baseball ANALYST

AUGUST 1982 ISSUE #2

This second edition of the Baseball Analyst begins with a plea. The research, gentlemen, is not pouring in. Some of you, I am certain, have ideas for research. Some of you have actually done research. This research is still in your notebooks. You are probably thinking that it isn't all that good. You are probably right. It doesn't matter; we still need it. If it makes you feel any better, the first five articles I ever published were a pile of crud. As long as I have articles in my file, I will continue to publish the Baseball Analyst. When I run out, it's over. So if you want to see this thing go--if you want there to be a place where a person who has something to say about sabermetrics can get his work published--then stop thinking about the things you could do, and hit the typewriters. Thanks.

Quantity has been a problem; quality has not. The number of man-hours of work that must be represented by these 20 pages in issue two is staggering. Dallas Adams begins on page three by reporting the results of

RE-PRINT

a massive study of age and skill patterns. Did you know that through 1968, 27-year-old players had played 152,490 games in the major leagues? One can imagine that this was not an easy thing to find out. Dallas studies the games played totals and averages of all players, contrasts that with the patters of Hall of Famers and focuses also on catchers to determine if there is any validity to the notion that a catcher matures more slowly or wears out more rapidly than other players. It makes for a detailed and fascinating study.

Craig Wright of the Texas Rangers responds to Paul Schwarzenbart's article in issue 1 on ballpark effects on fielding performance by offering some of his own research and some of his own ideas on the subject, beginning on page 10. Pages 13 to 16 contain two studies by Dick O'Brien, the first a short synopsis of a considerable amount of work on run production by batting order position, and the second contributing to the discussion of clutch hitting by an analysis of some of John Tattersall's data on home run production with varying numbers of men on base.

Jim Morrow is in search of the true slugging percentage on pages 17-19 (Marlin Perkins was invited to go along, but declined.) The article falls short of its goal, but still leaves us knowing more about the subject than we began with, which is the purpose of this magazine. And Dallas Adams concludes with a brief report on the effects of overwork on rookie pitchers. The help of all is appreciated.

SOME PATTERNS OF AGE AND SKILL

by Dallas Adams

When is a ballplayer at his physical peak? How long does it take for a young player to develop his skills up to the major league average? How rapid is a player's decline? This study was performed in the hope of answering these and other questions about players' ages and skills.

The investigation covered all non-pitchers who finished their major league career in the period 1901 through 1968. Using the MacMillan Encyclopedia (1969 edition) as a data source I tabulated, as a function of age, the number of games played by each player during each season of his major league career. A player's age as of June 30th each year was considered his age for the season. The tabulated results of all players are presented in Table 1. Separate tabulations were made for Hall of Fame players and for catchers; these data are shown in Tables 2 and 3, respectively.

TABLE 1

TOTALS AT EACH AGE FOR ALL PLAYERS WHO FINISHED THEIR MAJOR LEAGUE CAREERS BETWEEN 1901 AND 1968

Age	Games (Thousands)	Percentage of Games	Number of Players	Games/ Player	95 Games or More	Percent in 95 or More
16	.055	.003	2	27	0	0.0
17	.27	.017	18	15	0	0.0
18	2.19	.082	55	23	2	3.6
19	5.10	.320	150	34	20	13.3
20	13.01	.815	360	36	41	11.4
21	31.54	1.976	650	49	131	20.2
22	57.67	3.613	1040	55	278	26.7
23	88.78	5.562	1300	68	408	31.4
24	110.69	6.935	1560	71	587	37.6
25	134.77	8.443	1770	78	729	41.2
26	148.33	9.293	1820	81	827	45.4
27	152.49	9.553	1775	86	878	49.5
28	147.15	9.219	1670	88	842	50.4
29	138.41	8.671	1520	91	798	52.5
30	126.93	7.952	1360	93	729	53.6
31	106.89	6.697	1160	92	615	53.0
32	92.12	5.771	1020	90	526	51.6
33	73.16	4.583	820	89	411	50.1
34	54.44	3.411	650	84	290	44.6
35	40.15	2.515	480	84	209	43.5
36	28.60	1.792	365	78	146	40.0
37	18.39	1.152	260	71	77	29.6
38	11.17	.700	167	67	58	34.7
39	7.10	.445	115	62	33	28.7
40	4.10	.257	65	63	18	27.7
41	2.13	.133	40	53	9	22.5
42	.98	.061	22	45	4	18.2
43	.31	.019	15	21	1	6.7
44	.13	.008	3	43	1	33.3
45	.01	.0006	3	3	0	0.0
46	.01	.0006	1	10	0	0.0

Unfortunately, the tabulating system I used for Table 1 did not permit a completely accurate count of the number of players involved at each age; the "Number of Players" and "Games/Player" columns of Table 1 are approximately correct, the error should be less than 2% however. A different tabulation system was used in compiling Tables 2 and 3, on these tables all data are exact.

TABLE 2
TOTALS AT EACH AGE FOR ALL HALL OF FAME PLAYERS WHO FINSIHED THEIR MAJOR LEAGUE CAREER BETWEEN 1901 AND 1968

Age	Games	Percentage of Games	Number of Players	Games/ Player	95 Games or More	Percent in 95 or More
16	0	.000	0	0.0	0	0.0
17	45	.029	2	22.5	0	0.0
18	258	.167	7	36.9	0	0.0
19	904	.586	18	50.2	5	27.8
20	2022	1.312	34	59.5	10	29.4
21	4740	3.075	50	94.7	30	60.0
22	7413	4.809	59	125.5	48	81.3
23	8534	5.536	69	123.7	59	85.5
24	9189	5.961	69	133.2	64	92.7
25	9720	6.305	74	131.3	66	89.2
26	9975	6.471	75	133.0	69	92.0
27	10337	6.706	77	134.1	75	97.4
28	10151	6.585	77	132.8	71	92.2
29	9934	6.444	76	130.5	66	86.8
30	9885	6.412	74	133.6	70	95.6
31	9522	6.177	75	126.8	65	86.7
32	9483	6.152	74	128.2	65	87.8
33	8482	5.502	72	117.8	54	74.9
34	7958	5.162	71	112.0	51	71.8
35	6816	4.421	61	111.5	45	73.7
36	6211	4.029	59	105.2	43	72.8
37	4180	2.712	45	92.8	24	53.3
38	3005	1.949	36	83.3	20	55.6
39	1969	1.277	23	85.5	11	47.8
40	1456	.944	17	85.6	9	52.8
41	1075	.697	12	89.5	7	58.3
42	523	.339	8	65.4	4	50.0
43	232	.150	6	38.9	0	0.0
44	129	.084	2	64.5	1	50.0
45	0	.000	0	0.0	0	0.0
46	8	.005	1	8.0	0	0.0

One can make some immediate observations from the tables. Firstly, 27 is a significant age among major league ballplayers. More games are played by 27 year olds than by any other age. One reason is that the number of 27 year old players is greater than any other age except 26. A second observation is that 30 is the age at which the average number of games per player is the highest and also is the age at which the percentage of players appearing in 95 games or more reaches a maximum (see Figure 1). A final observation is that the number of players who appear in fewer than 95 games reaches its peak at age 25.

