By the Numbers

Volume 19, Number 3 The Newsletter of the SABR Statistical Analysis Committee August, 2009

Comment

Hitting Streaks: A Reply to Jim Albert

Trent McCotter

In a previous article in BTN, Jim Albert found that there was no significant difference in the expected and actual number of hitting streaks over individual seasons. Here, the author argues that when you aggregate all the single-season data, the result is *statistically significant, and comprise valid evidence that hitting streaks are indeed more frequent than expected.*

In the 2008 edition of SABR's "Baseball Research Journal.," I published an article showing evidence that hitting streaks in baseball occur significantly more frequently than they would occur if hitting was random from game to game. I used the random permutation method to determine whether the number of hitting streaks (of lengths 5+, 10+, 15+, and 20+ games) matched what an IID (independent and identically-distributed) model would look like. It turned out that it did not.

Later, in the November, 2008 issue of *By the Numbers,* Jim Albert analyzed the seasons from 2004 to 2008 using the same method that I used, but taking the seasons individually. Jim found high p-values for most numbers; that is, the number of streaks in real-life wasn't significantly higher than a random permutation would produce.

I have two issues with Jim's analysis and results. First, his results still show a tendency for there to be more hitting streaks in real-life than we'd 'expect' using a random permutation method – even at the single-season level. Out of the 20 matched-pairs that Jim generated (five years of data, with four different lengths of hitting streak for each year), 15 of those pairs had a higher value for the 'real life' streak total than for the average over the permutations. And the other 5 (where the real-life total was less than the average over the permutations) were pretty close to being even. So I'd say that – even at the single-season level – there is evidence that hitting streaks of pretty much every length are more-likely to occur in real-life than if the games were randomly permuted.

Second, even if Albert's results didn't show a tendency for there to be more streaks in real-life than over a random permutation, I'd still have a major qualm with his method of trying to show that there is little difference between streak totals in real life vs. the permutations. The qualm is that Albert split the 50 years of data that I used into single seasons and then said that there wasn't much significance at a single-season level. But that would be the case with almost every study. The entire purpose of conglomerating 50 seasons' worth of data is to find trends that might not be as obvious at a single-season level (although, per #1 above, I think there actually IS evidence that shows some significance at the single-season level).

If we look at each season individually and say that maybe there's a slight trend towards more hitting streaks, that wouldn't mean much; but if almost *every* season showed the exact same trend, then it would be very meaningful. In other words, the entire purpose of larger sample sizes is to smoke out trends that might not be apparent on an individual sample-by-sample basis; but if almost every sample tends to show the same pattern, then we probably have something significant going on. Of course it makes sense that – in any given season – there might not be that much evidence of a trend; the trend only becomes obvious when viewed from afar, when all the seasons are added together and their similar patterns become magnified.

Trent McCotter, treant985@triad.rr.com ♦

In this issue

"By the Numbers" mailing list

SABR members who have joined the Statistical Analysis Committee will receive e-mail notification of new issues of BTN, as well as other news concerning this publication.

The easiest way to join the committee is to visit http://members.sabr.org, click on "my SABR," then "committees and regionals," then "add new" committee. Add the Statistical Analysis Committee, and you're done. You will be informed when new issues are available for downloading from the internet.

If you would like more information, send an e-mail to Neal Traven, at beisbol@alumni.pitt.edu. If you don't have internet access, we will send you BTN by mail; write to Neal at 4317 Dayton Ave. N. #201, Seattle, WA, 98103-7154.

Submissions

Phil Birnbaum, Editor

Submissions to *By the Numbers* are, of course, encouraged. Articles should be concise (though not necessarily short), and pertain to statistical analysis of baseball. Letters to the Editor, original research, opinions, summaries of existing research, criticism, and reviews of other work are all welcome.

Articles should be submitted in electronic form, preferably by e-mail. I can read most word processor formats. If you send charts, please send them in word processor form rather than in spreadsheet. Unless you specify otherwise, I may send your work to others for comment (i.e., informal peer review).

I usually edit for spelling and grammar. If you can (and I understand it isn't always possible), try to format your article roughly the same way BTN does.

I will acknowledge all articles upon receipt, and will try, within a reasonable time, to let you know if your submission is accepted.

Send submissions to Phil Birnbaum, at birnbaum@sympatico.ca .

