assist record. However, the Oriole pitching staff was below average in righthanded innings (59.40% versys league average of 68.31%) and strikeouts per game (4.41 versus league average of 5.10). As a consequence, Ripken would be expected to have a greater number of opportunities for assists than other shortstops. Could his record be tainted?

To find out, we first compute the average number of assists per game for A. L. shortstops in 1984 (3.06). Second, we figure out the difference in percentage of righty innings pitched by Baltimore and the league as a whole (59.40 - 68.31 = -8.91) and multiply the difference by the handedness regression coefficient for A. L.shortstops (-.008), resulting in a correction for righty innings for Baltimore of .07 assists per game. Third, we figure out the difference in strikeouts per game for Baltimore versus the league (4.41 - 5.10 = -.69) and multiply it by the staff strikeout regression coefficient for A. L. shortstops (-.121), resulting in a correction for strikeouts by Baltimore pitchers of .08 assists per game. Fourth, we add the two corrections to the league assist average (3.06 + .07 + .08). The resulting 3.21 is the number of assists per game that the average A. L. shortstop would have made in 1984 if fielding behind the Oriole pitching staff, as represented by staff handedness and strikeouts.

As the accompanying table shows, Baltimore's 3.21 is the second highest "expected" assists per game in the league; thus, Ripken did have an "unfair" advantage over his rivals in the assist department. Nevertheless, Cal Jr. made 3.60 assists per game, which is twelve percent over his "expected" number (see the actual/expected [A/E] column). The twelve percent is the highest in the league, with only the White Sox shortstops (primarily Fletcher and Dyzbinski) close. I don't know it Ripken's record is tainted, but I know that his performance was impressive no matter how you look at it.

Data for American League Shortstops, 1984 (Mean = 3.06)

Team	% Righty Innings		Strikeouts/Game		Assists/Game		
	Amount	Correction	Amount	Correction	Expec.	Actual	A/E
Balt	59.40	+.07	4.41	+.08	3.21	3.60	1.12
Chic	70.50	-.02	5.19	-.01	3.03	3.33	1.10
Mil	73.32	-.04	4.88	+.03	3.05	3.21	1.05
Calif	71.93	-.03	4.65	+.05	3.08	3.22	1.05
Sea	59.48	+.07	6.00	-.11	3.02	3.15	1.04
Det	86.66	-.14	5.64	-.07	2.85	2.94	1.03
Cleve	76.40	-.06	4.96	+.02	3.02	3.10	1.03
KC	47.76	+.16	4.47	+.08	3.30	3.38	1.02
Tor	92.58	-.19	5.40	-.04	2.83	2.80	.99
NY	47.29	+.17	6.12	-.12	3.11	3.06	.98
Texas	69.05	-.01	5.36	-.03	3.02	2.89	.96
Oak	72.82	-.04	4.29	+.10	3.12	2.82	.90
Minn	63.64	+.04	4.40	+.08	3.18	2.81	.88
Bos	65.44	+.03	5.72	-.08	3.01	2.58	.86

I have two concluding statements. First, with two exceptions, it appears that the pitching staff handedness and strikeout biases are generally not much of a problem in making quick and dirty evaluations of infielders, as the actual assist column is highly correlated with the actual/expected column. The first exception is a team with extreme pitching biases. Kansas City is an example where bias could lead to performance overestimates; their shortstops (O. Concepcion, U. L. Washington, Biancalana) were second in the league in assists, but their pitching staff biases were such that their performance was barely greater than expected. In contrast, shortstops for Detroit (Trammell, Baker, Brookens) and Toronto (Griffin, Fernandez) would be underestimated if staff biases were ignored. These biases should be taken into account if one needs precise knowledge, such as if one was trying to rank-order the greatest fielding performances of all time. The second exception are third basemen; with correlations between handedness and assists per game around a substantial .5, evaluations may be problematic without taking handedness into consideration. Second, I must respond to Bill James' attack on those who prefer assists per game to Range Factor (1987 Abstract, page 16). A measure must be <u>valid</u>; it must actually measure what it is *that the scientist wants* it to measure. My understanding of Range Factor is that it is supposed to represent a fielder's ability to get to batted balls. As long as half of middle infielders' putouts are made on plays in which they have not fielded a batted ball, assists alone are a more valid measure of what I conceive Range Factor as trying to get at. The debate misses the point, however; the measure we really want is assists plus putouts per game <u>on balls hit to the fielder under examination</u>. Elias will not distinguish assists/putouts on balls hit to fielders and assists/putouts on force plays and the like, but Project Scoresheet gives us the opportunity to make the distinction ourselves. As I said in the earlier paper, I would like to see Range Factor redefined in this manner, and I would like to continue this project using this redefined measure.