My interpretation of the above points is that, up to age 26, an ever increasing number of young players are given major league trials. After age 25, the number of players appearing in fewer than 95 games declines; yet the total number of games played continues to increase up to age 27. Presumably this means that, while the number of players receiving trials is declining, some of these young players succeeded in establishing themselves as good enough to play regularly. After 27, however, the picture changes. If a player has not proven his ability by this age his playing time thereafter decreases more and more rapidly (see Figures 2 and 3); and eventually the unproven player finds himself out of the big leagues. But those who do prove themselves good enough by age 27, or earlier, begin to play in increasing amounts up to age 30 (see Figure 3).

TABLE 3

TOTALS AT EACH AGE FOR ALL CATCHERS WHO FINISHED THEIR MAJOR LEAGUE CAREER BETWEEN 1901 AND 1968

Age	Games	Percentage of Games	Number of Players	Games/ Player	95 Games or More	Percent in 95 or More
16	1	.0004	1	1.0	0	0.0
17	41	.018	5	8.2	0	0.0
18	72	.032	15	4.8	0	0.0
19	504	.223	36	14.0	0	0.0
20	1287	.569	62	20.8	3	4.8
21	2865	1.267	119	24.1	5	4.2
22	6481	2.866	180	36.0	14	7.8
23	9271	4.100	229	40.5	33	14.4
24	13450	5.948	291	46.2	47	16.1
25	17460	7.721	312	56.0	70	22.4
26	20788	9.193	353	58.9	85	24.1
27	20740	9.172	328	63.2	91	27.7
28	22860	10.110	346	66.7	90	26.0
29	21042	9.306	305	69.0	89	29.2
30	19176	8.480	291	65.9	74	25.5
31	16364	7.237	242	67.6	69	28.5
32	14381	6.360	223	64.5	56	25.1
33	10582	4.680	183	57.8	32	17.5
34	8813	3.897	152	58.0	30	19.7
35	6770	2.994	120	56.4	20	16.7
36	4959	2.193	86	57.7	17	19.7
37	3128	1.384	59	53.0	7	11.9
38	2408	1.065	50	48.2	7	14.0
39	1508	.667	31	48.6	1	3.2
40	643	.284	17	37.8	1	5.8
41	309	.137	14	22.1	0	0.0
42	194	.086	7	27.7	0	0.0
43	16	.007	4	4.0	0	0.0
44	3	.001	2	1.5	0	0.0
45	5	.002	1	5.0	0	0.0
46	1	.0004	1	1.0	0	0.0

After age 30, the average number of games played in a season declines due to a combination of deteriorating baseball skills and increased susceptibility to injury.

Figure 2, which shows the percentage of games played by each age, provides some insight into the effects of aging on performance and injury. The curve for all players peaks sharply at age 27 and then turns downward. But notice that the rate of decline accelerates each year after 27 until age 31 is reached. After 31, the decline continues at, essentially, a constant rate until age 34, after which point it begins slowing. Now look at the curve for Hall of Fame players on Figure 2. This curve also peaks at 27 and then shows a small but generally constant decline until about age 32; after this age the rate of decline begins to accelerate and becomes noticeably steeper after age 33 or 34.

I feel there is a reasonable explantion of the patterns of decline. The overall decline begins after age 27 when those players who haven't proven themselves begin playing increasingly fewer games each year. During this period (ages 28-30, roughly) the decline in games played by age is primarily due to those who failed to prove themselves being released in favor of younger prospects. By age 31 a large part of the "failure" group is gone from the big leagues; the number of players who appear in fewer than 95 games for the season drops by 40% between age 27 and age 31. Further, refer to Figure 4 which shows graphically the number of players who play in fewer than 95 games. This figure shows a steady decline up to age 31, after which age the rate of decline changes, slightly but perceptibly, to a slower (but still constant) pace to age 36. After 36, it slows again.

