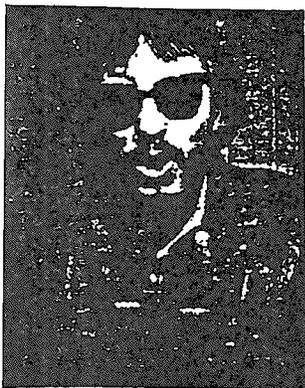


SAVE OUR GAME



Your Crusader

I, Ray Chegley, IV, have discovered an ugly truth about this thing we call baseball. It seems the men who run this game have taken it upon themselves to tamper with the very foundation of the game itself--the playing field! For years they have, in the name of profit, contrived to unweave the very fiber of our national pastime.

What is it that they are doing? For years I had been suspicious that the playing fields were shrinking. Things that once seemed far away were now somehow closer. I decided to investigate my suspicions. At a doubleheader in Detroit one 1975 day, I counted the number of seats at Tiger Stadium. I returned the next year and repeated the process. Much to my surprise, there were 18 more seats! I returned again in 1977 and my counting revealed an additional six seats. Where were they being put? Everything LOOKED much the same....and then a banjo-hitting shortstop powdered a ball over the centerfield fence and I realized the awful truth--my worst fears were an actual REALITY--the playing field was being shrunk to accomodate the extra seating for extra bucks in the pockets of the big league owners. Imagine my shock and dismay. How could they do this? Was this a local operation or a nationwide conspiracy? What could I do to stop it? I had to find out more....

Using my own funds, I undertook extensive travel to try and get the goods on the baseball bosses. I have been accumulating data on the subject: sabermetric proof that baseball records have been affected; measurements done with a ruler on the sly after hours; sworn testimony of "straight" groundskeepers; and extremely expensive aerial photography. I have amassed an impressive dossier on the subject, but have hit a brick wall in my quest to return the game to its natural state.

Simply put: I have exhausted the personal funds I had set aside for this task. I have the evidence; all I need are a battery of slick lawyers to bring this case to trial. That's where you, concerned baseball fan, come in. I need your financial support to make this crusade a reality. If you value baseball like I do, you'll see fit to send me the cash I need to carry on my fight. Look into your hearts! They're ruining your game America--BUT YOUR MONEY

CAN SAVE IT!!!



cut here
Yes!!! I want to save baseball from those that would do it harm. I believe firmly in the work that Ray Chegley, IV is doing on my behalf and in behalf of the game I love so dearly. I will do everything in my financial power to see to it that these bandits are stopped in their tracks. I do not want to have my children seeing major league baseball under a microscope by the year 2075. Go get 'em Ray Chegley, IV! and may God's speed be with you....

Please check appropriate donation box:

- | | |
|--|--|
| <input type="checkbox"/> \$10,000+ Saviour | <input type="checkbox"/> \$100 Keeper of the Faith |
| <input type="checkbox"/> \$5,000 Mahatma | <input type="checkbox"/> \$50 True Believer |
| <input type="checkbox"/> \$1,000 Saint | <input type="checkbox"/> \$25 Believer |
| <input type="checkbox"/> \$500 Fanatic | <input type="checkbox"/> \$10 Follower |

Please send CASH ONLY to Ray Chegley, IV c/o General Delivery/Juarez, Mexico. Do so in the knowledge that you are helping deliver your game from the very clutches of evil. Due to the covert nature of this program, donations are not tax-deductible.

Business Office:
Box 171
Winchester, Kansas 66097

Editorial Office:
20 Lincoln Street
Jersey City, NJ 07307
201-656-3892

IN THIS ISSUE:

Page

- 3 Letters
- 5 Base On Balls Abstract, II
.....Craig Wright
- 10,18 Dan Greenia's Freak Show
- 19 A Correction to "Some Further Aspects of the Distribution of Runs"
.....Dallas Adams

baseball-baseball ANALYST baseball-baseball

Editor: Jim Baker
Business Manager: Susie McCarthy
Statistical Consultant: John Borkowski
Founder/Publisher: Bill James

NOTES: PLEASE REMEMBER that all business aspects of the ANALYST are taken care of at the Kansas address listed above in care of Susie, and that all editorial comments, questions and submissions should be sent to the Jersey City address in care of John or me....don't forget to re-up with SABR, it's that time of year again...we're desperately hoping that Jersey City will build the stadium that has been threatened for New Jersey. We have picked out a spot about six blocks from here that we think would be ideal for major league baseball. It's a little cramped, but then so are Wrigley and Fenway, and everybody loves them....Jersey City's old International League team was called the Skeeters. Due to the intense pollution of the Meadowlands however, all the mosquitos are gone, so a new name will have to be in order...Craig Wright's impressive study on the walk continues in this issue and will conclude in February...Anyone wishing to attend the 1985 World Series at Shea Stadium with me, please contact me soon...can't see many tickets lasting past June....

LETTERS

Dear Readers,

Thanks to Dan Rappoport for commenting on our article on correlation and simple regression. We fear that careless use of the term "half" on our part caused the difficulty; we should have used the term "part." The denominator of the correlation equation has two "parts" to it, while the denominator of the regression formula contains only one of these two "parts", rather than "half" of them. Rappoport is quite correct in showing the implications of our wording.

We must, however, comment on the last paragraph of Rappoport's letter, where he warns us to "make sure that when one predicts performance, that the variables used are truly independent." The term "independent" has several interpretations, and we are unsure which of these interpretations Rappoport intends. If, by "independent" Rappoport wishes to caution us about "artifactual" relationships, or those involving variables which are actually alternative measures of the same thing, we are in agreement. For example, a correlation between games started and innings pitched would be uninformative, as they are both measures of pitching opportunity, while a correlation between hits allowed per inning and earned run average might be quite helpful, under the hypothesis that the former partly "accounts for" the latter. If, on the other hand, Rappoport means "statistical independence", then his claim leads to entirely different implications.

Charles Pavitt and Elaine M. Gilby

LETTERS SECTION CONTINUED

"LOOKING FOR MR GOODPITCH"

Dear Jim,

Since none of your reders ansered my challenge in the April '83 Analyst, Im gonna give you another zinger. Don't let me be underwhelmed with responses agin. Come on, you inellectual cadavers, speak up and speak out!

Coupla months ago I was visitin the General Custer monument at Little Big Horn Stadium, when I run into Professor Peter Shick A. Lee of the University of Southern North Dakota (or is it Northern South Dakota --- I never knowed since all my dealins with him was through his agent, B.D.Q. Bach of Euphoria, California). Anyway, Pete says to me, "Bert, I red your article in the Analyst and you got everything all screwed up." "How so, Pete," says I, measurin his physiogonomy for a tailor-made death mask. "You didn't say nothin about a pitcher's stuff and control, that's why, ding-bat." "Listen, dumb-dumb," I counters smartly, "BRA is control, H:IP is stuff. Ain't you got no smarts at all?" Well, he hems and haws and then puts his tail twixt his legs and skitters off without the curtsy of a fare-thee-well.

I kinda suspect most Savvymetricians would agree that pitchun effectiveness is based on two things: (1) good stuff and(2) good control --- plus a damn good defense. Now that that's settled, how often does a starting pitcher have good stuff? Good control? And not stand with my anser to Professor Pete, how do you define the two?

Two years ago, or somewhere in that nayborhood, Larry King aksed Warren Spahn on his Neutchooal talk show how ofen he (Spawn) thought he had good stuff. The crafty southpaw didn hesitate a minit before anserin, "75% of the time." Vin Scully aksed Joe Garage-io-la this past season how ofen he thought the average starter had good stuff. Joey ansered that he thought about 50% of the time. Terry Felton cited the Fifth Ammendmunt in declinin to anser.

We know a pitcher can win without good stuff, ifn he's lucky. He can also win without good control, but it ain't all that easy. How wood you define good stuff? Good control? I been ponderin this considerable.

How does this grab you? Good stuff is when the starter allows fewer hits than innings pitched. Thuswise, you include both power pitchers and finesse boys (cutey pies for you uninisiates) without resortin to a peck of strikeouts that some fellers use as a criteryon of "stuff". In the words of the immortal Festus Haggen, "Don't pay them no never-mind." They's good only to impress the baseball functional illiterates.

Now, what about good control. Well, yours truly says that any dude who allows 3 or fewer walks per 9 innings, qualifies. What about WP, PB, and HB? Nosireesir. Include them out. Why? Because I say so, that's why. They just gum up the works. Besides, ifn you got some gluteous maximus, stone-fingered backstop who couldn't catch my maiden aunt Tilly Grossbeak's fastball with a flutterby net, it ain't no pitchers fault. No, I like my KISS definitions, plus one other qualifier. Your startin pitcher has got to work a minimum of five (5) innings to be included in my survey.

Eyeball these 1983 stats:

- 21 % of all starters failed to go at least 5 innings.
- 57 % had good stuff (fewer hits than innings pitched)
- 39 % had good control (less than 3 walks per 9 innings)
- 26 % had good stuff and control simultaneously (that means at the same time)

These here stats is all well and good for Mr Joe Average, but what about the best pitchers? Based on W-L records? Minny on your tin-type. Bestest is by ERA, and I ain't slappin Phyllis Schlafly. now.

Looking at the top fifteen qualifyers in ERA leadership in both leagues last year, their performances show considerable likewise comparability:

	AL	%	NL	%
Number of games started	461		482	
Starters with 5 plus innings	408	.89	431	.89
Good stuff	243	.60	255	.59
Good control	205	.50	227	.53
Good stuff and control	128	.31	138	.32

Draw your own conclusions. Based on these one season excavations, the bestest pitchers have a much higher good control percentage than in the good stuff category above the average pitcher. Nothing out of the blus there, neither. These stats go a long way in provin the importance of good control , irregardless of how much smoke is thrown ---- Dewey Hoyt's 1984 performance, notwithstanding.

Only two of the NL starters (McMurtry and Rogers) could be considered cutey pies --- less than 5 Ks per 9 innings --- while 6 of the fifteen AL pitchers was (Honeycutt, McGregor, Haas, Young, Zahn, and Darwin). Which proves? Nothing. Cept that the saying that the AL is now a breaking-ball league while the NL is poplated by swifties may be just a tad true based on the better pitcher performances.

The best performance percentages in each of the four categories is shown down under for those fellers with 12 or more starts,

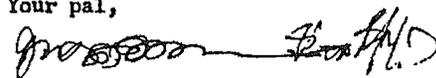
	5 Plus Innings *	Stuff	Control	Stuff & Control **
J. De Leon	100	Welch 85	Hoyt 91	Hoyt 55
Haymaker	100	Berangr 80	Cox 80	Candelabra 52
Sohio	100	Boddiker 77	Honeydew 77	Boddiker 50
Gidry	97	Penya 77	McGregor 77	Haymaker 48
Seaver	97	Ryan 76	Candelaba 74	Perez 46

* Four of these 5 fellers is NL boys --- DH effect dont't show up --- course, they's fuer runs skored in the NL, too.

** Their total W-L record is a super-mifty 46-9 (84%) in this here category.

Well, that's it, Jim.

Your pal,



Guthbert Magnolia, Esq., Ph D.
John Wesley Hardin Professor Emeritus
of Aberrant English and Lycan thropology
Southwest Correctional Institute

Craig Wright

BASE ON BALLS ABSTRACT, PART II

II. WALKS IN CAREER CONTEXT

A number of common theories about the base on balls have emerged from looking at walks from the perspective of career context. Again I have separated six of those traditional theories for study.

- A. Walk averages tend to be steady and characteristic of the individual during his career.
- B. In general there is a slight gradual rise in walk average over the career path.
- C. Players who already walk a lot in their early years experience less of a rise in their walk averages over their careers.
- D. Drawing walks is more of an innate skill, thus difficult to teach or learn.
- E. Power-hitters tend to experience a greater rise in their walk averages during their careers than other players.
- F. Some hitters are such good "bad-ball" hitters and have such a need to hit aggressively that their performance suffers when they walk more.

My investigation into these theories will not be as conclusive or as convincing compared to my research in Section I. My methodology for studying walks in the career context is complex and does not lend itself well to studying large samples of players. I produced individual career charts of 41 post-WWII players plotting their walk averages relative to the league mark against each player's age in that season. Chart 1 serves as an example of that form.