Runs Scored in Division III Baseball: A Preliminary Study

John F. McDonald

The author computes a run-expectancy matrix for a certain Division III (college) baseball team, and compares it to the MLB version.

Introduction

The classic study of strategies in baseball by George Lindsey (1963) introduced the study of runs scored from the 24 situations that can exist when a new batter steps up to the plate. Since then the tabulation of this data has become a standard tool for analyzing strategies in major league baseball, and is known as a run expectancy matrix or expected run table. It turns out that the numerical elements in this table are reasonably stable over time. Lindsey's original table compiled from the 1959 and 1960 major league seasons (1963, p. 485) is shown in Table 1.

A basic fact in the table is that the average major league team scored 4.14 runs in nine innings (0.46 times 9). This table led Lindsey to reach conclusions such as the inadvisability, on average, of sacrificing an out to move a runner from first base to second base. The table shows that the average number of runs scored when there is a runner on first with no outs is 0.81, but the runs scored with a runner on second and one out is 0.67. However, note that a sacrifice out (with no outs) that moves runners on first and second to second and third increases the number of runs scored from 1.47 to 1.56. In this case the manager must factor in the chance that the batter will fail to execute the sacrifice successfully (which is about 11%).

The table also shows the importance of getting the leadoff batter on base. The average runs scored with one out and bases empty is only 0.24, compared to 0.81 runs scored with a runner on first and no outs. See Click (2006) for the run expectancy table for the major leagues in 2004. Average runs scored per nine innings had increased to 4.86, so many of the numbers in this table are somewhat larger than the numbers in Lindsey's original table.

Table 1 – Run Expectancy Matrix, 1959-60

Runners	0 Outs	1 Out	2 Outs
None	0.46	0.24	0.10
1st	0.81	0.50	0.22
2nd	1.19	0.67	0.30
1st & 2nd	1.47	0.94	0.40
3rd	1.39	0.98	0.36
1st & 3rd	1.94	1.12	0.53
2nd & 3rd	1.96	1.56	0.69
Loaded	2.22	1.64	0.82

Source: Lindsey (1963, p. 485)

Researchers are asking whether the methods developed for studying major league baseball can be applied to baseball and softball at lower levels. For example, Jean Marro and Thomas Pfaff (2007) studied the ability of the Bill James runs-created formula as applied to Division III college softball. This paper uses the Lindsey run expectancy table to study runs scored in Division III baseball.

Context of the Study

This study uses the scorebook from the 2009 season of one Division III college baseball team in the Midwest. This team played 36 games in a season of 64 days that started in early March and ended in early May. Some games were played on a Spring Break trip to the South, but otherwise the games were played in the Midwest – sometimes under challenging weather conditions. The team had a winning record.

Baseball at the Division III level is different from major league baseball, of course. Use of aluminum bats (as opposed to wooden bats) means that balls often are hit more sharply. I played Division III baseball in the days of wooden bats, and to me the difference is quite striking (pun intended). Even smaller Division III players can put quite a charge into a hit. When combined with aluminum bats, other factors can also mean that many runs are scored. Some of these factors can include:

- Schedules that put a strain on pitching staffs,
- Fielding of variable quality, and
- Smaller dimensions of some Division III ballparks.

A team often scores ten runs or more in a game.

The Preliminary Study

The study uses a sample of 20 games played by this team to compile a Lindsey table, which is shown as Table 2. The table pertains only to runs scored by the team, not the opponents, on the grounds that this is the table that the coach needs to think about offensive strategy.

One of the first results to note is that this team scored an average of 6.84 runs per nine innings (0.76 times 9). (From Table 1, major league teams scored 4.14 runs per nine innings in 1959-60.) Next, note the harm done by an out. Runs drop off sharply in all base runner situations as outs increase. A third basic fact is that getting the leadoff batter on first base is a very big advantage. Average runs scored with a runner on first and no outs is 1.29, compared to 0.31 runs scored with nobody on and one out, a difference of 0.98 runs. If this team could get the leadoff batter on first in every inning, it would score almost nine runs more per game! Why does this happen? Some factors that apply to baseball at all levels are:

- The pitcher has to pitch from the stretch and worry that the runner will attempt to steal second.
- The first baseman has to hold the runner on and therefore covers less ground.
- The second baseman or shortstop has to shade towards second to cover the base.

Combine these factors with the aluminum bats used in Division III, and having the leadoff batter on first can open the floodgates. And it gets worse (or better, depending on your point of view) if the team in the field is using its number three or four starting pitcher.