FIGURE 1

THE PERCENTAGE OF PLAYERS APPEARING IN 95 OR MORE GAMES, AS A FUNCTION OF AGE

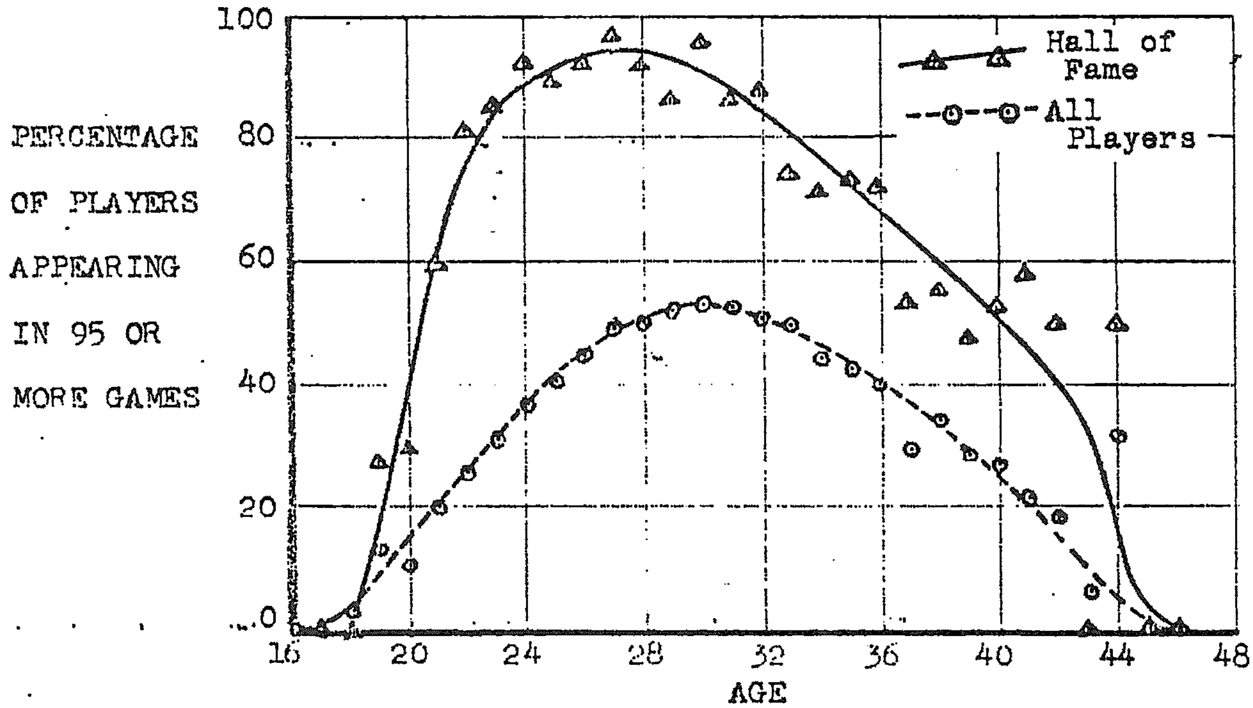


FIGURE 2

GAMES PLAYED AT EACH AGE AS A PERCENTAGE OF TOTAL GAMES PLAYED

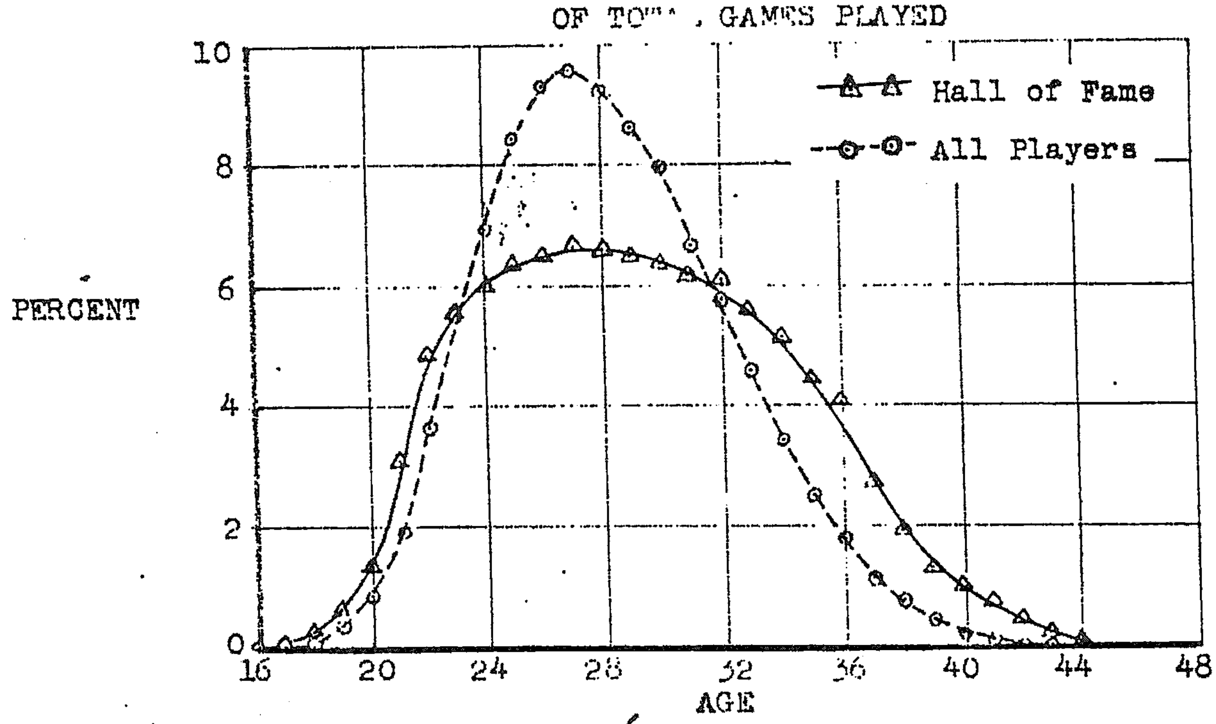
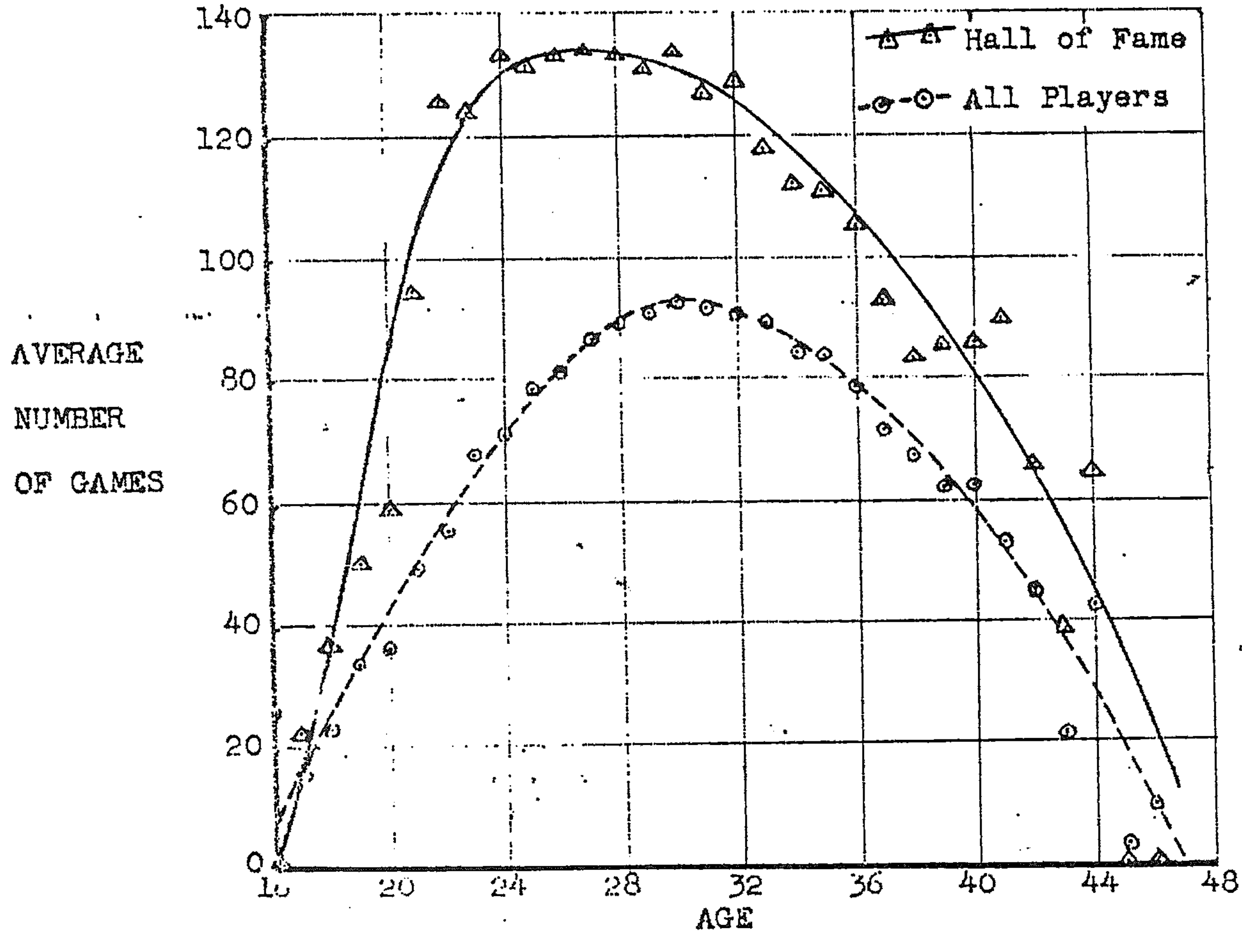