The selection of the players was not random. Most were selected because the length of their careers made them better subjects. That also means they were mostly star players as is usually the case for players who are major league regulars for a dozen years or so. Other players were selected simply because of my curiosity about some aspect of their career - hardly scientific. However, even with these weaknesses this study should be a major addition to the cursory observations and impressions that formed these traditional theories.

Are walk averages fairly steady and characteristic over the course of the individuals career? By my research it is a difficult call. Chart 2 shows three players who walked less than most and were pretty consistent about it. Chart 3 shows three others who maintained walk averages around the league norm their whole career, and Chart 4 is three more players, this time heavy walkers, who deviated relatively little over their careers.

The exceptions are equally plentiful. Chart 5 is three batters who raised their walk averages significantly during their careers. All three are power-hitters so it may be Theory E in action, but Chart 6 shows three players with average power following the same trend. Chart 7 is three players with the less common "down" trend. Actually, Dave Parker is demonstrating the "falling-off-the-cliff" technique. Chart 8 is three players demonstrating "I'm-going-up-but-its-hard-to-tell" form.

My version of Theory A would be, "Walk averages over the course of a career tend to be steady and characteristic of the individual for the near majority of players, but the exceptions are neither rare nor mild in their deviations."

Twenty of the 41 players in my study have charts complete from age 20 to 35 with minimal gaps due to injury or military service. Chart 9 has a composite graph of those players during that 16-year age period (center line) as well as two composite graphs which separate the group into equal halves according to their walk averages in their early years.

There are three things worth noting on Chart 9. First, notice that the peak walk averages come after age 30, past the physical prime years and what we know as generally the peak periods for baseball offensive performance. This is yet another piece of evidence that walk averages are not chiefly side-effects of superior offensive performance. Second, note that the walk averages are indeed rising over the course of the career. Third, the players who walked more in their early years continued to improve their walk averages as much or more than the other players.

It appears that Theory B is essentially correct in predicting a general rise in walk averages over the course of the career, although the description "slight and gradual" matches less closely to Chart 9. For reasons I will explain later, I do believe that the grade of advancement on Chart 9 is unusually steep if compared to a more average group of hitters.

Theory C seems totally contradicted by Chart 9. The group who walked more at an early age raised their relative walk average from a low of 1.38 to a high of 1.99 (+.61) while those who walked less in their early years rose from .55 to only .91 (+.36).

Theory C seems to have been based on the reasonable assumption that the performance has a ceiling and the closer you start to that ceiling, the less room for improvement. In reality, though, the ceiling is so far away from the walk levels of the majority of young players that there is plenty of room to freely advance. Of the 20 players only two (Morgan and Singleton) had relative walk averages over 2.00 at a young age. In their exceptional cases Theory C did seem to apply as neither raised their walk averages significantly over their careers.

Theory D's assumption that drawing walks is an innate skill that is difficult to learn or to teach is probably the most widely held and strongly believed of the six traditional theories being studied. Yet when you stop to actually think about it, doesn't it seem a bit ridiculous? Innate skill? Everything a human can do is partially differentiated by some inborn ability. The ability to learn is an innate skill. Is drawing walks more of an innate skill than those baseball performances relying on abnormal size, strength, coordination, speed, vision, or an ability to throw hard? Would you be more successful trying to teach people to run faster? How many baseball skills allow for the scope of improvement found in walk averages? Aaron as a 20-year-old minor leaguer has a .52 relative walk average and eventually he is walking twice as frequently as his league. Clemente never reaches a .75 relative walk level in his first 10 pro seasons and then the next

CHART 1

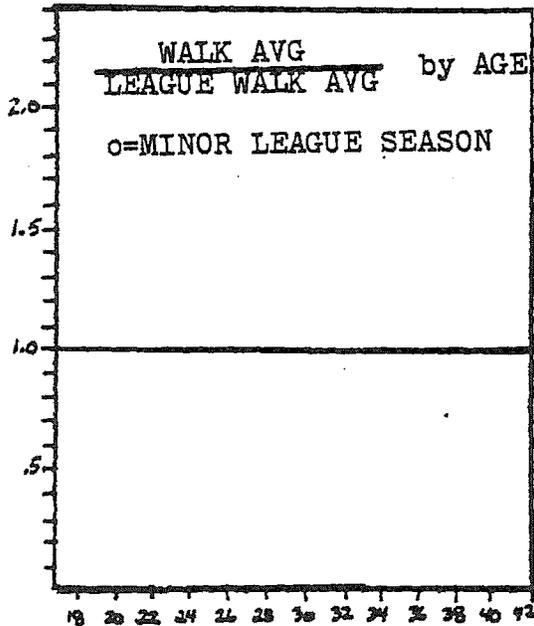


Chart 1

Minor League Season is where Minor League plate appearances out-number Major League plate appearances.

Age is based on July 1 of that Season.

Seasons of fewer than 100 plate appearances due to injury or military service are not plotted (Example, Mays at age 22 was in Military Service).

Back-to-back part-time seasons are averaged together and plotted between the two seasons (Example, Harvey Kuenn at ages 34-35).

In cases where a player is active in 1984 I have plotted his season as of July 15 if the player was on pace for 500+ plate appearances.

CHART 2

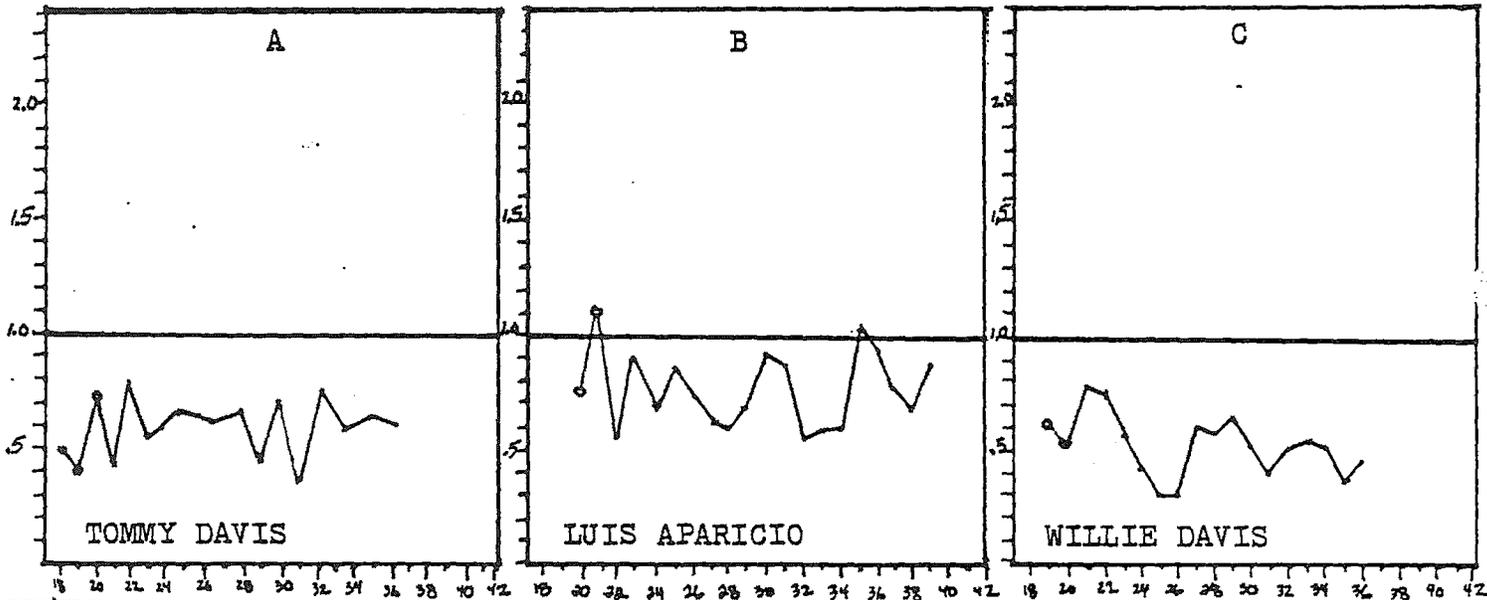


CHART 3

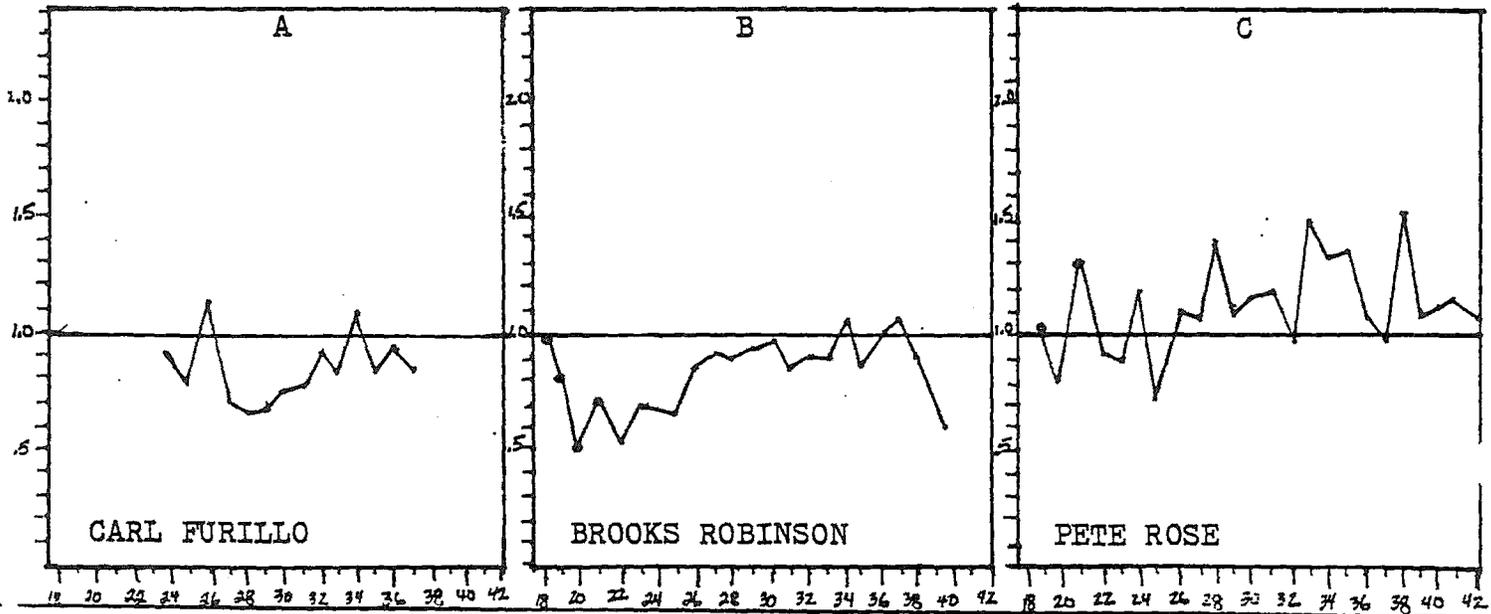


CHART 4

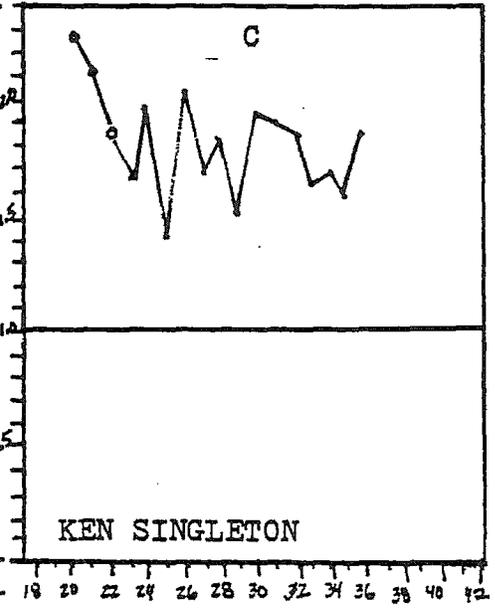
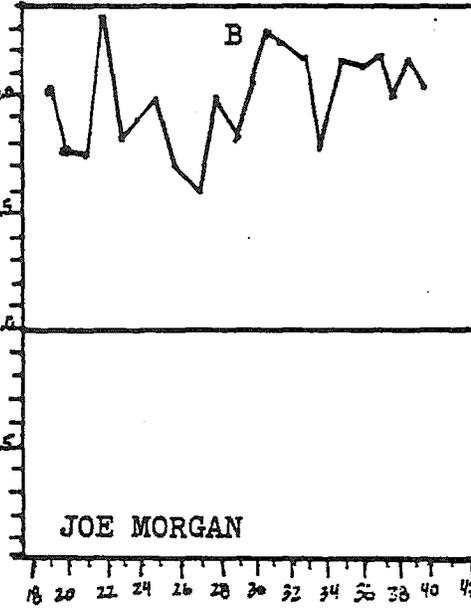
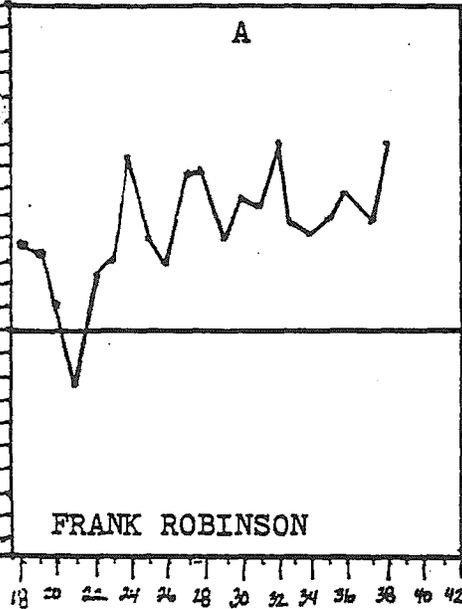