LATE HOMERUN HITTERS
by Steve Roney

For the last couple of years, I have been looking (when I have time) at the records of home run hitters. I had started after programming Brock2 to run on my Apple. I noticed errors on big homerun hitters, and from that gathered information on homerun hitters in general. While the main object of my study is not currently in a form for publication, it did give me quick access to some of the answers raised a couple of months ago in the discussion of Frank White's homerun hitting.

Indeed, White is the oldest player to hit twenty homers for the first time and then hit twenty again, other than Luke Easter (who was black and did not play until 35) and Cy Williams (who was 34 in 1922 -- and was actually 2 months younger, anyway). Only Charlie Gehringer, John Lowenstein, Mickey Vernon and Easter were older than White the first time they hit 20. Buddy Bell last year joined them, JKuhel and GCrowe as 34 year old first-timers.

Not too surprisingly, White is not the leader in percentage of homers after 30, even throwing out dead-ball players like Williams and early black players, like Minnie Minoso. More surprisingly, White is no better than third among <u>active</u> players, being beaten out by Brian Downing and Davey Lopes. Below is a list of players who hit 65% of their homers after their 30th birthday. Most of these players hit twenty for the first time after age thirty. This list is certainly not complete, but includes a lot. Bill Robinson turned 30 in the middle of a 24 homer season, so was not included in this list. Players born in April, May and June may be ranked too high, those in July, August and September too low.

		Mo	Yr	Bef30	After	Pct
BCerv		May	26	9	96	91.4
EJoost		Jun	16	13	121	90.3
MMinoso		Nov	22	24	162	87.1
HSauer		Mar	19	42	246	85.4
DLopes	***	May	46	24	130	84.4
MVernon		Apr	18	41	131	76.2
CMaxwell		Apr	27	38	110	74.3
BDowning	***	Oct	50	47	119	71.7
EAverill		May	02	69	169	71.0
FWhite	***	Sep	50	39	92	70.2
HBauer		Sep	22	51	113	68.9
FMcCormick		Jun	11	62	128	67.4
JKuhel		Jun	06	44	87	66.4
CGehringer		May	03	63	121	65.8

The other comment about the greatest homerun hitters not hitting that many homeruns when they were 22 or 23 is also wrong:

Age	21	22	23	24
Aaron	27	26	44	30
DiMaggio	29	46	32	39
Foxx	33	37	30	58
Horner	33	35	15	35
RJackson	1	29	47	23
Mays	4	DNP	41	51
Mantle	21	27	37	52
Mathews	47	40	41	37
Murphy	2	23	21	33
Murray	27	27	25	32
Ott	25	29	38	23
Rice	1	22	25	39
FRobinson	29	31	36	31
Strawberry	26	26	29	27
TWilliams	23	37	36	DNP

USING A BASEBALL SIMULATOR PROGRAM TO CALCULATE BATTER RUNS CREATED

By Dallas Adams

Some years ago Pete Palmer described to me in qualitative terms the method underlying his baseball simulator computer program. It does not resort to a Monte Carlo approach, but rather is an exact solution. It computes the probability of each possible sequence of events, then combines them to give the final result (in runs scored per game). Recently, starting with my recollection of his general method, I worked out the details and programmed my own version. While doing so an interesting application occurred to me.

A conventional approach might be to input, say, offensive data for nine typical batters and then use the program to examine differences in team run production as a function of various batting orders. However, the thought which struck me was to input the same data into each of the nine batting-order slots. The resulting run production would then be a direct measure of the runs created by the specific batter whose batting statistics were employed. For example, input Babe Ruth's 1920 statistics into each batting-order slot; the results of the computation will give the runs created by him in 1920.[1]

The program calculates (and prints out) a table: the probability (P_N) of scoring each exact number of runs (N, where N=0,1,2,...) in a full nine-inning game, and the probability (Q_N) of scoring exactly N runs in an inning. From this table the program calculates the average runs per game ($S=\Sigma P_N N$, which forms EQ. 1) as well as, for single batter cases, the total number of runs created (RC) by that batter's input statistics. Since those statistics create S runs for each 27 outs expended, then RC=(AB-H)(S)/27 (EQ. 2).

In addition, one can go a step further. Make a second computer run utilizing, instead of Ruth's 1920 batting record, the American League batting totals for that same year. Then, by taking the output scoring probabilities from both computer runs it is possible to determine the likelihood of Ruth (in 1920) winning a game played against the 1920 American League:

let P_N = Ruth's probability of scoring exactly N runs in a game
Q_N = Ruth's probability of scoring exactly N runs in an inning
P'_N = 1920 AL probability of scoring exactly N runs in a game
Q'_N = 1920 AL probabilIty of scoring exactly N runs in an inning
m = maximum possible number of runs (set to 50 in the simulator)

then, the outcome of a game between 1920 Babe Ruth and the 1920 AL is:[2]

$$\text{Ruth's chance of winning in 9 innings} = W = \sum_{k=0}^{m-1} P_{k+1} \sum_{j=0}^{k} P'_j \qquad \text{(EQ. 3)}$$

$$\text{Ruth's chance of losing in 9 innings} = L = \sum_{k=0}^{m-1} P'_{k+1} \sum_{j=0}^{k} P_j \qquad \text{(EQ. 4)}$$