FIGURE 3

THE AVERAGE NUMBER OF GAMES PLAYED AS A FUNCTION OF AGE



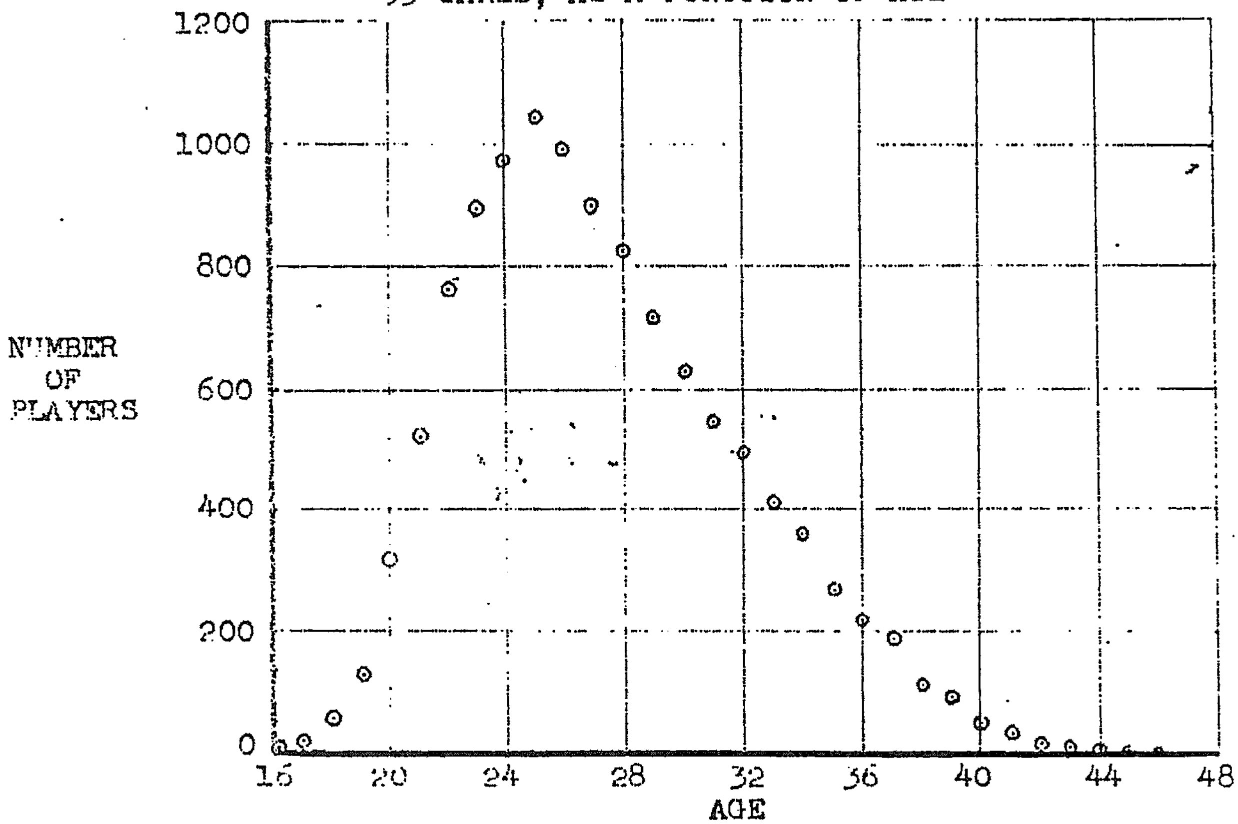
After age 30 however, most of that group of players who had been playing regularly, and who had been playing more games each year (see Figure 3), now begin to play less each year. This results in a steady decline in games up to age 34 (see Figure 2). After 34, even the best players (see the Hall of Fame curve on Figure 2) begin to play increasingly fewer games. This starts an acceleration in the overall decline. Now examine Figure 2.

Notice particularly how the decline among all players after age 33 parallels the decline in Hall of Fame players from age 35 onward. It would appear that about two years of talent erosion due to normal aging is required to reduce the skills of the greatest players down to the level of more ordinary veteran players. This suggests a qualitative estimate of the difference in the inherent baseball skills between the very best players and the "good enough to play regularly but not good enough for the Hall of Fame" group of players.

To speculate, perhaps it could be said that a veteran player with Hall of Fame credentials is still as good a ballplayer as the typical major leaguer two years his junior. Or perhaps a young player who will prove himself good enough for Cooperstown, is as skilled as his more normal contemporaries would be if they had continued to develop their skills for two additional years. Continuing the trend of this latter conjecture, Figure 3 shows that most Hall of Fame players had established themselves as regulars by age 22, or by age 24 at the very latest. From Table 3, it is seen that, of the 77 Hall of Fame Players in the study, 48 were playing in 95 games or more at age 22; and 64 were playing that many games by age 24. If it is assumed that such players begin developing their big league skills at age 19 or 20, then their development period would cover about three or four years. If less gifted players would need an additional two years of development to become as talented, a total of five or six years perhaps; then it might be concluded that future Hall of Famers are about 50% more talented than the average ballplayer. That conclusion is very speculative, I admit, but never having seen any other estimates to compare mine with, I feel 50% is not unreasonable.

FIGURE 4

THE NUMBER OF PLAYERS APPEARING IN FEWER THAN 95 GAMES, AS A FUNCTION OF AGE



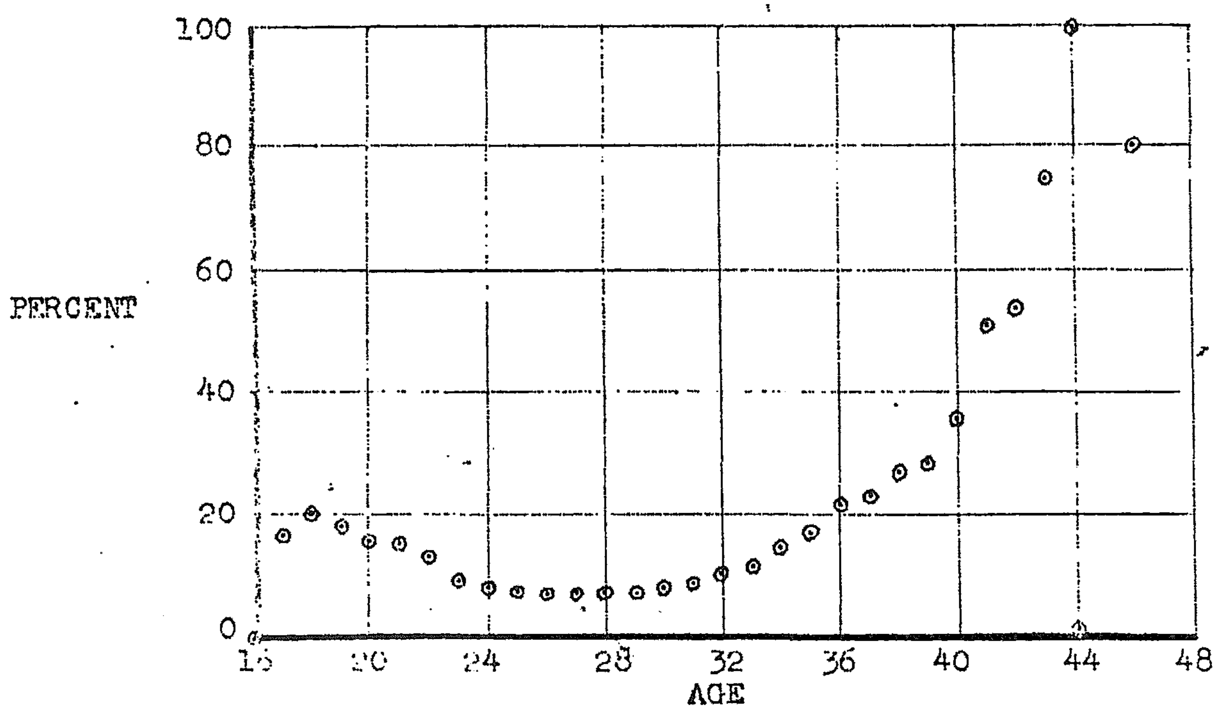
Just as Hall of Fame players last longer, they also begin sooner and establish themselves sooner (see Figure 3). This can also be seen on Figure 5 which portrays the games played by future Hall of Fame members as a percentage of total major league games. Nor is it surprising to learn that the greatest players average approximately the same number of games, about 130, each year from age 23 through age 32 (see the Hall of Fame curve on Figure 3). A Hall of Fame player stays near his peak and plays full time for roughly a ten year period.