CHART 5

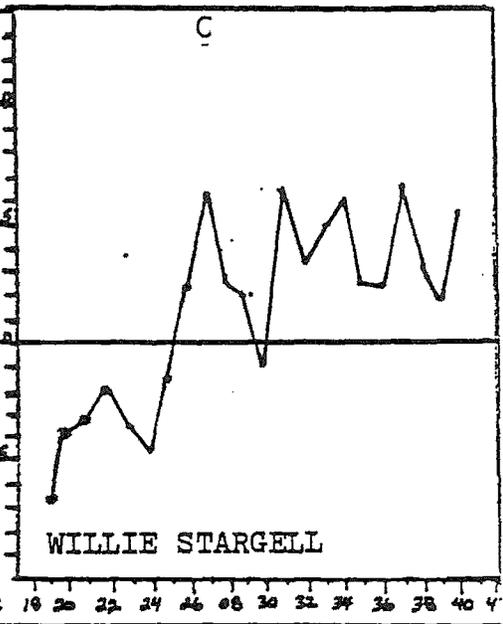
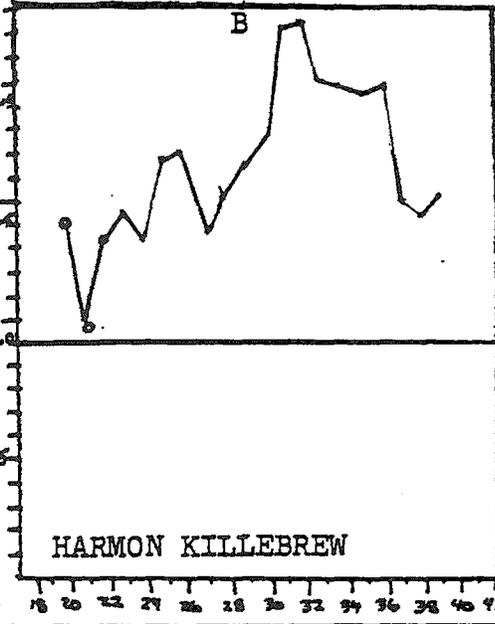
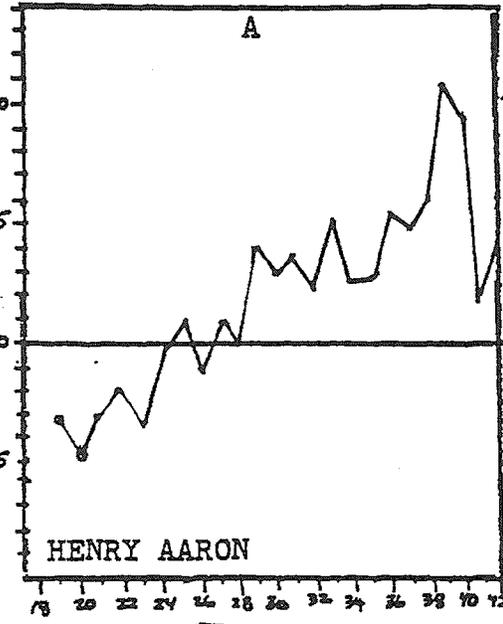


CHART 6

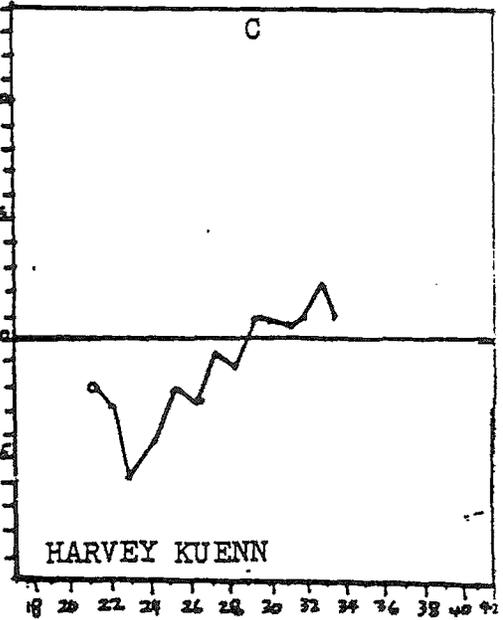
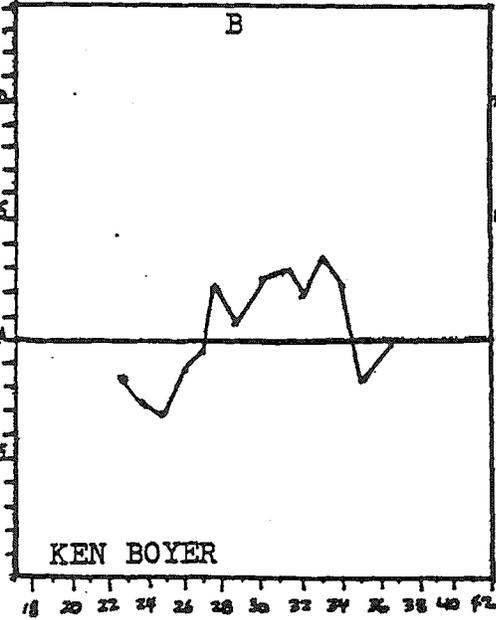
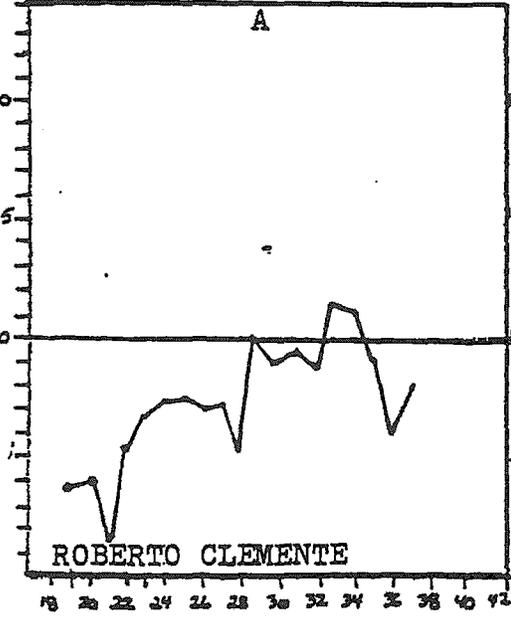


CHART 7

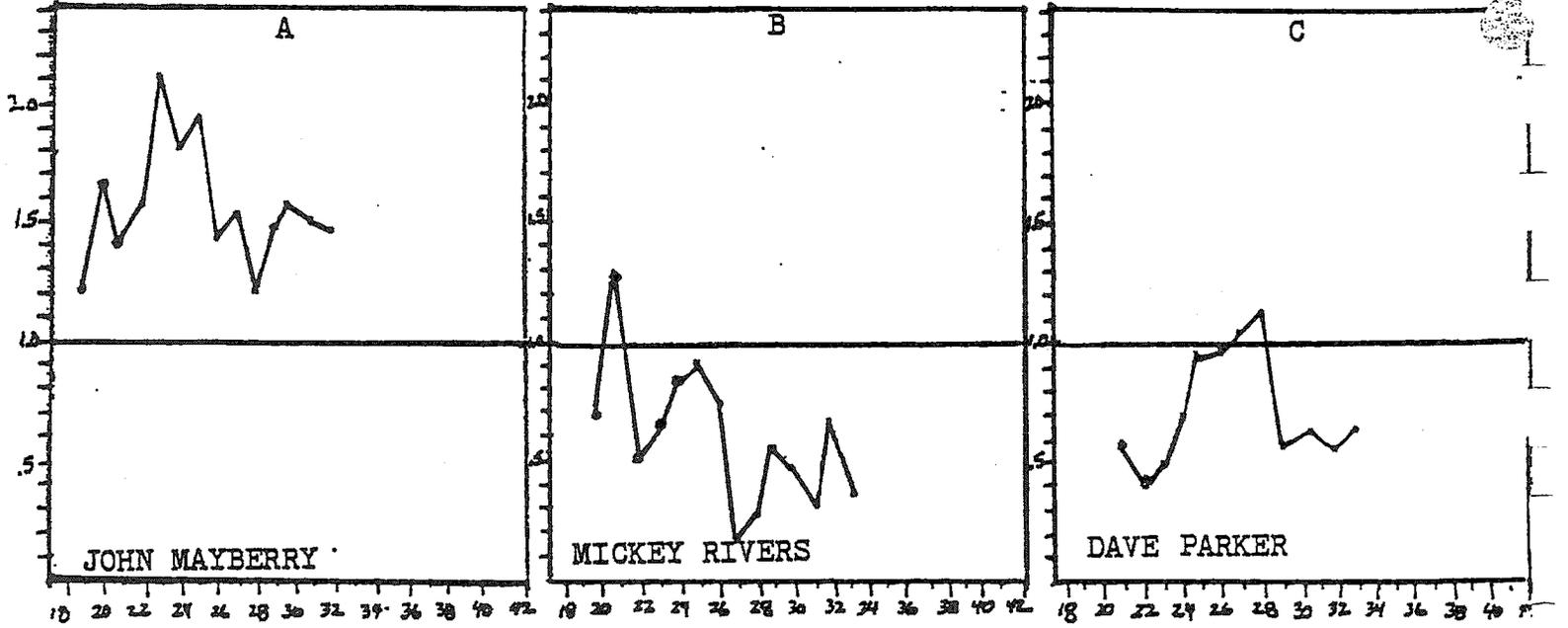
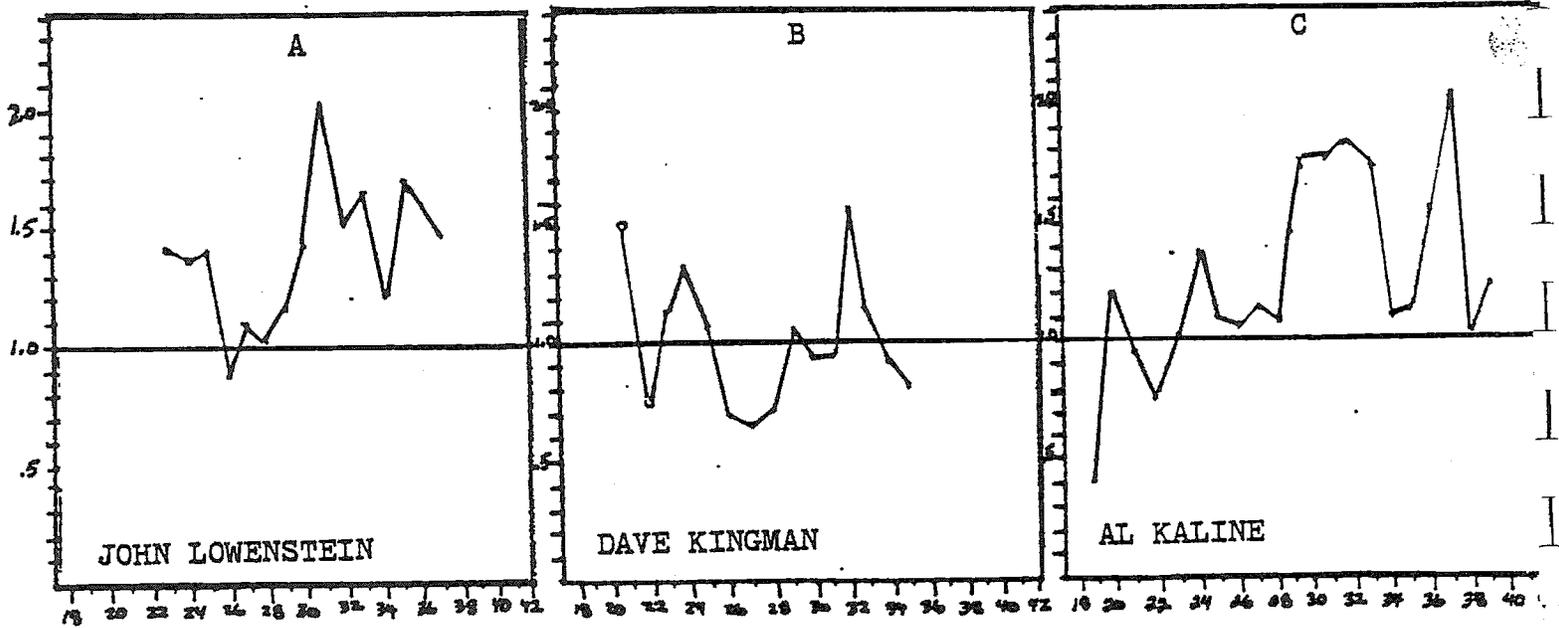