Table 1 illustrates that the advantage of getting the leadoff batter on first base is 0.57 runs for major league teams in 1959-60. However, if major league teams in 1959-60 had scored as many runs as this Division III team, the comparison is much closer. The major league teams scored 4.14 runs per nine innings in those years, compared to 6.84 runs per nine innings for this Division III team – a ratio of 1.65. Multiplication of the figures in the first three lines of Table 1 by 1.65 produces Table 3.

Table 2 – Run Expectancy Matrix, Division III Baseball Team

Runners	0 Outs	1 Out	2 Outs
None	0.76	0.31	0.20
1st	1.29	0.80	0.25
2nd	2.15	0.81	0.74
1st & 2nd	2.09	1.11	0.21
3rd	*	1.33	0.25
1st & 3rd	*	1.50	1.00
2nd & 3rd	*	1.87	1.08
Loaded	2.15	1.85	0.58

* Situation occurred too few times for reliable results

Source: 2009 Team Scorebook

This adjusted table shows that the advantage of getting the leadoff man on first base would have been 0.94 runs, compared to 0.98 runs for the Division III team. The other elements in the first two lines of Table 2 and Table 3 are quite similar.

Table 3 – Projected MLB Run Expectancy for Divison III Offense

Runners	0 Outs	1 Out	2 Outs
None	0.76	0.40	0.16
1st	1.34	0.82	0.36
2nd	1.96	1.11	0.50

Source: Table 1 with cells multiplied by 1.65

The inadvisability for this Division III team, on average, of sacrificing a runner to second is pretty powerful. In Table 2 average runs scored with a runner on second and one out is 0.81, compared 1.29 runs scored with a runner on first and no outs – a difference of 0.48 runs. In the major leagues the difference from Table 1 above is only 0.14 runs. (The adjusted difference is 0.23 runs in Table 3.) Why the big difference? Is it the aluminum bats used by the Division III team, coupled with the extra out? This is a question that needs further study.

What about stealing second base? Detailed analysis of stealing in major league baseball produces pros and cons, but in Table 2 the utility of stealing second is clear. Having a runner on second with no outs, for this team on average, produced 2.15 runs compared to 1.29 runs for a runner on first and no outs – a gain of 0.86 runs. Combine this with the fact that the ability of Division III catchers to throw out runners at second may not be very good, and you get a green light. However, the figure of 2.15 runs with a runner on second and no outs looks to be questionable (perhaps a result of the small sample size of 20 games) because the team scored 2.09 runs with runners on first and second and no outs (and 2.15 runs with the bases loaded and no outs). Using the same comparison of situations, the effect of stealing second in the major leagues from Lindsey's Table 1 is to add 0.38 runs, on average. (The adjusted effect of stealing second in Table 3 is 0.62 runs.)

Let's look at stealing second with no outs more closely. Suppose that the Division III catcher is excellent and can throw out the runner 40% of the time. In this case this team on offense can expect to score the following number of runs when the steal sign is flashed;

```
(0.60) x (2.15) + (0.40) x (0.31) = 1.41 runs.
```

This is compared to 1.29 runs if the steal is not attempted. Suppose p is the probability that the catcher throws out the runner at second. The "break-even" point in this case is:

```
(1-p) x (2.15) + (p) x (0.31) = 1.29, or p = 0.47.
```

Is any catcher in Division III baseball that good? I doubt it. From Lindsey' table, a 40% failure rate for stealing second means that the average runs scored with a steal attempt is;

```
(0.60) x (1.19) + (0.40) x (0.24) = 0.81 runs
```

compared to 0.81 runs if the steal is not attempted. This is the "break-even" point. Click (2006) reported that the actual break-even point in the major leagues varied between 30% and 25% from 1982 to 2004 (except for extreme values of 34% in 1985 and 22% in 2001).

Table 2 is based on a relatively small sample of games (20), so some of the eight situations with no outs occurred only a few times (six or fewer); these are noted in the table. The entries in the table for these situations are not reliable and are not reported.

Conclusion

A basic conclusion is that the run expectancy table can be a valuable tool for baseball coaches at the Division III level. The results of this study show that sizable differences may exist between the tables compiled for Division III teams and major league teams – and therefore that strategies should be different at different levels of baseball. I think that baseball coaches can benefit from this kind of analysis. Furthermore, they are dealing with smart college students (some of whom are math, business, or economics majors) who will benefit from learning about and actually doing sabermetrics for their own teams. It should be part of the education of college baseball and softball players, who will become the next generation of sabermetricians. Get your players in the classroom, coach!