If it is assumed that a manager tries to play his best players as much as possible, then it is likely that the average major leaguer reaches his peak at age 27. Why else do 27 year olds play more games than any other age? Since managers over a 68 year period have preferred to play 27 year olds (in a total of 152,490 games) more than, for example, 26 year olds (148,330 games) or 28 year olds (147,150 games), it must mean that those managers felt the 27 year olds were better players.

How long an average player remains near his peak is somewhat difficult to decide. Figure 3 does not show as well-defined a peak period for average players as it does for Hall of Fame players with their ten year period. A reasonable judgment might be that normal players remain near their peak from age 27 through age 33; a seven year period. But note that the average player appears in about 90 games per year during this seven year interval; a Hall of Fame player gets into 130 games a year during his 10 year peak period. If we ignore the difference in games per season and consider only the seven and ten year lengths, then the three year difference in length is roughly (and maybe only coincidentally) consistent with the earlier conjecture that there was a two year talent differential between the two groups.

FIGURE 5

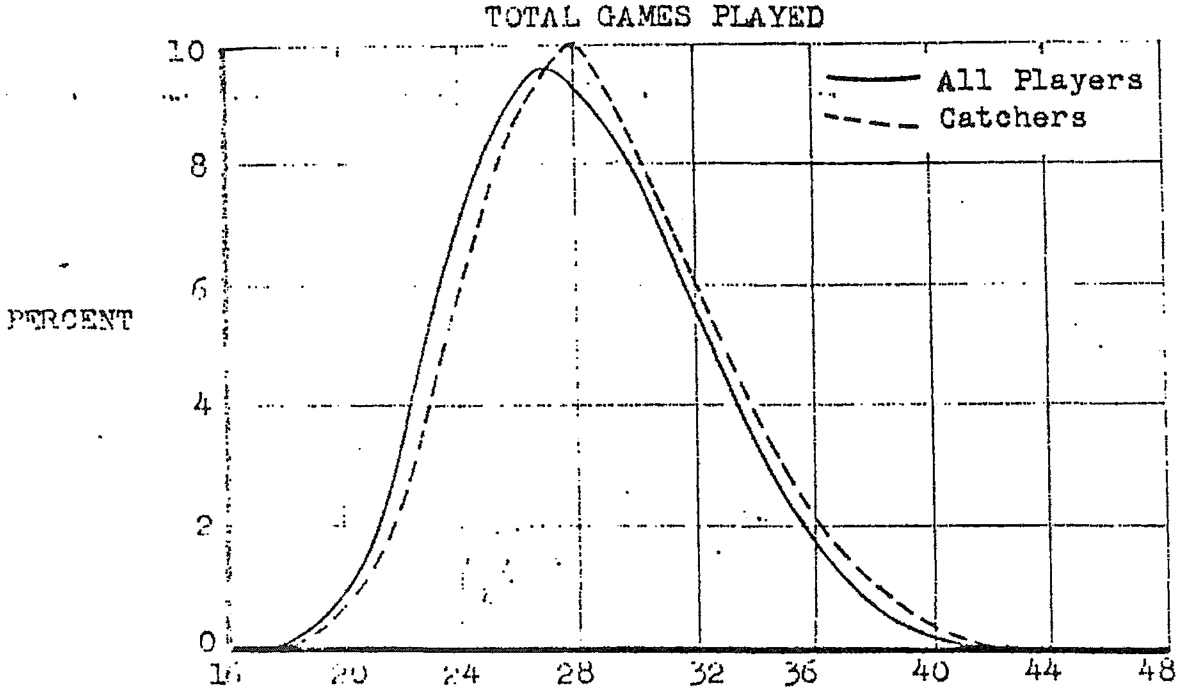
GAMES PLAYED BY HALL OF FAMERS AS A PERCENTAGE OF TOTAL MAJOR LEAGUE GAMES



Lastly a few observations about catchers. This position demands mental development more than physical skills. It is not a requirement that a catcher be able to run or to hit. Not surprisingly the mental demands of catching require longer to develop (by about one year it turns out) than do the physical demands of infielders and outfielders. Figure 6 reproduces the All Players curve of Figure 2 and adds to it a curve of games played by catchers at each age as a percentage of total games by catchers. The two curves are nearly identical, except that the catchers curve is shifted about one year to the right. Catchers, as a class, come up one year later than infielders and outfielders, reach their peak one year later and last one year longer. But the rate of development and the speed of decline for catchers are essentially no different from those of other players.

FIGURE 6

GAMES PLAYED AT EACH AGE AS A PERCENTAGE OF TOTAL GAMES PLAYED



BALLPARK EFFECTS ON FIELDING PERFORMANCE: FURTHER EVIDENCE

Craig Wright
Sabermetrician
Texas Rangers

Paul Schwarzenbart's article on ballpark effects on double plays and errors was interesting to me as I have done similar studies myself. His basic conclusion that fielding percentages are higher on artificial turf is, of course, backed by every study I've seen on the subject. The spread of his evidence for infielders is a little large, though. You may have noticed in your own studies of fielding statistics that there are overall more infield assists (minus double plays) on NL teams with grass home fields than those with artificial turf. More chances mean more errors for infielders.

I have yet to see a really solid study to show that more ground balls are hit on grass fields, but I believe the logic for it is there. I theorize that the National League is a fastball league because 50% of their fields are artificial and that the American League is a breaking ball league because only 21% (now 29% with the Metrodome) of their fields are artificial. Why? Breaking balls get hit on the ground which is more likely to result in an out on grass than on artificial turf where the ball is faster. I realize the batter also may be trying a little harder to put the ball on the ground for artificial turf, but the edge belongs to the pitcher I think.