Chart 8



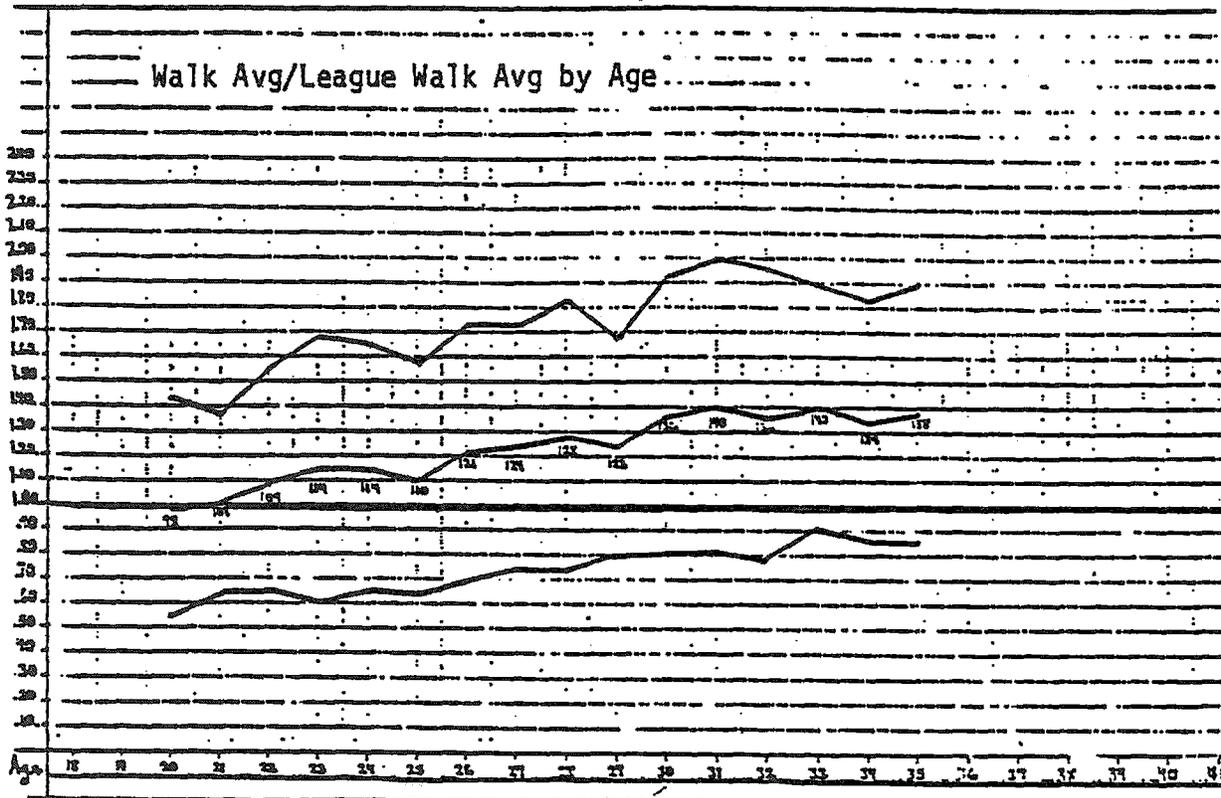


Chart 9

20 All-Stars from post WWII: Rose, Aparicio, F. Robinson, T. Davis, Aaron, Reese, McCovey, Clemente, Stargell, B. Robinson, Mays, Singleton, Mantle, Kaline, Killebrew, W. Davis, Snider, Yastrzemski, Oliver, Morgan.

Top Line: 10 who walked the most in their early years.

Middle Line: Entire group.

Bottom Line: 10 who walked the least in their early years.

Chart 10

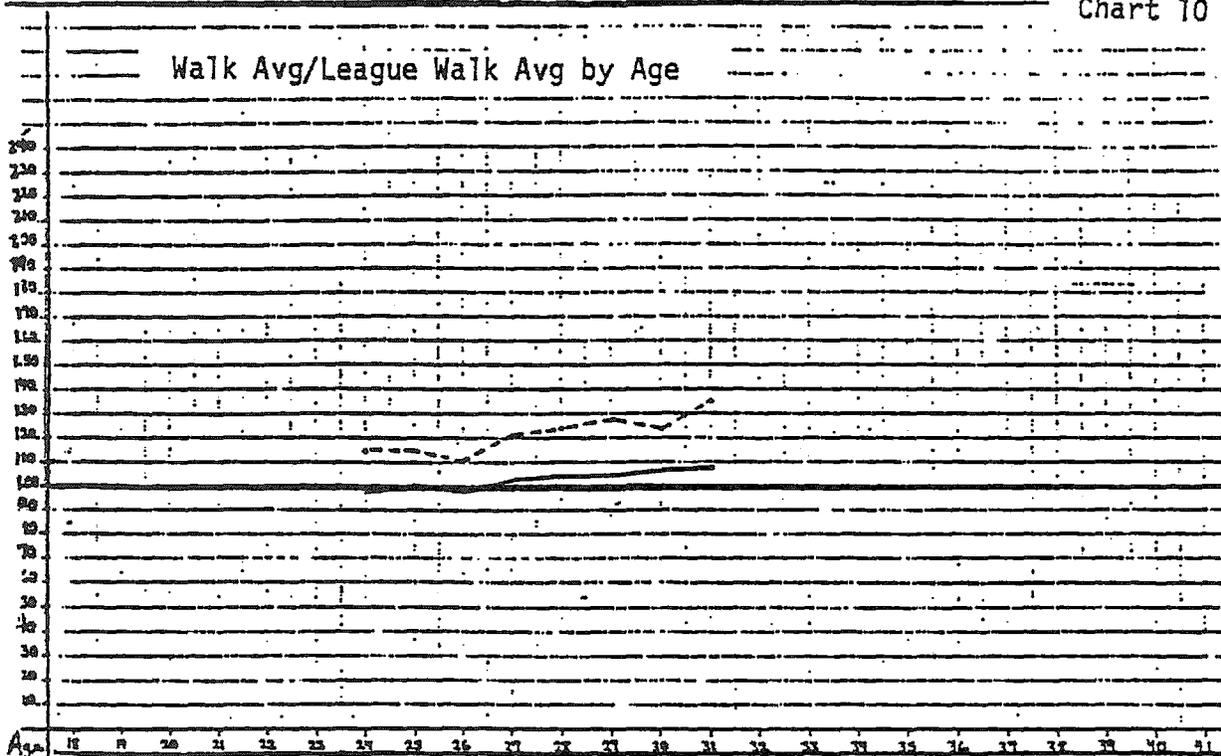


Chart 10

—: Group of 17: Buddy Bell, Ken Boyer, Billy Cox, Carl Furillo, John Groth, Gil Hodges, Dave Kingman, Harvey Kuenn, John Lowenstein, Roger Maris, John Mayberry, Dick McAuliffe, Dave Parker, Doug Rader, Mickey Rivers, Bobby Thompson, Ernie Whitt.

- - - : Group of 20 from Chart 9

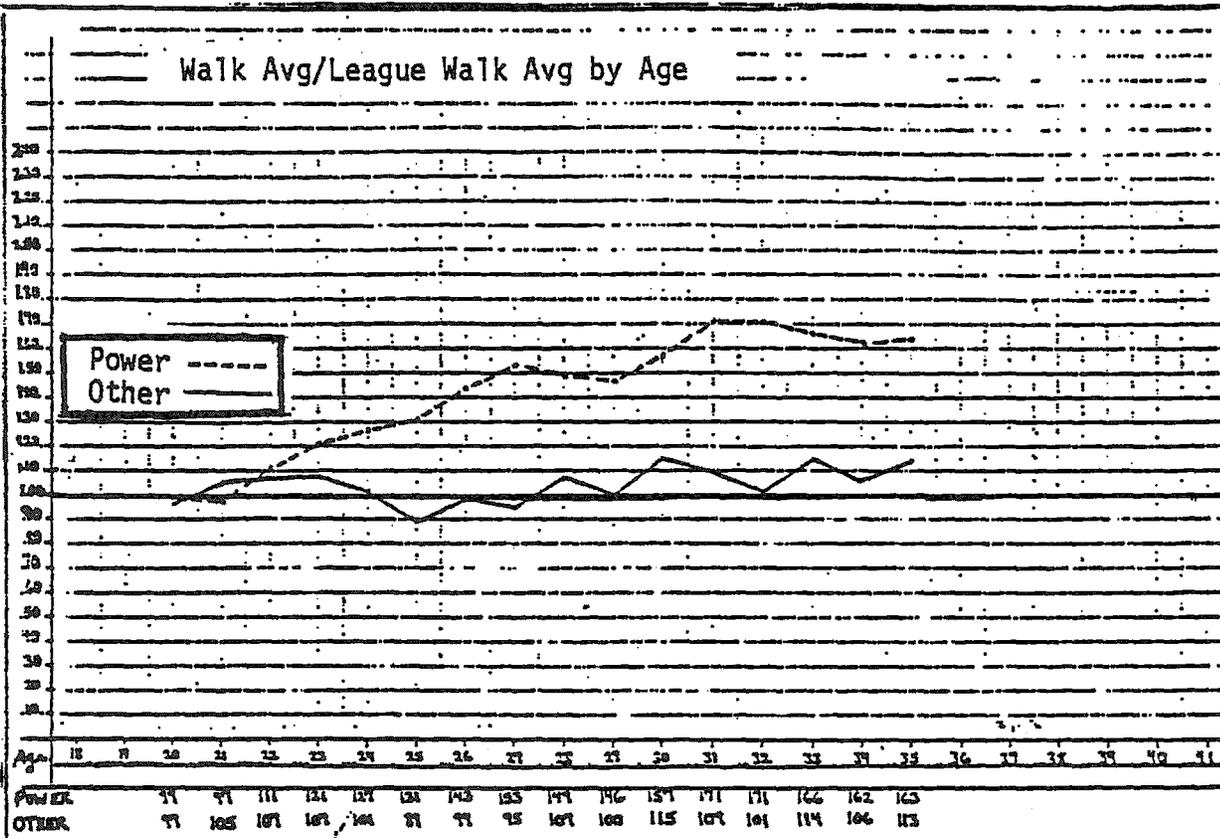


Chart 11

POWER GROUP: Frank Robinson, McCovey, Yastrzemski, Mantle, Kaline, Snider, Killebrew, Aaron, Stargell, Mays.

OTHER: Rose, Aparicio, Tommy Davis, Reese, Clemente, Brooks Robinson, Oliver, Singleton, Willie Davis, Morgan.