References

Click, J. (2006) "What if Rickey Henderson Had Pete Incaviglia's Legs?" In J. Keri, ed., *Baseball Between the Numbers*, New York: Basic Books.

Lindsey, G. (1963) "An Investigation of Strategies in Baseball," *Operations Research*, Vol. 11, pp. 477-501.

Marro, J., and T. Pfaff (2007) "Does the Runs Created Formula Work for Division III Softball?" *By the Numbers*, August 2008, pp. 13-15.

John F. McDonald, Roosevelt University, 430 S. Michigan Ave., Chicago, IL 60605, jmcdonald@roosevelt.edu ♦

The Greatest Base Thief

Tom Hanrahan

Who was the greatest base stealer of all time? The author discusses why the standard methods – most steals, most steals per game, most steals per plate appearance – are misleading, and offers his own measure.

Who was the best base stealer in Major League history?

Most "best ever" questions generate a lot more heat than definitive answers, but in this case, most people would agree that Rickey Henderson was the best thief, by virtually any metric. Except for the one I'll use in this paper; and I will propose this metric may be the most informative measure regarding who was the greatest base thief.

First, let's look at the career SB leaderboard. It says:

```
R. Henderson  1406
L. Brock       938
B. Hamilton    912
T. Cobb        892
T. Raines      808
V. Coleman     752
```

Not even close, is it? Now, career totals are not everything. Most 'best ever' discussions also consider rate statistics; was the best slugger ever Hank Aaron, who has the most total bases and RBI, or was it Ruth and his eye-popping career-best .690 slugging percentage? Rate statistics for hitters are typically denoted in success per plate appearance, or sometimes per year. This makes sense, because hitters take turns batting, and so this more or less evens out opportunities. And this should take the wind of Rickey's sails, right? I mean, he is 4th all-time in games played and in plate appearances, behind Rose, Yaz, and Aaron; definitely a long career! So, I could create a table of most steals per years-in-majors, or steals per plate appearance (with some minimum # of SB required) … and again, Rickey dwarfs most others.

However, these are not the right measurements to put in the denominator when it comes to stealing bases. Batters do not take 'turns' when stealing; they *choose* to run. And, not everyone gets the same opportunity to steal a base every game; obviously, you have to get on base. Primarily, first base, since most stolen bases are going from first to second. Yes, third base can be (and was) stolen as well, but most of time it is second base that is pilfered, while reaching first also gives you the opportunity to steal second *and* third. Furthermore, the number of doubles players hit is dwarfed by the number of times they reach first, and is not a significant differentiator between players (no great base thief consistently hit more than 30 doubles per season).

I could alter this by only counting steals of second base, in order to directly correlate with Times Reached First Base (TRF). However, that would discount the value some players had by stealing other bases. Also, it is very difficult to obtain data for steals of third and home for each player. Conversely, I could use some other metric that also incorporates times reached second base (by hitting a double), but then I would have to concoct a special formula using some weighted variables for how much more it "counts" to get to first than second, and there is no perfect way to come up with such a number. And again, the differences in batter doubles are not large; for example, Rickey Henderson hit about 5 doubles more per year than Vince Coleman.

I submit that the correct metric for creating a rate statistic for stealing would be stolen base success per TRF. And first base only; hitting a triple does not leave much opportunity to steal.

Some might suggest here that even better than using TRF would be using play-by-play data to count how many times a player was on first base *with second base open*; this would correct for those times when a steal was unlikely (Tim Raines on first with some slow pitcher in front of him). And I agree, this would be a more accurate measure; it also involves a boatload of extra work, and I judge the gain to be marginal. Even if accomplished, one would have to decide whether being on first with second base open for an entire inning (none out, one out, two out) would count as three opportunities to steal, as opposed to one, and there may not be a definitive answer to this question. If someone wishes to go through the Retrosheet database (yes, Retrosheet *is* a wonderful thing) for the entire careers of many players, it's all yours.

Henderson reached first base very frequently; his career OBP is .401. He was not a big slugger, either, so most of his times on base were on first base, from where most steals occur. When you put all of those together, Rickey was standing on first base more than almost any other player in MLB history. Table 1 shows the career leaders in Times Reached First (singles plus walks plus HBP; I have ignored reaching on errors, catcher's interference, fielder's choices, etc).