The error factor for outfielders should favor the artificial turf outfielder a little bit more than Schwarzenbart's data suggests as by following the same logic we have more chances for the artificial turf outfielder. I am surprised at Schwarzenbart's reluctance to accept that artificial turf does positively affect fielding perdentages for outfielders. Besides the added accuracy to bounced throws, a true is <u>important</u> to an outfielder. My observation would be that an outfielder <u>running</u> to pick up a ball is less prepared to handle a bad hop than is an infielder. In fact, if I were to characterize the frequency of outfield errors, the order would be:

1) Errors charged to outfielders on their initial fielding of base hits,

2) Throwing errors, and

3) Dropped flies.

Schwarzenbart's study looked at double play frequencies without studying double play <u>opportunities</u>. The number of runners on first is affected by ballparks, too, and might also have been noted.

In my five-year study of ballparks (1976-1980), the six NL grass parks had significantly (.005) more runners on first per out, not a strikeout, than games played on the road involving the same sets of teams. This data is given in the chart at the top of page 12:

1976-80 Data for LA, NY, SF, Chi, Atl, SD

	Home Games	Road Games
Outs = AB - Hits	123,074	122,586
Strikeouts	- 24,351	- 23,899
Outs - Strikeouts	98,723	98,687
H + W - Extra Base Hits	46,991	45,872
Men on first per out not a strikeout	.476	.465

This evidence by itself would predict an increase of 2.3% in double plays in these grass parks versus their teams road games. Schwarzenbart's actual firgures for these six parks with grass fields show an average increase of only 1.3%. So in fact there are fewer double plays per opportunity in these parks than on the road. But the difference remains small, and like Schwarzenbart I am inclined to dismiss the influence of turf on double plays as insignificant.

But there is still that suggestion that ground balls are more prevalent on grass fields. If so, it may very well be true that there are significantly more double plays turned on artificial turf per ground ball per double play opportunity. I personally suspect that this is true, and, as mentioned, what data there is available hints that it is true. But we still wait on more controlled data to be reasonalby sure.

RUN PRODUCTION BY BATTING ORDER POSITION

Generally speaking, any leadoff batter who scores a run 14% or more of his total at-bats and walks, and a cleanup batter who has a 16% or better RBI mark on his at-bats, are doing a damn good job. This is a pretty good measurement for individual performance, but it really tells us little about what to expect from total team production.

A review of the 1981 box scores for both leagues sheds an interesting light on what to expect from each position in the batting order. This study is, admittedly, a small and perhaps distorted sampling, since it takes into consideration only one season --- and a shortened and split one, at that. But it yields data that should approximate, if not precisely isolate, the validity of its conclusions (A follow-up study will be made of the entire 1982 season). The following listings show the percentage of RBI production for each position, and does not include RBIs accounted for by pinch hitters or substitute players who bat in positions other than their predecessors since their totals would distort the normal rhythm of the batting order philosophy.

NATIONAL

Pos	RBI	RBI %	
1	376	.0868	
2	394	.0924	
3	662	.1550	.3342
4	738	.1726	
5	652	.1529	
6	501	.1175	.4438
7	432	.1013	
8	366	.0858	
9	153	.0359	.2230
	4265		
PH & Subs	423	.0992	

AMERICAN

Pos	RBI	RBI %	
1	466	.0887	
2	549	.1045	
3	738	.1404	.3336
4	834	.1587	
5	681	.1296	
6	619	.1178	.4061
7	530	.1008	
8	478	.0909	
9	361	.0687	.2604
	5256		
PH & Subs	456	.0868	

Note the DH effect on the distribution of RBI production. Whereas the middle of the NL batting order accounts for 44% of production and the bottom third for 22%, the AL equivalents show a corresponding loss and gain, respectively.

Looking at these percentages in terms of "quotas" to be achieved, we can carry it one step farther and assess a player's contribution in terms of how well he approaches, meets or exceeds these standards, proportioned of course, to the number of games he plays in in respect to his team's total games. The problem is that relatively few players remain in one batting order slot consistently. And yet some players do spend a great majority of their time in one spot ---- Garvey, Murray, Simmons, Parker, etc. We've ranked the following players according to their plus or minus percentage in meeting the above expectancy rates. And as a not entirely parenthetical afterthought, it should be noted that on superior run-producing teams, the third and fourth place batter's RBI percentage could very well be less than expected since other positions will be producing at a greater rate than projected for them. (See Willie Stargell and Jim Rice)

Player	Bat Ord Position	Yrs Xper Consid	Proj RBI	Act RBI	Perf %	RBI Bat Ord %	AB/RBI %
Dave Parker	3	7	602	617	2.52	.1589	.1629
Dave Winfield	3 (1981) 4 all oth	8	674	682	1.16	.1746	.1583
Reggie Jackson	4	12	1192	1205	1.07	.1604	.1785
George Foster	4	7	742	749	.93	.1742	.1907
Willie Stargell	4	11	955	944	-1.16	.1706	.1939
Jim Rice	4	7	743	718	-3.40	.1533	.1780
Eddie Murray	4	5	493	476	-3.40	.1533	.1684
George Brett	3	8	654	621	-5.06	.1333	.1447
Mike Schmidt	4	8	872	823	-5.62	.1629	.1902
Ted Simmons	4	10	949	886	-6.66	.1611	.1622
Fred Lynn	4 (1981) 3 all oth	7	581	542	-6.70	.1310	.1664
John Mayberry	4	9	783	721	-7.94	.1461	.1633
Steve Garvey	4	7	751	683	-9.04	.1570	.1531

All figures shown above include the 1981 season.

Comments:

Steve Garvey - His teammates claim Reggie Smith is a better cleanup batter because he hits with a lower AB-HR ratio and a higher extra-base hit percentage. Vin Scully says Lasorda likes to use Garvey on the hit and run, something he cannot do with Smith. This thinking supposedly accounts for Garvey's poor W-SO ratio, and his less than average RBI percentage for a cleanup batter. Does this make sense to anyone?

Dave Winfield - Anyone who says he isn't deserving of being the highest paid player has only to check his two RBI percentages shown above for starters. He is the only player listed who has a higher batting order RBI% than AB/RBI% who has played for a perennial non-contender. And also note that his RBI% in both departments is higher than Steve Garvey's. At a cocktail party the other day, a local sportswriter was belittling Winfield's ability citing his relatively poor W-SO ratio. It sounded like someone criticising Brahms because he didn't write an opera.

Check the RBI differentials between Rice and Murray. Is this a reflection of RBI timliness? In the last 5 years since Murray became a regular, Baltimore has had only 88% of Boston's RBI total, yet Murray has accounted for 95% of Rice's total RBIs.

Steve Garvey - Part II - Whatever was said above about Reggie Smith would apply equally as well to Ron Cey. The inescapable fact is that Garvey should be hitting in the #3 slot.