**Largest Metropolitan Areas With
No Local Major League Baseball
by Daniel Greenia**

Rank	Population *SMSA 1980	Minor League	ML Teams Within 50 Miles
7) Washington, DC	3,060,240		Orioles
11) Nassau/Suffolk, NY	2,605,813		Mets-Yankees
17) Newark, NJ	1,965,304		Yankees-Mets
21) Miami, FL	1,625,979	FL St.	----
22) Denver, CO	1,619,921	AA	----
24) Tampa, FL	1,569,492	FL St.	----
25) Riverside/SB/Ont., CA	1,557,080		Angels-Dodgers
26) Phoenix, AZ	1,508,030	PCL	----
30) San Jose, CA	1,295,071	Calif.	A's-Giants
31) Buffalo, NY	1,242,573	East.	----
32) Portland, OR	1,242,187	PCL	----
33) New Orleans, LA	1,186,725		----

Smallest Areas With Teams:

23) Seattle	1,606,765
27) Cincinnati	1,401,403
28) Milwaukee	1,397,143
29) Kansas City	1,327,020

*Standard Metropolitan Statistical Area: A large population nucleus together with adjacent communities that have a high degree of economic and social integration with that nucleus.

seven years he never falls below .85 and three times walks more frequently than the league. Kaline spends 4 years (ages 25-28) with relative walk averages from 1.05 to 1.14 then goes to a new career high of 1.44, then 1.72, then 1.75, and then 1.83. Stargell at ages 24-27 had consecutive relative walk averages of .52, .86, 1.22, and 1.62. Snider, Killebrew, and Mantle each raised their relative walk averages in 5 consecutive seasons with an average overall gain of +.80.

Drawing walks involves a complex interaction of two thinking beings affecting a larger contest between two teams. Drawing walks has not been associated with any physical skill. Clearly it is more a mental than physical adaptation. What possible reason is there to believe it cannot be a conscious mental adaptation?

If it seems that drawing walks is difficult to learn or teach, it is probably because that possibility has generally been denied. It is difficult to learn on your own and have to learn everything anew. It is difficult to teach with no field of knowledge or techniques to fall back on.

It is my conviction that drawing walks is a learnable skill, and that is also my basis for believing that the rise in walk levels for a more average group of players would not be as high as those plotted in Chart 9. I can conceive of players having such strong physical gifts that they can "star" despite personality or mental weaknesses, but to "star" over the long haul generally seems to require strength in all three areas. The group in Chart 9 strikes me as unusually fit to advance in this skill according to this new perspective.

Of the 21 players not plotted on Chart 9, seventeen have complete charts from age 24 to 31. Chart 10 identifies those 17, (clearly not the All-Star caliber of the group on Chart 9) and plots the walk averages of the 17 as well as the Chart 9 group for this 8-year age period. This graph shows the Chart 9 group is rising faster going from 1.14 to 1.36 (+.22) while the 17 went from .98 to 1.08 (+.10) during the same ages.

Chart 11 examines the claim of Theory E that power-hitters raise their walk averages more than others. Chart 11 clearly supports Theory E, and, although I have not included it as an extra chart, the same effect emerges from splitting the 17 players on Chart 10 (power-group gains .15 versus +.10 for the non-power group).

It is interesting to note on Chart 11 that the two groups had similar walk levels at a young age and really did not begin to separate until age 23. What is the basis of this relationship between drawing walks and power-hitting? Again, let me make the point that the walk level is chiefly a separate entity from the offensive threat of the long-ball. In the case of this particular group they averaged 32.3 homers per 550 at-bats during ages 22-24 which escalated, along with their walk averages, to a peak of 34.5 homers per 550 at-bats for ages 29-31. For ages 33-35 they were back to averaging 32.3 homers. The same home run threat existed at ages 22-24 as at ages 33-35 but the relative walk averages for those periods went from 1.20 to 1.64.

To understand why Theory E works one needs to recognize that the two performances are based on some similar skills. To hit for more power a batter generally needs

to be able to adjust not just mechanically but also to the strategy between him and the pitcher. That strategy involves patience, selectivity, knowledge of the strike zone, discipline, confidence, and calculated anticipation. How could such a player not learn to walk?

It would not be far-fetched to suggest that each assists the other. The homer rate should and does decline sooner because of its close link to physical gifts that would diminish with the earlier passing of the physical prime. The home run frequency of this group actually peaked at age 31 which is unusually late for a baseball performance relying on physical skill. Perhaps this reflects the portion of their power traceable to this approach which promotes both their walks and their power.

Next to the claim that you can't teach a player to walk, the most popular of the six theories is that an aggressive "bad-ball" hitter can hurt himself by walking too much. Has there ever been a free swinger who missed saying, "I'd hit .200 if I swung at just strikes", or something to that effect?

To check Theory F out we will examine the seven charted players who had at least four seasons where their walk average was less than 60% of the league average. The focus will be on the seasons that they walked the most.

First is Luis Aparicio. At age 30 Aparicio set a new career high in the majors for walk percentage (.90) relative to the league. That year he hit over his career average and slugged a career high 10 homers and also had his career high in P%. At age 35 he set a new career high in relative walk average (1.03) and also hit a new career high in average (.280). The next year at age 36 he had the second highest relative walk average of his career (.093) and stunned baseball by hitting .313 and slugging .404 with a career-high 29 doubles. This is all before Aparicio was traded to Boston.

Next is Tommy Davis who gives us little opportunity to see him as a walker. Tommy never walked more than .78 of the league average in either the majors or the minors. When he did top the .70 barrier he had his best minor league season and two fairly average major league seasons for the different stages in his career. He did have his best seasons in the .55-.59 range which was a little below average for his career.

Willie Davis is the next subject. Willie walked the most when he was in his first two full major league seasons. The first year he hit only .254 but had excellent power with 37 extra-base hits (12 homers and 6 triples) in only 339 AB. His sophomore season he hit .285, led the league in triples, and belted a career-high 21 homers. His next highest relative walk average would come at age 29 when he would hit a career-high .311. When his walks were lowest he hit .238 at age 25.

Roberto Clemente is an interesting case because of his reputation as the best "bad-ball" hitter in history. Roberto hardly walked at all in the minors or his first two years in the majors. In those first four pro-seasons he hit .256 in the minors with only 7 homers in 622 AB, and in the majors it was .285 with only 11 homers in nearly 1,000 AB (994). With his walk averages rising he hit .289, then .296, and never hit under .300 again. At the age of 29 Clemente established a new career-high relative walk average of 1.01 versus his previous high of .73. Over

the seven seasons from age 29 to 35 his walk average never fell below .85, and he enjoyed a phenomenal level of performance for a player his age. He took three batting titles (.339, .329, .357) set his career highs in doubles (40), triples (14), homers (29), runs (105), RBI's (119), BA (.357), and slugged .515 for the whole period.

Al Oliver is another interesting case particularly because of his astounding 1982 season. Until 1982 Oliver's relative walk averages tended to fluctuate from the low .50's to the mid-.70's and as an offensive force was a poor-man's All-Star, an outfielder who hit .300 with average power and never led the league in a single offensive category. For the record, he had been a slightly better player with his higher walk averages.

In 1982 a 35-year-old Oliver did two very surprising things, he drew a relatively astonishing number of walks and had a hell of a year. He hit a career-high .331 to take the batting title. He drove in 109 runs to lead the league. He also led the league in hits and doubles and set career highs in doubles (43), homers (22), and slugged over .500 for the first time in his career. That is not your expected season for a 35-year-old Al Oliver. But then neither was drawing 61 walks when his previous high was 40.

True, Oliver was walked intentionally 15 times but that is still 46 non-intentional walks, a 48% increase from his previous high in that category. The next season Oliver played three fewer games and saw his non-intentional walks plummet to a more normal 27. We also got a typical Oliver-year, .308, 8 homers, 84 RBI's.

Our next player in this category is Dave Parker who raised his relative walk average every year from age 22 to 28 which covers his climb to and reign as the best player in the game. Then the bottom fell out of both his walk average and his performance. In both categories it looked like 1975-1979 never existed.

The final player is the ever-confusing Mickey Rivers. Unlike the rest, Rivers actually walked quite a bit in some of his early seasons. At age 21 he led the Texas League with a .343 mark and a surprising 49 extra-base hits in just 449 AB. But the astounding fact is that Rivers also walked 65 times; his walk average was 28% above the league average. He earned a September trial and hit .320 in 25 at-bats. He also had 3 walks which is rather startling for a player who six years later would average 136 at-bats per 3 walks. After that season Rivers' walk averages were very poor his first two years at AAA. In his third year at AAA he began to walk more again (.85) and had a very successful September trial, hitting .349 and slugging .457 in 129 AB.

Rivers' best season for drawing walks in the big leagues was his first full season in 1974. It was also a good offensive year. He hit .285 in a tough offensive park and led the league with 11 triples while playing only 118 games. His last season in California involved another big drop in his walk average. He hit .284 and experienced a drop-off in power, his P% going from .107 to .074.

From ages 21 to 25 he walked at 21, 24, and 25 (1.01) and in the other three he did not (.64)

Major Leagues	AB	BA	P%
Ages: 21, 24, 25	620	.300	.106
Ages: 22, 23, 26	1,043	.268	.072

It is after age 26 that Rivers established his reputation as the toughest regular player to walk in the big leagues. He was successful, .301 and a .111 power percentage, despite a relative walk average around .40. He hit well in a very tough park when he walked the most, and he hit just as well when he walked the least. He did not hit well in his relative middle.

Overall Theory F appears to be backwards. In five of the seven cases (Aparicio, W. Davis, Clemente, Oliver, and Parker) the free-swingers actually hit significantly better when they were drawing more walks. The two exceptions, Tommy Davis and Mickey Rivers did not seem to be adversely affected when they walked more than usual.

After the results of the research into Theories E and F it is only natural to formulate and examine a seventh theory, Theory G, that "learning" to walk, i.e., increasing walk averages, enhances a player's other offensive abilities, prolongs their peak, and slows their decline.

Before investigating such a theory we need to be sure we are free of the "chicken or the egg" arguments. If we see rising walk levels keeping pace with improved performance, are we going to question Theory 1, the corrected theory which now states that drawing non-intentional walks has minimal correlation with superior performance as a hitter?

The major evidence for the correction was the fact that composite players A, B, and C (separated by descending non-intentional walk averages) were basically similar offensive players except when it came to drawing walks (Player A did have a slight home run edge). Just to be sure, we will repeat the study for the 1982 American League Season. This time there are 98 players with 400+ plate appearances; we will put 34 in groups A and C and 33 in group B. The 98 players averaged 569 PA and below are the statistics for the three groups projected into that average.

	AB	H	BA	2B	3B	HR	Non-Int. Walk	HRC%
Group A	501.4	138.0	.275	24.1	2.8	16.6	67.6	.165
Group B	524.8	144.8	.276	24.1	3.9	16.0	44.2	.163
Group C	541.7	144.5	.267	24.5	4.6	14.3	27.3	.149

Again composite Player A resembles closely Player B, but this time Player C is noticeably weaker than either A or B. This may be an unusual case of a number of weak performances netting 400+ PA that they normally would not earn.

Contributing to Player C's poor showing are:

	PA	BA	Slug %		PA	BA	Slug %
Brookens	432	.231	.352	Foli	528	.252	.308
Cabell	482	.261	.323	Milbourne	453	.257	.327
T.Cruz	519	.230	.376	Richardt	435	.241	.289

The next season only 2 of the 6 managed 400+ PA (Cabel cleared with only 419) even though the group's level of performance stayed much the same, .237, .335 slugging percentage.

Whatever, combining 1982 and 1983 the three composite players looked like this per 550 PA.

	AB	H	BA	2B	3B	HR	Non-Int. Walks	HRC%
Player A	485.9	132.5	.273	23.6	3.1	15.7	64.1	.161
Player B	506.5	140.5	.277	24.9	3.8	14.3	43.5	.164
Player C	522.4	141.2	.270	24.3	4.1	14.0	27.6	.153

The home run gap stands out a bit more, Player A 10% ahead of B, 12% ahead of C. The triples increase in the opposite direction, and Player C is weaker as a hitter than A or B. But the real key is that the difference in average and power between A and C cannot realistically lead one to believe that the pitchers would walk him two and a third times more frequently than Player C. The pitchers have no reason to walk Player A more than Player B.