A player needs talent, health, on-base ability, as well as the ability to steal, in order to be the only major leaguer to amass over 1,000 stolen bases, and Rickey indeed brought those key areas together. Only Pete Rose stood on first base more often in his career. By dividing stealing prowess by TRF, we find a rate statistic akin to batting average, slugging, etc.; it does not measure volume, but it does measure proficiency. As with any rate stat, some minimum number of steals should be used, so that we do not anoint someone like Matt Alexander (103 career SB, with only 36 career hits; he was mainly a pinch runner) as the best thief.

The next item to be addressed is that a count of stolen bases (SB) does not fully measure success; we need to account for the cost of being caught stealing (CS). Over the course of time, there have been about two successful steals for every runner caught. Also, most students of the game have come to general agreement that it takes two successful bases stolen to equal the cost of one caught stealing; 2/3rds is about the consensus 'breakeven' point. Thus, I propose that a measurement of "SB minus 2 * CS" is a good measure of the net total of base thievery. I will refer to SB – 2 * CS as "Net Steals".

Table 1 – Times Reached First Base (TRF)

	Singles	Walks	HBP	Total
Rose	3245	1556	107	4908
Henderson	2182	2190	98	4470
Cobb	3053	1249	94	4396
Bonds	1495	2558	106	4159
Yastrzemski	2262	1845	40	4147

I will make one other adjustment; because it is possible to achieve a negative number of Net Steals, and because either at the beginning or end of a player's career he may not be as efficient a base thief, I will delete any seasons where a player's Net Steals are 0 or below. This will only be done for leading or trailing seasons; if a player has one poor season of being caught stealing very often in the middle of his career, that is his fault and will hurt his record. But it seems incorrect to penalize a player as if he were 'worse than zero' for poor seasons if his manager kept sending him after his stealing ability was gone.

Table 2 – Career Net Steals (SB – 2 * CS) Leaders, 1950 to 2009

	SB	CS	Net Steals	Deleted Years?
R. Henderson	1406	335	736	
T. Raines	808	146	516	Last season
W. Wilson	668	134	400	
V. Coleman	752	177	398	Last season
J. Morgan	689	162	365	
D. Lopes	557	114	329	
L. Brock	917	295	327	Last season
K. Lofton	622	160	302	
B. Campaneris	643	192	259	Last season
O. Nixon	620	186	248	
E. Davis	335	54	227	

At this point, it would be helpful to print a table of career Net Steals leaders…. but we have a problem: MLB did not always record CS. Before 1950, we are missing a lot data. So, this study will only on post-integration players. Yes, this misses Ty Cobb and Slidin' Billy Hamilton, but we ought to recognize that before the live ball era (1920), baseball was a *very* different game, and it may be unwise to be comparing base thieves of 100 years ago with modern counterparts anyway. And between 1920 and 1950, there were no stellar base thieves, so this analysis will not miss much by skipping that early portion of the live ball era.

So now, we're about to (finally!) ask and answer the question, "who was the greatest base thief of the past 60 years?" There will be two measures; a career total of Net Steals, and a career rate of Net Steals per TRF. First, Table 2 gives a Net Steals leaderboard. The right-most column shows any seasons that were deleted from the player's record as mentioned above; in this case, it only applied to a few seasons as older players became less able to steal bases.

Once again, Rickey has a commanding lead. Tim Raines, a very efficient base thief, is a clear second.

Before proceeding, I should note and discuss the obvious; many of these players were contemporaries. Should we discount the base-stealing exploits of those who played in the 1980s, when MLB was generating great base thieves like spring rabbits? I would argue that we should not. Yes, more players with speed skills were reaching the majors in that time. Yes, managers were running more often; but actually not *that* much more often. Stolen bases have been about even (± 10% per year) from 1977 through the late 1990s, and even as they have declined in the last decade, success *rate* has gone up. A premier base thief should have been a premier base thief regardless of when he played. Managers of Luis Aparicio, Maury Wills, and Carl Crawford recognize their player's talents and use them accordingly. It is true that artificial turf played some role in the game, and it is true that speed was the "in" thing for a number of years; but I fail to see how Rickey Henderson would have been a different creature who had less base-stealing success had he made the majors in 1960 or 1995. Yes, he might have run a little less often, and so his career stats might look more like 1150 SB and 200 CS instead of 1406-335; but the value would have been about the same. And if a player existed in 1955 or 2005 who had Rickey's skills, I have no doubt that he would have been a monster on the base paths. So, while acknowledging that game conditions were such that players like Omar Moreno and Frank Taveras might have had shorter MLB careers had they been born a generation later or sooner, I maintain that when assessing the very greatest thief, this person would equally have existed in 1960, 1985, or 2010.