CLUTCH HITTING

Re: pp 134-140, 1980 Abstract

John C. Tattersall's 1975 magnum opus, The Home Run Handbook, discloses that a home run is worth 1.64 runs (54% with the bases empty, 31% with one man aboard, 12% with 2 men on base and 3% with the bases loaded). Then he shows us (pp 104-114) a breakdown of all home runs hit by players with a grand total of 100 or more through the 1974 season.

A player with a relatively low percentage of bases empty home runs must, in a small way at least, be hitting homers in a more timely fashion than someone with a high percentage of solo shots. Surely, we'd all admit that the timing of home run hitting isn't entirely random. The figures that Tattersall uncovered pretty well advance this proposal. And as a recently retired Quality Control Chief Inspector for the El Paso County White Slavers Association, believe me when I say, I can, on occasion, recognize good figures.

If we take the 1.64 average runs produced by the home run and compare our inactive players' records with that mark, we can partially assess their slugging timliness, if nothing else. This comparison, as well as the percentage of solo shots, can serve as rudimentary index of their "clutch" homers, regardless of the context of their occurrence, and become even more meaningful when we look at their position in the batting order.

I think we'd agree that a player with a strong hitting team would have more opportunities to hit more home runs with runners on base than a similarly skilled player with a less than proficient offensive club. So when we see someone like, say Harlond Clift (1.674), batting cleanup with teams that never finished better than sixth place, one can't help but wonder if his clutch slugging wasn't considerably better than Duke Snider (1.592), who played with perennial contenders. Nevertheless, the real key to performance appears to ly in the position the batter occupies in the batting order as is evidenced by averaging out the cumulative run production for those players appearing in Tattersall's select list.

Bat Order Pos	Runs per HR	% of Cleanup Batter
1	1.478	.85
2	1.432	.83
3	1.664	.96
4	1.731	---
5	1.588	.92
6	1.575	.91
7	1.570	.90
8	1.524	.88

Virtually all players move around in the batting order depending on that stage in their careers when they are either acquiring or losing skills; the whims, intuitive revelations or occult divinations of management; trades or injuries. Joe DiMaggio batted 3rd and 4th, so did the Babe. So we simply consider their totals in that position where they batted most of the time. Quite arbitrarily, the players are shown ranked according to their low home run frequency rate.

Player	Bat Order Position	AB/ HR	Bases Empty HR %	Runs by HR
Babe Ruth	4	11.76	.49	1.693
Ralph Kiner	4	14.11	.51	1.688
Harmon Killebrew	4	14.22	.48	1.692
Ted Williams	4	14.79	.45	1.760
Mickey Mantle	4	15.12	.55	1.606
Jimmie Foxx	4	15.23	.48	1.768
Willie McCovey *	4	15.73	.54	1.651
Hank Greenberg	4	15.69	.48	1.737
Lou Gehrig	4	16.23	.47	1.773
Hank Aaron *	4	16.38	.52	1.653
Willie Mays	3	16.49	.55	1.574
Hank Sauer	4	16.65	.53	1.628
Eddie Mathews	4	16.67	.53	1.602
Willie Stargell *	4	16.68	.50	1.702
Frank Howard	4	16.98	.53	1.607
Frank Robinson *	4	17.08	.55	1.591
Reggie Jackson *	4	17.10	.61	1.528
Roy Campanella	5	17.38	.46	1.760
Rocky Colavito	5	17.39	.51	1.682
Gus Zernial	5	17.43	.48	1.738
Duke Snider	4	17.59	.54	1.592
Norm Cash	4	17.79	.56	1.552
Johnny Mize	4	17.95	.48	1.710
Richie Allen	4	18.03	.52	1.639
Ernie Banks	4	18.40	.51	1.688
Mel Ott	4	18.50	.46	1.728
Roger Maris	5	18.55	.54	1.615
Joe DiMaggio	3	18.89	.47	1.765
Gil Hodges	5	19.00	.53	1.695
Wally Post	5	19.08	.55	1.623
Johnny Bench *	4	19.30	.50	1.703
Hack Wilson	4	19.51	.48	1.713
Bob Allison	5	19.66	.57	1.574
Joe Adcock	4	19.66	.54	1.631
Boog Powell	4	19.71	.47	1.703

* Record through 1974 only.

The top ten leaders in run production by means of the home run are:

Ty Cobb	1.915
Chet Laabs	1.846
Hank Leiber	1.842
Travis Jackson	1.800
Zeke Bonura	1.798
Granny Hamner	1.798
Rudy York	1.791
Dixie Walker	1.790
Vern Stephens	1.785
Heinie Manush	1.782

Curiously, only four of the above top ten batted regularly in the cleanup slot.

In Search of the "True" Slugging Percentage

by Jim Morrow

One all-time favorite baseball arguments is the offensive value of different players. Cobb vs. Ruth, DiMaggio vs. Williams, Schmidt vs. Brett, these comparisons draw much dispute and more statistical indices to justify one or the other player's superiority. Batting average, slugging percentage, and runs created per out expended (for the sophisticated) are all constructed measures which suffer from a similiar problem; the weights attached to outcomes (singles, walks, homeruns, etc.) are all set arbitrarily. Obviously, every hit is not worth the same amount as batting average assumes, but is a homerun worth more than, less than, or exactly two doubles?

To answer this question, I examined the batting records of every National League hitter in 1979 who had 100 or more plate appearances (except for Phil Niekro, who I arbitrarily left out), a sample of 156 players in all. Using multiple regression, I tried to explain each player's run production, measured by runs scored plus rbi's divided by plate appearances, by his batting statistics, in terms of singles, doubles, triples, homeruns, and walks per plate appearance. Multiple regression, for the statistically uninitiated, fits a line (or plane) through a set of data to give the best fit as measured by the sum of squared errors. The estimated coefficients have a particularly simple interpretation in this case; the coefficient gives the number of runs scored and driven in on average each time a player gets that type of hit. For example, if the coefficient for singles production is .5, each time a player hits a single, he should score or drive in one-half of a run. By comparing, the coefficients for different types of results, we will have a direct comparison of the values of different types of hits. Finally, as the dependent variables are heteroskedastic, GLS was used, correcting for the number of plate appearances. In layman's terms, the runs-produced statistics for players with 600 plate appearances are less affected by chance than those statistics for players with 100 plate appearances. To compensate for this, we weight the former set of statistics more heavily when we calculate the estimates.