Of course, one has to wonder if this same evidence argues against Theory G. Shouldn't there be more players with rising walk levels in group A than B or C and B than C, and if Theory G results in significant differences, shouldn't there also be larger differences between the three groups? One, we do not know the rate of occurrence or strength of Theory G to predict its impact in a large common grouping. Two, rising walk levels is a relative measure, thus, such players could be liberally distributed through the three groups although Group C would be expected to have the lowest concentration. Three, we have reason to assume that Players A, B, and C do not have similar natural abilities to draw on.

Triples are youth-related and that suggests that Player A may suffer an age handicap. It does indeed turn out that the average age of Player A is 29.9, Player B is 29.3, and C is 28.4. We know that, in general, offensive performance declines after age 28, so it is more difficult for Player A to match B's or C's performance if all other things are equal. The fact that A is able to keep pace or have superior performance to B and C is consistent with Theory G. The fact that Group C's performance is weaker than A's or B's is also consistent with Theory G.

The best evidence I can presently offer for Theory G are simply consistent examples, but they are rather powerful. Consider these two groups of players. Group I is Henry Aaron, Roberto Clemente, Harmon Killebrew, and Willie Stargell. Their impressive record of rising walk averages is documented on Chart 12. In assembling Group II, I wanted 4 players with long careers whose early performances marked them as comparable hitters (batting average and power) to Group I, but the Group II players would have relatively consistent walk averages over their careers. Fortunately, I was able to find a near perfect match in Frank Robinson, Tommy Davis, Joe Morgan, and Willie Davis. Their fairly constant walk levels are documented on Chart 13, and Chart 14 compares their composite relative walk

averages from their common ages of 20 to 36 (Willie Davis breaks the string by going to Japan at age 37). Note how Group I eventually outwalks Group II in 11 of the 17 ages, but never got closer than .17 to Group II in their first five common ages (20-24).

The similarities between the two groups include the facts that (1) each has 3 outfielders and one infielder, (2) each has a positive major home park change in mid-career, Aaron going from Milwaukee to Atlanta at age 32 and Morgan going from Houston to Cincinnati at age 30, (3) they have similar injury histories, Killebrew's elbow injury (1965) and ruptured hamstring (1968) combines with Stargell's broken ribs (1975) and knee problems (1972 on) to offset Morgan's minor knee injury (1968) and T. Davis' broken ankle (1965), and (4) the groups have similar composite physiques through age 25 according to the baseball register, (6' 1/4", 188 pounds versus 6' 0", 180 pounds). Because the Davis' did not continue to develop their power it is easy to overlook their size, but Tommy was 6' 2", 195, and Willie was 6' 2", 185. Despite their slugging feats, Clemente was listed as 5' 10", 175 pounds, (he was jumped to 5' 11" at age 27), and Aaron was consistently listed as 6' 0", 176 pounds, a full inch and 7 pounds less than Robinson.

Some readers with a cursory knowledge of Tommy Davis' career may object that his serious ankle injury affected his career to such an extent that he should not be included in this comparison. The general theory is that Tommy never came close to duplicating his superb 1962 season because of the ankle injury in 1965. However, if you look at Tommy's career record you see that his decline from his 1962 heights was well under way before the ankle injury. In 1964 he played a healthy 152 games and hit .275 with a slugging percentage of .397. Tommy's ankle was broken early in 1965, given over a year to heal, and his first season back he hit .313 in 100 games. His second year back he played 154 games and hit .302 with a career high 32 doubles and his 16 homers was his highest in any season other than 1962. Four years after the ankle injury he stole a career high 20 bases. I could accept that the severe ankle injury limited his speed and defensive ability over his career, but I cannot see it greatly hampering his ability as a hitter and barring him from this study.

Of greater concern to most readers will be my comparison of hitting ability between Group I and Group II. Looking at their careers in retrospect, Aaron was clearly superior to Robinson as was Clemente to Tommy Davis and Killebrew to Willie Davis, and when you realize we are not considering Morgan's walks, speed, or defense, Stargell is seen as far more valuable. Still, it is a fact that Group II was not only comparable to Group I through age 24, they were better.

Two players, A and B, both get to the majors at age 20. In that first major league season player A hits 10 points higher than B, has exactly the same number of doubles and triples and hits 25 more homers than B. After 3 years in the majors their records look like this:

	AB	H	BA	2B	3B	HR	Slug %
Player A	1,183	363	.307	56	11	67	.543
Player B	1,070	320	.299	65	15	40	.500

Chart 12

71

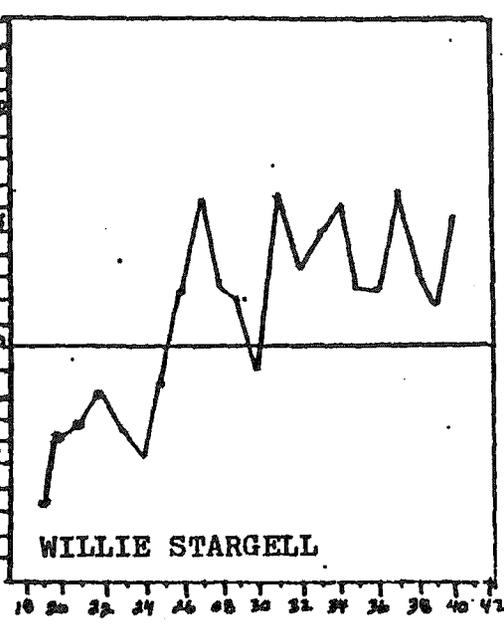
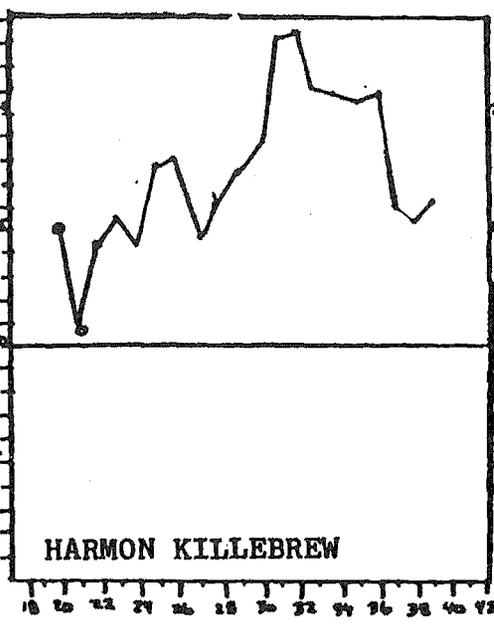
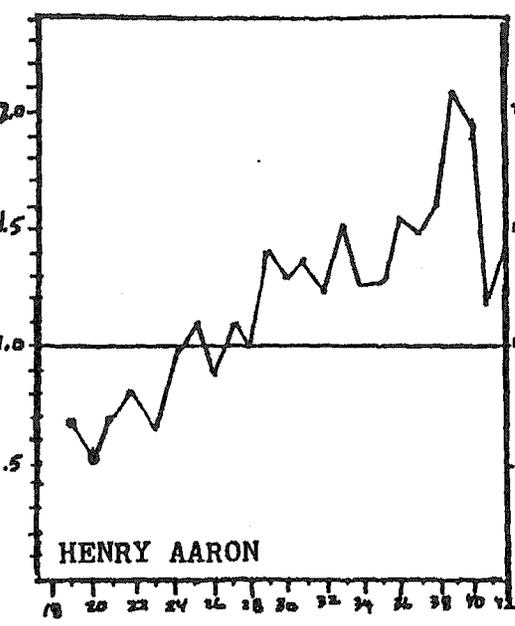
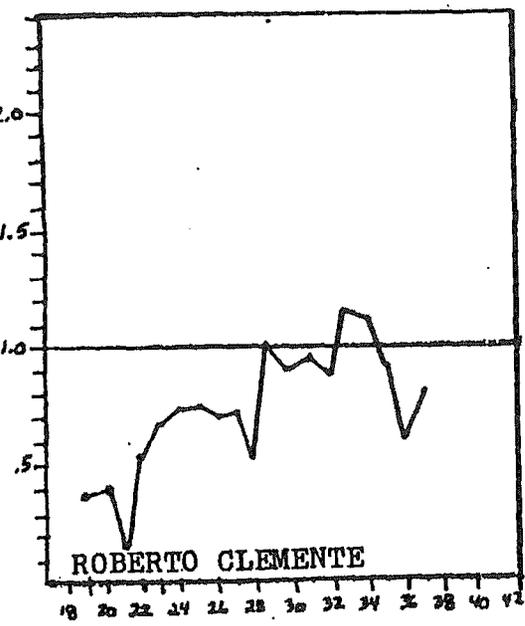


Chart 13

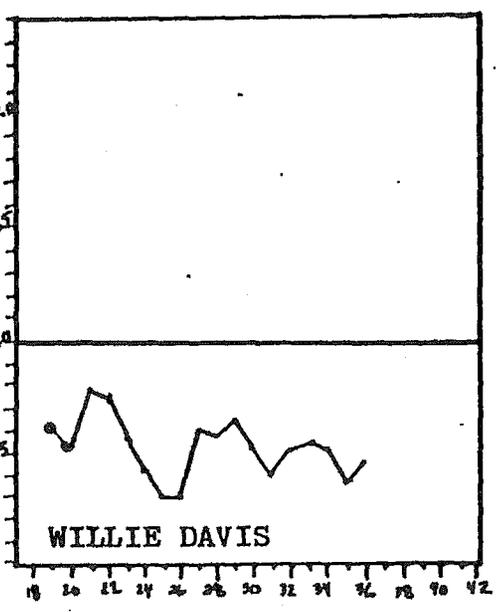
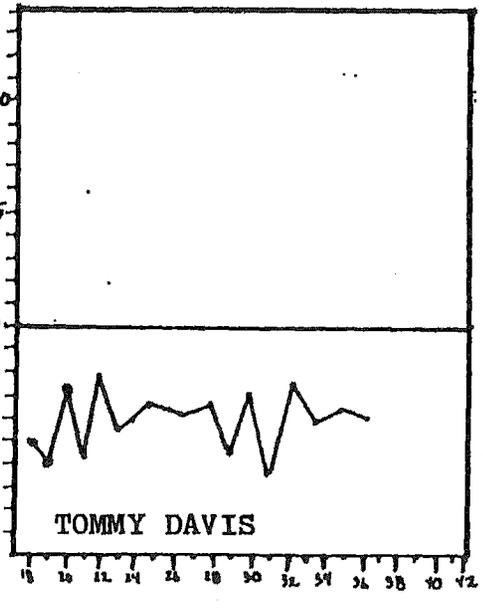
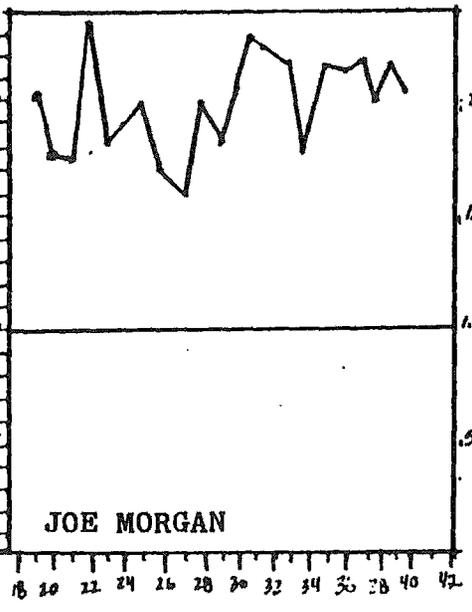
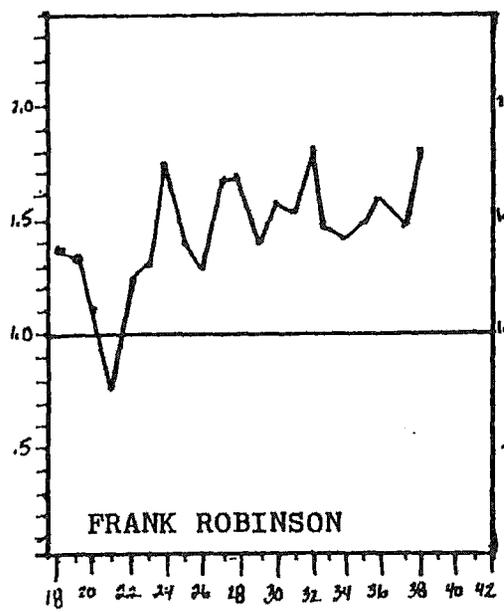