When there is a tie score after 9 innings (the probability of this equals 1-W-L), the chance of Ruth going on to win in extra innings is:[3]

$$X = \frac{\sum_{k=0}^{m-1} Q_{k+1} \sum_{j=0}^{k} Q'_j}{1 - \sum_{k=0}^{m} Q_k Q'_k} \qquad \text{(EQ. 5)}$$

Hence Ruth's overall chance of winning is W+(X)(1-W-L), which forms EQ. 6. This overall chance of winning is what Bill James calls "Offensive Winning Percentage", although his method[4] of computing it is a lot more indirect.

With regard to the calculation of runs created, the drawback of the simulator method at present is that the program somewhat simplifies the game of baseball. The program has no provision for errors, stolen base attempts, double plays, or runner advancement on outs. Omitting these things made the initial version of the program a lot easier to write. They should be included, of course, and I intend to work them into future versions.[5] At this moment, though, they are missing; and for that reason I have not put much effort into computing runs created and offensive winning percentages for large numbers of batters.

Because of the limitations of the present version of the program I restricted myself to a test case of 1920 Babe Ruth versus the 1920 American League. For both of them, the inning and game probabilities of scoring each exact number of runs are shown in Table-1. From this it can be found that the average runs per game, from equation 1, equals 4.36 for the AL and 18.29 for Ruth. Then, from equation 2, one finds that the AL input statistics produce a total of 4837 runs and Ruth a total of 194.

With regard to the latter two numbers, the American League in 1920 actually scored 5867 runs and Bill James' best determination for Ruth is 211 runs created.[6] The dominant resaon for the differences is that the program, at present, ignores errors. Of course, in the case of Ruth particularly, the program is giving us a piece of interesting information: it is telling us how many runs Ruth created strictly by his own efforts untainted by any gifts from the opposing fielders.

To date the best offensive ratings systems have been derived from league and/or team data, because these are the only data which include created runs along with the number of hits, walks, home runs, etc. which led to the scoring of those runs. These rating systems derived from league and team data are then applied to individual batters to find the number of runs their hits, walks, home runs and so forth would create. League and team runs totals, however, also benefit from the effects of opposition errors and so this effect automatically gets incorporated into the ratings equations. When such a rating system is then applied to an individual player, the player's calculated runs created total is a function not only of his personal offensive statistics but also of that league average error rate which was built into the equations. Often, of course, it is desired to have player RC's which include the effects of errors. On the other hand, there will sometimes be applications where it is wished to eliminate error contributions from an individual's RC's. Indeed, I anticipate making both types of calculations once the simulator program is modified to allow for the inclusion of errors.

As previously mentioned, the probabilities of Table-1 can be used with equations 3-6 to compute Ruth's 1920 Offensive Winning Percentage. Making the calculations, one finds that it equals .971.[7]

This .971 is rather different from the .934 figure computed using the method of Bill James.[8] I feel that the simulator's value is the more accurate determination. The method it uses (equations 3-6) is direct and logically unassailable. Whereas Bill, in executing his approach, is forced to depend on the accuracy of assumptions like a 26-out game for 1920[9] and that his "Pythagorean Equation" (i.e. that the ratio between a team's wins and losses equals the ratio between its runs scored and runs allowed) functions accurately for extremely large ratios between runs scored and runs allowed.

Regretably there are serious grounds for questioning the validity of the latter assumption. The Pythagorean Equation was derived from major league team data. I've made an analysis of the equation's accuracy as a function of scoring ratio (i.e. a team's runs scored divided by runs allowed). I am forced to report that there is a clear relationship: the equation is most accurate for scoring ratios between 1.20 and 1.30, and the farther one gets from that range the larger the standard error between a team's predicted won/lost percentage and its actual value.