Next, we computer the rate statistic Net Steals per TRF. I will use only the 11 men in Table 2, so as not to populate the rankings with players who did not last very long in the majors. The results can be seen in Table 3. Some observations from the Table:

Table 3 – Net Steals per TRF

	Net Steals	TRF	Net Steals per TRF
V Coleman	398	1622	.245
W Wilson	400	2275	.176
E Davis	227	1327	.171
R Henderson	736	4470	.165
T Raines	516	3228	.160
D Lopes	329	2098	.157
O Nixon	248	1789	.139
L Brock	327	2932	.112
K Lofton	302	2776	.109
B Campaneris	259	2404	.108
J Morgan	365	3609	.101

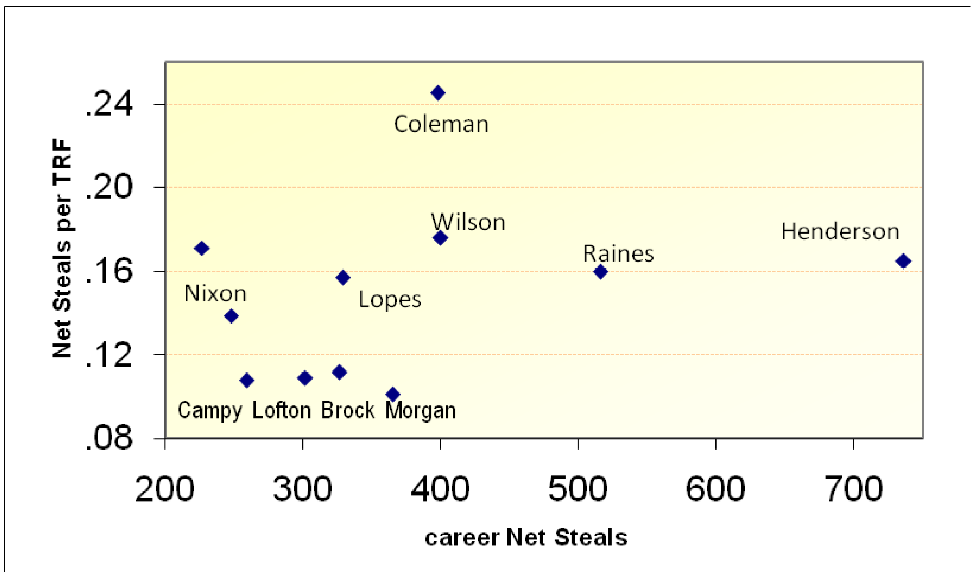
Some current players who would be among the leaders in Net Steals per TRF are Carl Crawford (.153) and Jose Reyes (.163), but they do not have the career totals of these others. Speed is a young man's strength, more than most any other skill, and most players steal many fewer bases as they age; it is likely their rates will decrease the longer they play.

Eric Davis hit fewer than 700 singles in his entire career, and still pilfered a lot of bags.

On the other hand, Joe Morgan is known as a great base thief, but much of his success came from standing on first so often.

And once again, we have a very clear leader dominating the others; but this time it's a new name: Vincent Van Go. A few facts about Vince Coleman's career:

Figure 1 – Career vs Rate Net Steals



- He led the NL in steals in each of his first 6 seasons. During that time, he had an on-base percentage of only .326. And, he never played more than 154 games in a season.

- In 1986, Coleman had a poor year, in that he hit only .232. Somehow, he stole 107 bases. And was caught only 14 times. That is *by far* the most incredible single-season base thievery accomplishment ever.

- From 1985 to 1987, three players stole 90 or more bases in a season. Coleman, Coleman, and…. Coleman. 110, 107, and 109.

- Coleman's baseball prowess was pretty much limited to his speed. He never drove in more than 43 runs in a season.

- Coleman's last season of over 100 at-bats was 1995. But he is more than three years younger than the recently retired Julio Franco.