Chart 14

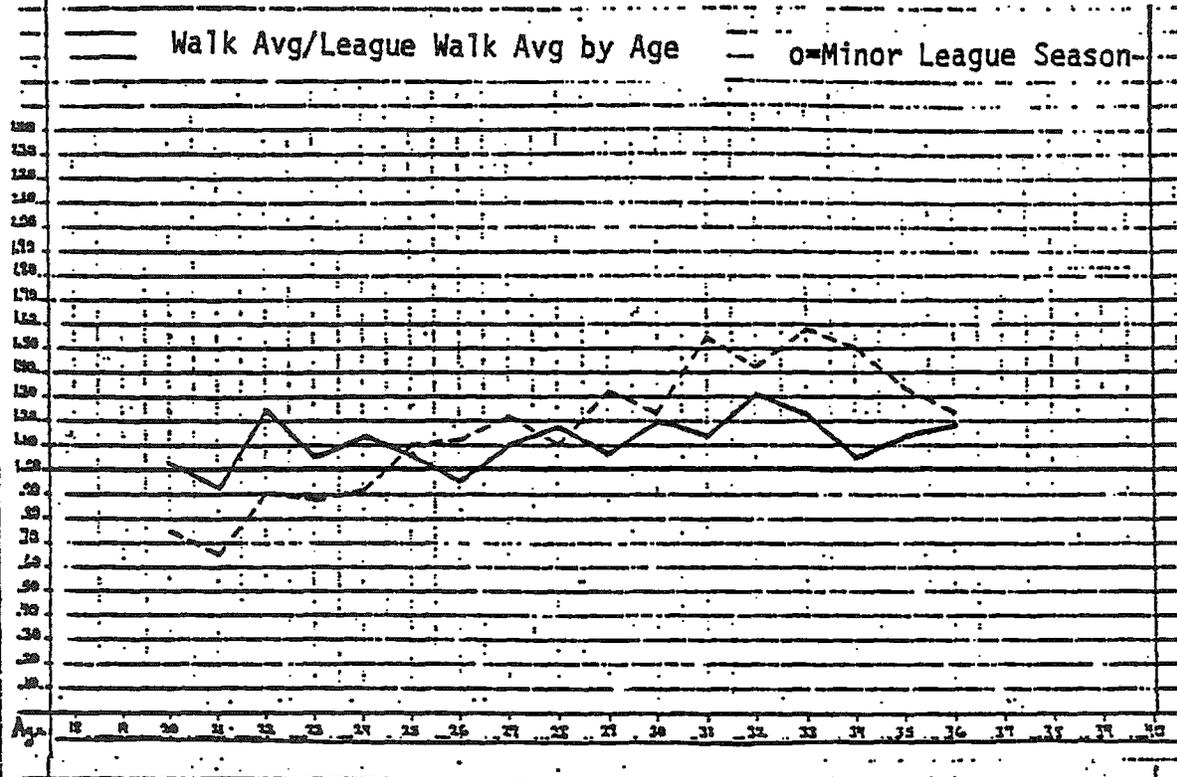


Chart 14

— T. Davis, Robinson, W. Davis, Morgan
 - - - Aaron, Killebrew, Clemente, Stargell

Chart 15

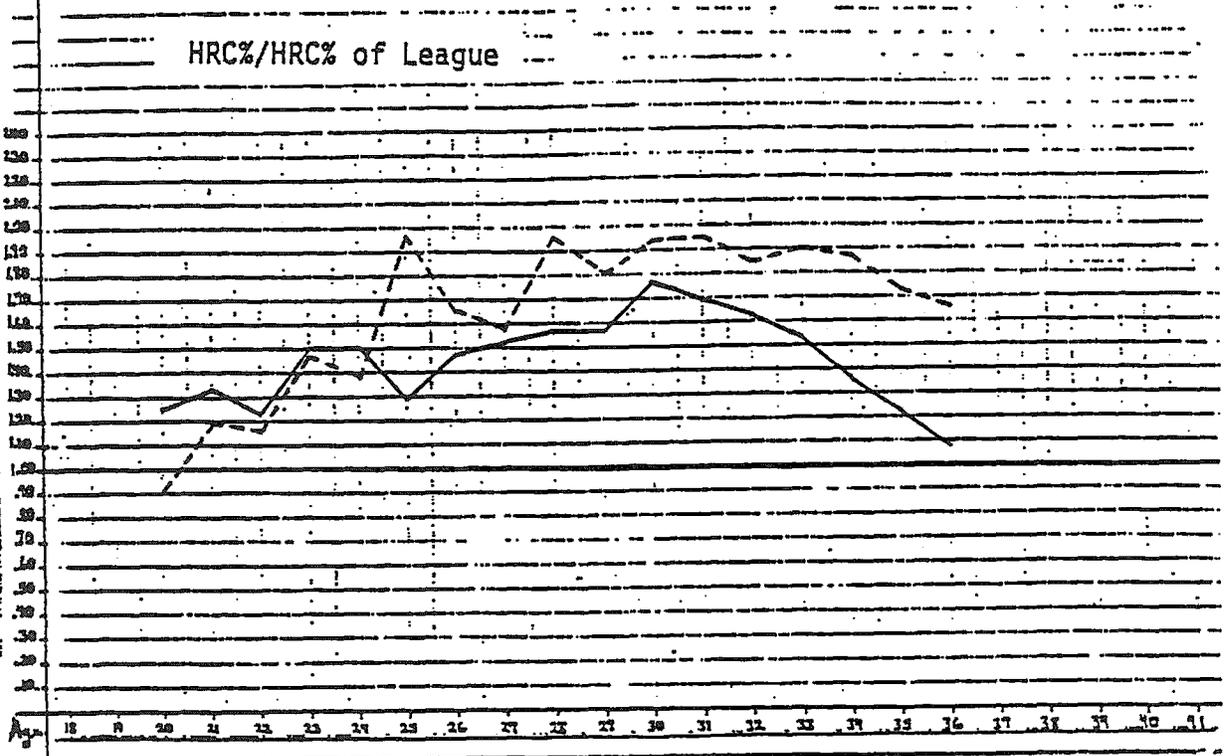


Chart 15

— T. Davis, Robinson, W. Davis, Morgan
 - - - Aaron, Killebrew, Clemente, Stargell

Player A is Frank Robinson; B is Henry Aaron.

Two more players to consider, C and D. At age 19 C hits .257 with 2 homers in 148 at-bats at AAA while D is playing in a high A league and leads in batting average (.365), triples (16), doubles (40), and hits 15 homers to also lead the league with 71 extra-base hits. The parent club of C is a last place club which lost over 100 games and is converting a .239 hitting outfielder to a pitcher. C goes to the majors at age 20 and hits .255 with only 5 homers in 473 at-bats. D, whose parent club is a first division team, jumps to AAA and again leads in hitting (.346), triples (26), and extra-base hits (81). In a September trial in the majors he hits .318 with 9 extra-base hits in 88 at-bats giving him 90 extra-base hits for the year.

At age 21 C has a relatively big sophomore season hitting .311 with 7 homers in 543 at-bats. D, now in his rookie season, plays as a reserve outfielder. D hits only .254 but has good power with 12 homers, 6 triples, and 19 doubles in only 336 AB.

At age 22 C struggles and reverts to his rookie form, .253 with 4 homers in 451 at-bats. D becomes a regular and hits .285 with 21 homers and leads the league with 10 triples.

Player D is Willie Davis and C is Roberto Clemente.

As an 18-year-old, E is a highly touted prospect signed to a large bonus which forces him to stay on the major league roster where he gets 13 at-bats. F is not highly sought after and goes to college. At 19 E stays in the majors but hits .200 with 4 homers in his 80 AB. F signs, plays well at low-A, moves to high-A in mid-season. F hits .310 and slugs .498 in the minors and .240 in a September trial.

At age 20 E goes to the minors and hits .325 with 15 homers. F plays AA-ball and finishes second in the batting race with a .323 mark and slugs .512. At age 21 E goes to AA and hits .279 with 29 homers. F is a major league regular hitting .271, slugging .398 in a tough home park, and is named Rookie of the Year.

At age 22 E plays AA again and hits .308 with 17 homers in 299 AB. Jumped in mid-season to AAA, E struggles and hits .215 with 2 homers in 122 AB. F remains a major league regular hitting .285 in his sophomore season.

Player E is Harmon Killebrew and F is Joe Morgan.

Players G and H are both in the minors at age 19. G is at AA and hits .304 with 14 homers. Player H is at A-ball and hits .260 with 11 homers. At age 20 G jumps to AAA and takes the batting title (.345) while hitting 18 homers and 9 triples. Player H moves to high-A league and hits .289 with 22 homers.

At 21 G is in the majors and hits .276 with 11 homers in 352 AB. Player H is at AAA hitting .276 with 27 homers. At 22, G is hitting .278 with 15 homers in his sophomore season. Player H is now a rookie and hits .243 with 11 homers.

At 23, player H has a pretty good sophomore season hitting .273 with 21 homers, but G is the best hitter in the league as he leads in average (.346) and RBI's (153) and hit 27 homers.

G, of course, is Tommy Davis and H is Willie Stargell.

An equally startling revelation is that the home run power of Group II was superior to Group I through age 22. Group I did out-homer Group II in minor league homers per 550 AB through age 22 (22.2 to 17.8), but Group II led in major league homers per 550 AB (21.0 to 15.3) and per all professional at-bats (19.4 to 18.9).

The best way to show how Group I overtook Group II is with a chart of their HRC% relative to the league each season. The problem is how to treat the early years when they were playing at different levels. Individual walk averages tend not to change as drastically when going from A to AA to AAA to MJL near as much as hitting statistics. Having a relative walk average of 1.00 in the minors is a lot closer to a 1.00 in the majors than a minor league relative HRC% of 1.00 is to a 1.00 in the majors.

I evolved a system that equated an A-ball HRC% as 85% equivalent to a AA performance. AA was 80% of AAA and AAA was 60% compared to the majors. Essentially that means that Killebrew's relative HRC% at A-ball of 2.01 would be comparable to a .82 performance at the major league level. This system would need to apply only through age 23 when Killebrew would be the only one of the eight not yet in the majors for good.

Chart 14 shows that Group I, the ones with the increasing walk averages, passes Group II at age 25 and never looks back. In a classic example of Theory G, their performance strictly as hitters (hitting for average and power) became more efficient; they peaked faster, prolonged that level of performance, and declined slower than Group II, the ones with the relatively consistent walk averages.

There are a number of other fine examples that seem to fit Theory G. Ken Boyer had a number of fine seasons past the normal peak years, and they corresponded with rising relative walk averages. His walk averages actually begin to rise after age 25, and at age 28 he consistently began walking more frequently than the league average for the first time in his career. From age 28 to 33 he hit .302 and had a slugging percentage of .501 compared to .286 and .459 for ages 24-27. His highest relative walk average came in his MVP year at age 33.

When Al Kaline took his batting title at age 20 he adopted not only Ted Williams' belief in conditioning but also Ted's conviction that a batter should anticipate pitches, swing only at strikes, and with less than 2 strikes generally swing only at what you anticipate. The previous year Kaline hit a respectable .276 with only 22 walks in 504 AB. The year he took the title he drew 88 walks in 588 AB; his relative walk average went from .42 to 1.19. Unfortunately Kaline did not feel comfortable walking that much and his walk averages declined to .95, .75, and 1.01 the next three years with his hitting falling off a bit, too. But at age 24 he set a new career high with a 1.35 walk average while hitting .327 and leading the league with a .530 slugging percentage. Kaline went on to walk more after age 30 and had a number of excellent seasons relative to his age. In 1967, a real pitcher's year with the league ERA at 3.23, Kaline was third in the league with a .308 average and fourth in slugging percentage (.541). He hit 25 homers (fourth in the league) in only 458 AB. Only Vastrzanski, Killebrew, and Howard had more HR's per at-bat. Kaline was 32 and had another new career high in relative walk average (1.83). In 1972, at the age of 37, he set his final career high in relative walk average (1.97) while

hitting .313 and slugging .475 in 300+ PA. It was his highest average since he was 26 and his highest slugging percentage in five years. He also led the league with 10 pinch-hits (10 for 24, .417).