The results are as follows:

$$(1)\quad \underset{}{RP} = \underset{(.02)}{.00} + \underset{(.08)}{.56}\ SP + \underset{(.16)}{.86}\ DP + \underset{(.36)}{1.68}\ TP + \underset{(.13)}{2.74}\ HRP + \underset{(.06)}{.26}\ BBP \qquad R^2=.78$$

where RP, SP, DP, TP, HRP and BBP are runs produced, singles, doubles, triples, homeruns and walks per plate appearance. The numbers in parentheses beneath the coefficients are the standard errors of the coefficients. R^2 is a summary statistic for goodness-of-fit; it gives the proportion of variance in run production explained by the regression. Two more regressions were run to separate the effects on runs scored and runs driven in:

$$(2)\quad RUNP = \underset{(.01)}{-.01} + \underset{(.06)}{.31}\ SP + \underset{(.12)}{.49}\ DP + \underset{(.29)}{.58}\ TP + \underset{(.09)}{.97}\ HRP + \underset{(.02)}{.13}\ SBE + \underset{(.05)}{.21}\ BBP \qquad R^2=.55$$

$$(3)\quad RBIP = \underset{(.01)}{.02} + \underset{(.06)}{.16}\ SP + \underset{(.14)}{.43}\ DP + \underset{(.30)}{.40}\ TP + \underset{(.11)}{1.81}\ HRP \qquad R^2=.68$$

SBE is stolen base efficiency which is the number of steals divided by the number of times on first (singles and walks). I chose this measure for stealing as a way to look at the frequency of successful steals as a producer of runs. Of course, this measure ignores times caught stealing.

To interpret the results, examine the coefficients for homeruns first. Each homerun produces one run scored and around 1.75 rbi's for the batter. When a homerun is hit, there are .75 men on base on average, according to the equation. Examining just equation 1, walks are worth roughly about half a single; what we should expect as for scoring runs, a walk is as good as a single, walks cannot drive in runs except when the bases are loaded which is rare, and there are just about as many rbi's as runs scored. Doubles are not worth as much as is generally thought. Triples are difficult to judge; because of the large standard error caused by their rarity and wide deviation among players' triple frequency, it is difficult to estimate where the true value of a triple lies. Adding the coefficients from equations 2 and 3 should give us roughly the coefficients for equation 1; for triples, it's not even close. My guess is the true value of a triple is around 1.3-1.5, which implies a runner on third will score 55 to 75 per cent of the time (remember that there are .75 runners on base on average who must be driven in by a triple). Another interesting conclusion we can draw from the results is that you need a success rate of about two-thirds when stealing to break even as the value of a stolen base is about one-half that of a man on first.

My main interest was to deduce rough weights to replace total bases used in slugging percentage to calculate an offensive value index. By eyeballing the results, I arrived at the following weights:

Walk-1 Single-2 Double-3 Triple-6 Homerun-10

To check the value of these weights, I calculated the index for each team in the majors in 1979 and 1980 and correlated the results with runs scored. Team batting average correlates with team runs scored at .84, slugging percentage at .90, implying that slugging pecentage is a better measure of offensive ability. My index, however, came in at a disappointing .89. Considering that it includes walks which slugging percentage does not, this result was a severe disappointment. Nevertheless, these results have convinced me that doubles are overweighted and homeruns underweighted by counting just total bases. The implications here for the runs created statisic should be noted.

This analysis also provides another look at clutch hitting; by examining the residuals, that is the differences between the actual and the predicted rbi production, we should be able to determine which players are clutch hitters. If a player drives in more runs than his statistics say he should, we can assume he is hitting in the clutch. Then if a player showed consistency from year to year in either direction, this record would establish him as either a clutch or choke hitter. As I worked with only one year, I do not have these results. I believe "clutch" hitting is only a reflection of small sample variation and expect such a study to show there is no significant consistency in clutch performance. Two points that bolster this argument are first that those players with high clutch scores are players with few plate appearances and good pinch-hitting records (i.e. Del Unser and Jerry White). We expect these players to outperform their statistics as they have better-than-average numbers of runners on base when they come to bat. Second, the R^2 of the regression is relatively high, implying there is little random variation in the sample (roughly the same as a correlation of .83 between the predicted and actual rbi's).

Finally, I would like to comment on one limitation of this study I am certain a number of readers have noticed. By working with individual statistics instead of team statistics, a bias may be introduced. As teams tend to place more powerful hitters in the center of the lineup, those hitters have better opportunities to score and drive in runs, inflating their statistics. These inflated statistics are what the estimates are based on; thus the estimates may overvalue power hitting. To some extent they probably do, but I tried to correct for this problem. By introducing a dummy variable for 3, 4 and 5 hitters, I hoped to pull out this effect in equation 2. In simple terms, I put in an additional variable to "tag" 3, 4 and 5 hitters and pull out the extra help they received by batting in the middle of the lineup. This dummy had no effect; it appears there is no great effect by batting in the center of the lineup.

In conclusion, the conventional total base valuations for hits undervalues homeruns and overvalues doubles. Walks are worth one-half of a single as we expect. By extending this study to larger samples of players, we can expect greater accuracy in our estimates and more confidence in the above conclusions.

THE EFFECTS OF OVERWORK ON ROOKIE PITCHERS

by Dallas Adams

A young pitcher's career can be significantly lessened if that pitcher is overworked as a rookie. That appears to be the most logical conclusion which can be drawn from the data presented below. In this study, a pitcher's "rookie year" is considered to be the major league season in which he pitches 45 or more innings for the first time. The pitcher's "second year" is considered to be the season following his rookie year, and so on.

I began by compiling rookie, second, third and fourth year major league data for all pitchers whose rookie season occurred in the interval 1969 through 1975. Later, to increase the data base, I expanded the study to cover all the rookie pitchers (with a minimum of 40 games as rookies) in the years 1964 through 1968. The pitchers were divided into three groups based on the number of games they pitched as rookies: fewer than 40 games, 40-59 games, and 60-79 games. This grouping produced the composite data shown in Table-1.

TABLE-1
MAJOR LEAGUE INNINGS PITCHED IN ROOKIE, SECOND, THIRD AND FOURTH YEARS

GAMES AS ROOKIE	ROOKIE YEAR INNINGS	SECOND YEAR INNINGS	THIRD YEAR INNINGS	FOURTH YEAR INNINGS
under 40 (1969-75)	21885	19769	19225	18549
40-59 (1964-75)	9676	8466	7339	6983
60-79 (1964-75)	2227	1778	1433	1468

As can be seen, each group shows a continuing decline in major league innings pitched as successive years occur. But the key point is this: the more games a pitcher appeared in as a rookie, the faster is that decline. This trend is seen more clearly in Table-2 where the Table-1 data are normalized as a function of rookie innings.

TABLE-2
MAJOR LEAGUE INNINGS PITCHED IN ROOKIE, SECOND, THIRD AND FOURTH YEARS, NORMALIZED TO ROOKIE INNINGS

GAMES AS ROOKIE	ROOKIE YEAR INNINGS	SECOND YEAR INNINGS	THIRD YEAR INNINGS	FOURTH YEAR INNINGS
under 40 (1969-75)	1.000	.904	.878	.848
40-59 (1964-75)	1.000	.875	.759	.723
60-79 (1964-75)	1.000	.798	.644	.660

The most apparent explanation of the above pattern is that excessive work early in a pitcher's major league career will result in shortening the career. The more games a rookie pitches, the fewer major league innings he can be expected to pitch over the next three years. Seemingly, the long-range physical, and perhaps mental, well-being of a young pitcher requires that he be nurtured carefully in his first major league season.