For most major league teams the equation is reasonably accurate. The problem is that no major league team this century has exceeded a scoring ratio of 1.85 (1906 Chicago Cubs). Whereas when applying Bill's Offensive Winning Percentage method to Babe Ruth and other great batters one is often dealing with scoring ratios 3.00 or even 4.00 (c.f. the simulator's 1920 values of 18.29 R/G for Ruth and 4.36 for the American

TABLE-1

PROBABILITIES OF 1920 BABE RUTH AND 1920 A.L. SCORING EACH EXACT NUMBER OF RUNS IN AN INNING AND IN A GAME

	------BABE RUTH-----		---AMERICAN LEAGUE--	
RUNS	RUNS/INN	RUNS/GAME	RUNS/INN	RUNS/GAME
0	.39337	.00023	.75350	.07830
1	.15077	.00078	.11968	.11193
2	.12926	.00186	.06570	.13255
3	.10106	.00363	.03296	.13525
4	.07428	.00622	.01562	.12489
5	.05219	.00965	.00710	.10709
6	.03543	.01390	.00313	.08665
7	.02342	.01882	.00134	.06686
8	.01515	.02422	.00057	.04957
9	.00962	.02985	.00023	.03552
10	.00602	.03541	.00010	.02470
11	.00372	.04062	.00004	.01673
12	.00228	.04524	.00002	.01107
13	.00138	.04906	.00001	.00717
14	.00083	.05191	.00000	.00455
15	.00049	.05373	.00000	.00284
16	.00029	.05449	.00000	.00174
17	.00017	.05422	.00000	.00105
18	.00010	.05302	.00000	.00063
19	.00006	.05101	.00000	.00037
20	.00003	.04834	.00000	.00021
21	.00002	.04515	.00000	.00012
22	.00001	.04161	.00000	.00007
23	.00001	.03786	.00000	.00004
24	.00000	.03404	.00000	.00002
25	.00000	.03026	.00000	.00001
26	.00000	.02661	.00000	.00001
27	.00000	.02316	.00000	.00000
28	.00000	.01996	.00000	.00000
29	.00000	.01704	.00000	.00000
30	.00000	.01442	.00000	.00000
31	.00000	.01209	.00000	.00000
32	.00000	.01006	.00000	.00000
33	.00000	.00830	.00000	.00000
34	.00000	.00680	.00000	.00000
35	.00000	.00553	.00000	.00000
36	.00000	.00447	.00000	.00000
37	.00000	.00358	.00000	.00000
38	.00000	.00285	.00000	.00000
39	.00000	.00226	.00000	.00000
40	.00000	.00178	.00000	.00000
41	.00000	.00139	.00000	.00000
42	.00000	.00108	.00000	.00000
43	.00000	.00084	.00000	.00000
44	.00000	.00064	.00000	.00000
45	.00000	.00049	.00000	.00000
46	.00000	.00037	.00000	.00000
47	.00000	.00028	.00000	.00000
48	.00000	.00021	.00000	.00000
49	.00000	.00016	.00000	.00000
50	.00000	.00012	.00000	.00000

League, a Ruthian scoring ratio of 4.2).

It should be specifically mentioned that the difference between the simulator's .971 Offensive Winning Percentage for Ruth and the corresponding Jamesian value of .934 is not due in any significant degree to the fact that the simulator at present lacks certain sophistications like errors, doubleplays, stolen base attempts and runner advancement on outs. For it must be remembered that both Ruth and the 1920 American League will be similarly (to a first approximation, at the very least) affected.

If, for example, I attempt to compensate for the lack of a direct provision for errors by assuming that 76% of the 1920 American League's 1719 errors put batters on first base (i.e. if I treat those errors as the equivalent of singles) and assume that the other 24% of errors allow an extra base to be taken (i.e. if I consider them to turn singles into doubles) then the simulator predicts 5867 league runs, exactly the correct total. Assuming that Ruth benefits from errors at the same rate, his calculated RC's equal 207, which is very close to the Jamesian value of 211. Evaluating equations 3-6 with the inning and game scoring probabilities produced by these "error-compensated" cases, the resulting Ruthian 1920 Offensive Winning Percentage" is .968, not much different from the .971 found for the "no-error" case.

---

1. Actually, it isn't necessary to input Ruth's batting statisitcs nine consecutive times. Once is sufficient, for the program recognizes that this signifies a "game" simulation using the same batter over and over. Furthermore, Palmer's procedure greatly simplifies itself when a single batter is used. For example, it no longer is required to compute an inning's expected run production for each of nine different men leading off the inning; that computation only has to be made once.

2. For a brief discussion of equations 3 and 4, see my paper "Team Won/Lost Percentage as a Function of Runs and Opponents' Run's", The Baseball Analyst, April 1983, page 10.

3. For a derivation of equation of equation 5 see my paper "A Correction to 'Some Further Aspects of the Distribution of Runs' ", The Baseball Analyst, Dec. 1984, pages 19-20.

4. Bill James, The Bill James Historical Baseball Abstract, pages 295-296.

5. When GIDP, CS, SH and SF possibilities are incorporated into the simulator, then the (AB-H) term of equation 2 will be suitably adjusted.

6. James, ibid., page 445.

7. Results for a 9-inning game are: W =.962362 for Ruth winning, L =.0256107 for Ruth losing, and .0103959 for a tie score. The probability of Ruth going on to win a 9-inning tie game in extra innings is X =.808677.

8. James, ibid., page 445.

9. James, ibid., page 295.