John Lowenstein usually walked in the 1.35 range in the minors and matched that his first full year in the majors. Unfortunately, he hit .212 in 151 AB although he had pretty good power (.185 P%), but either the Indians or Lowenstein, himself, convinced him that his walking hurt his hitting. The next year he hit .292 while walking only 23 times in 305 AB. He was happy even though his power fell off (.079). Because of that initial success Lowenstein continued to swing more freely but the results were considerably worse. He followed that .292 year with .242, .242, .205, and another .242. Then at age 31 he tried walking a lot. He hit only .222 that year but he caught the eye of Baltimore with his .357 on-base-average and .165 P%. His first year with Baltimore he hit .254, his highest since the .292, and had a P% of .228. His second year he hit .311 and had an OBA of .408. He was 33-years-old. The next season his relative walk average fell to 1.22, his lowest as an Oriole and his lowest in five years. He hit .249 with average power. Then in 1982 he raised the relative walk average to 1.67 and had the most amazing season for a 35-year-old player who had been a utility man with a career batting average of .238 at age 32. Lowenstein hit .320 and slugged .602 in 322 AB. The next year his walk average fell slightly to 1.53 as he hit .280 and slugged .479 in 311 at-bats. At ages 35-36 he had hit .300 and slugged .542 in 633 AB, over 700 PA. That is not luck. Take his five seasons with the highest relative walk averages and he hits .283 and slugs .488 versus .240 and .360 for the rest of his career.

Willie Mays never experienced the dramatic jumps in his relative walk average characteristic of so many other power hitters. For the most part it was a slow gradual rise, but his performance, too, seemed to improve when he was walking more. The first time Willie hit 50 homers (51 at age 23) it was a new career high in relative walk average (1.26). When he hit .347, the highest average of his career, it was another new high in walk average (1.39). When he had the back-to-back home run crowns at ages 33 and 34 (47 and 52 homers) his walk averages were the two highest of his career before age 40. The period where Mays raised his walk averages the most was after age 29. It is startling how strong his performance would be for ages 30-40. He hit .295, slugged .550 and averaged 35 homers per 550 at-bats in the eleven seasons. Before age 30 he averaged only 33 homers per 550 AB.

As for walks as a side-effect of superior performance, Mays only once led the league in walks, 112 in 1971. He hit .271 and slugged .482 which is super when you are 40 years old, but not near as terrifying to the pitchers as the 21-year-old Mays who hit .345 and slugged .667 - and drew 66 walks in over 100 more plate appearances than he had at age 40.

Willie McCovey's emergence as a prolific walker and slugger is rather interesting. From ages 23 to 25 he was quite successful with a relative walk average of 1.09, a .280 average and .549 slugging percentage. At age 26 he raised his relative walk average to the dizzying height of 1.95 and had a horrible year (.220, .412). But McCovey stayed with his increased plate discipline and never fell below 1.55 in relative walk average as he ruled the league until his major knee injury at age 33. From age 27 to 32 his relative walk average was 1.90 while McCovey hit .291, slugged .581, and had an OBA of .403.

Brooks Robinson raised his relative walk averages slowly and topped 1.00 only twice in his career. The first time, he hit .272 with 20 homers at the advanced age of 34. The second time, he had his last good year by hitting .288 at age 37. It was his highest average since he hit .297 at age 28.

Doug Rader always walked a little bit below or above the league average until his last two seasons. At age 31 he hit .257, his highest average in any season with 502 PA. At age 32 he hit .251 and slugged .437, his highest mark since age 25. At first one is tempted to say that his performance was aided by leaving the Astrodome, but the Astros used the short fences for much of Rader's tenure. Doug was not the only Astro busting the ball. 1969-Wynn, 33; Morgan, 15; Rader, 11. 1970-Wynn, 27; Rader, 25. 1971-Morgan, 13; Rader, 12. 1972-L. May, 29; Wynn, 24; Cedeno, 22; Rader, 22. 1973-L. May, 28; Cedeno, 25; Rader, 21; Wynn, 20. 1974-Cedeno, 26; L. May, 24; Rader, 17. 1975-Watson, 18; Cedeno, 13; Rader, 12.

Truth is that San Diego's Jack Murphy Stadium was a tougher offensive park than the Astrodome in that period. It is also true that prior to his disheartening trade to Toronto, Rader was having his best offensive season ever (.271, .441 slugging %) while drawing an astounding 33 walks in 170 AB. That is a .163 walk average compared to his career mark of .089 going into that season.

Pete Rose slowly raised his walk averages in a choppy fashion during his career. It stands out that his performance was usually superior when he walked the most. Rose had five seasons through age 40 where his relative walk average was over 1.20, ages 28, 33, 34, 35, and 38.

Ages 22-40	AB	H	BA	2B	3B	HR	Slug %
28, 33-35, 38	3,234	1,036	.320	207	33	40	.442
Other	8,676	2,661	.307	465	89	115	.421

Carl Yastrzemski had seven seasons where his relative walk average exceeded 1.60. Those seven include his three batting titles, his two seasons of .300-40 homers, and his 2 highest averages after age 30 (.296 and .301 at ages 33 and 34).

We will close with another Lowenstein-type example to demonstrate clearly that Theory G seems to apply to all levels of talent. Ernie Whitt's walk averages in both the minors and majors were very average through age 30. Offensively, he had a major league career average of .240 with 19 homers in 837 AB. In 1983 at the age of 31 he turned in a relative walk average of 1.42 and hit .256 with an amazing 17 homers in 344 AB. As of August 26, Whitt's relative walk average for 1984 was 1.49 and Whitt was hitting .252 with 11 homers in 246 AB, roughly the same ratio as in 1983.

MJL Career	AB	H	BA	2B	3B	HR	Slug %	Relative Walk Avg.
1976 - 1982	837	201	.240	40	4	19	.366	.95
1983 - 8/26/84	590	150	.254	26	3	28	.451	1.45

At present Theory G stands up well. It is consistent with what evidence there is. There is no firmly based theory that it contradicts. It is not illogical.

Remember that the theory does not claim that the walks themselves automatically make the hitter better. The theory of the theory, so to speak, is that the techniques and/or strategies applied by a hitter that cause more walks to be drawn also tend to improve the batter's offense as a hitter.

The conclusions emerging from the research in Section II are:

- A. Walk averages tend to be steady and characteristic over the career for the near majority of players, but the exceptions are neither rare nor mild in their deviations.
- B. Walk averages tend to gravitate upwards through most of the average career.
- C. Among young players, those who walk the most will tend to raise their walk averages even more during their careers. The exception would be any young player already walking at twice his league's average.
- D. Drawing walks is less of an innate ability than most baseball skills and can be both learned and taught.
- E. Power-hitters tend to raise their walk averages more during their careers than other hitters.
- F. No player-type has been shown to suffer a consistent loss of offensive performance as a hitter when drawing more walks than usual.
- G. Increasing walk averages enhance a player's performance as a hitter, prolong his peak, and slow his decline.

NEXT ISSUE: PART THREE (CONCLUSION) OF CRAIG WRIGHT'S
BASE ON BALLS ABSTRACT

Hall of Fame Candidates
by Daniel Greenia

These are the highest ranking players in various offensive categories who, although eligible, are not in the Hall of Fame:

Rank	HR	RBI	BA
18) B. Williams	426	31) B. Williams 1475	3) J. Jackson .356
F. Howard	382	G. Davis 1435	P. Browning .343
O. Cepeda	379	O. Cepeda 1365	R. Stephenson .336
	R		SA
21) B. McPhee	1684	32) V. Pinson 2757	20) R. Allen .534
J. Ryan	1643	B. Williams 2711	B. Herman .532
G. VanHaltren	1639	D. Cramer 2705	K. Williams .531
	G		SB
25) B. Williams	2488	19) B. Williams 4599	8) M. Wills 586
V. Pinson	2469	V. Pinson 4264	C. Milan 495
B. Dahlen	2443	O. Cepeda 3959	J. Sheckard 460

A CORRECTION TO
 "SOME FURTHER ASPECTS OF THE DISTRIBUTION OF RUNS"

by Dallas Adams

In its February 1983 issue the Baseball Analyst published the second half of my paper "Some Further Aspects of the Distribution of Runs Scored". Shortly after its publication I noticed that what I had said at the end of Section III (page 12 of the referenced issue) was in error and ought to be corrected.

The equation I had used to represent the home team's probability of winning at the beginning of an extra inning was really the equation for the team's chance of winning from the start of the game, which is not the same thing at all. Instead of what I wrote, the final paragraph of Section III should read:

When the two teams are of unequal strength, they will each have a different set of per out scoring probabilities. When computing a PW (or PL) per equations 9-15, use the home team's R_i 's when that is the team which is batting; and use the visiting team's set of R_i 's when it is the visiting team which is batting. The only other item which will change is the coefficient $\frac{1}{2}$ in equation 9. This fraction must be adjusted to reflect the home team's new probability (it no longer is $\frac{1}{2}$, for the two teams now are of unequal strength) of losing when the game enters extra innings. The new fraction, call it f , is:

$$f = \frac{\sum_{i=1}^{19} V_i \sum_{j=0}^{i-1} H_j}{1 - \sum_{i=0}^{19} V_i H_i} \quad \text{(equation 16)}$$

where V_i = the visiting team's probability of scoring exactly i runs in an inning
 H_j = the home team's probability of scoring exactly j runs in an inning

The derivation of equation 16 begins by noting that f is defined as the probability that the home team will lose when a game enters the top of the 10th inning (or any extra inning).

When an extra inning game enters the bottom of the extra inning, the home team will either trail by a margin of i runs or it will be tied (in which case $i=0$). If the home team then scores more than i runs in the inning, it will win; if it scores fewer than i runs it will lose; and if it scores exactly i runs then it will force the game to go another inning, in which case its chance of losing is, by definition, equal to f . This means that the home team's chance of losing when it enters the bottom of an extra inning trailing by i runs or tied ($i=0$) is

$$fH_i + \sum_{j=0}^{i-1} H_j \quad \text{(equation 17)}$$

where f and H are as defined above.

In the top half of the extra inning the visiting team will score zero runs V_0 percent of the time; thus sending the game into the bottom of the

inning with the score still tied, which means there is an fH_0 (from equation 17) probability of the home team losing. If the visiting team scores exactly one run in the top of the inning, then (again from equation 17) there is a fH_1+H_0 probability of the home team losing. And so forth. The general expression, then, for the home team's chance of losing when the game enters an extra inning is

$$V_0(fH_0) + V_1(fH_1+H_0) + V_2(fH_2+H_0+H_1) + \dots + V_{19}(fH_{19} + \sum_{j=0}^{18} H_j)$$

But the home team's chance of losing when a game goes into extra innings is, by definition, equal to f . Hence,

$$f = V_0(fH_0) + V_1(fH_1+H_0) + V_2(fH_2+H_0+H_1) + \dots + V_{19}(fH_{19} + \sum_{j=0}^{18} H_j)$$

which simplifies to equation 16.

Thank you for clearing that up for us Dallas. I have one comment on your clarification: Logically, the argument presented is correct. However, the notation in equation 16 is mathematically incorrect. As I perceive it, H_j = the home team's probability of scoring exactly j runs in an inning; therefore, the denominator should appear as

$$1 - \sum_{i=0}^{19} V_i H_i$$

(the probability that 2 teams do not score the same amount of runs in an inning).---John Borkowski

* * * * *

I'd like to attach a postscript expressing my appreciation to Charles Hofacker for his paper "Functions for Predicting Winning Percentage from Runs" which appeared in the December 1983 issue of the Baseball Analyst. His work extends the frontiers of a subject which interests me greatly.

I was especially impressed with his explanation of why the Pythagorean equation of Bill James is a valid predictor of a team's winning percentage. Very nice.

In conjunction with his Figure (2), which expresses the distribution of leads after nine innings, Hofacker notes that with the addition of "a small unspecified amount for the probability the home team wins in extra innings" the team's overall probability of winning could be determined from that figure. Well, my equation 16, above, gives an expression for f , the home team's probability of losing when a game goes into extra innings. Hence, Hofacker's required probability that the home team wins in such a situation is $1-f$. My equation 16 requires V_i and H_i , the distribution of runs in an inning by visiting and home team respectively; these can be computed from each team's distribution of runs in a game (his Figure 1) by the equations given in my paper in the December 1982 issue of the Baseball Analyst.

--Dallas Adams