So, how do we bring these two measures, career totals and career rates, together? First, a graph, plotting them against each other, is shown in Figure 1. Efficiency (rate) is on the Y axis, and Volume (totals) is on the X axis.

This shows what we already knew; Henderson is far and away the best career thief, and Coleman is king of proficiency. At this point, some readers are thinking the main reason Coleman's rate stats are so good is that he missed much of his 'decline phase'; after all, he was essentially finished by age 33. How would he have done at age 40? What would Rickey's numbers look like if he had retired a decade or so sooner? I'm so glad that you asked.

Figure 2 – Rickey Henderson and Vince Coleman: Steal rate by age

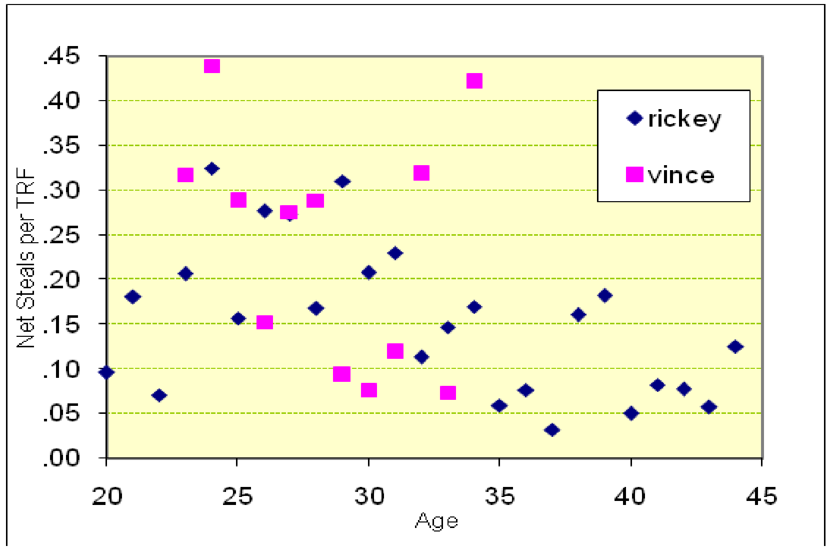
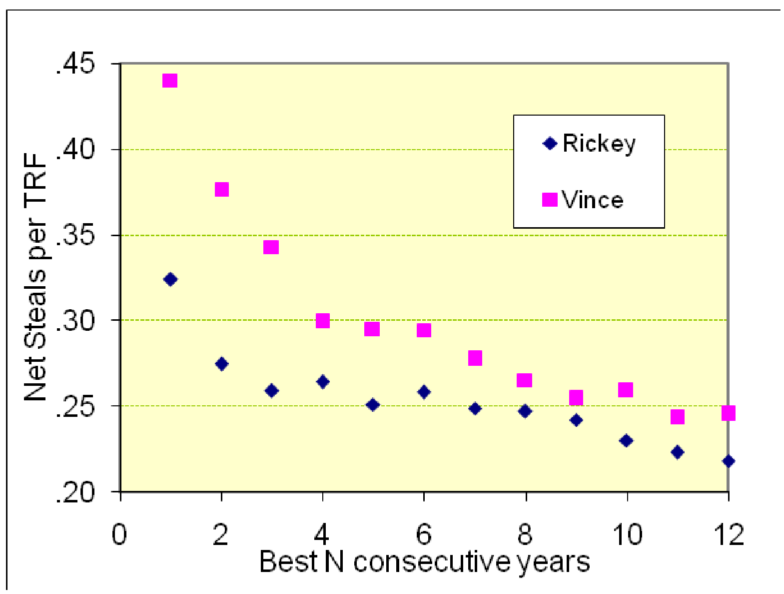


Figure 2 shows Coleman's and Henderson's Net Steals per TRF each year as the aged. Rickey was still a fine base thief in his 30s, although not *as* good as he was in his 20s. Coleman's last full year was age 33; the data for his age 34 and 35 seasons are based on very few games.

Figure 3 – Rickey Henderson and Vince Coleman: Steal rate by consecutive years



So how do we compare these two thieves? I propose that we take the best N consecutive years for each man. That way, Henderson will get credit for whichever years he was at his best, whether or not they were within the same age range that Coleman played. For example, Coleman's best 2 years were in his age 24-25 seasons, in which he garnered 139 net steals while getting to first base 369 times, for a rate of .377. In comparison, Henderson's best 2 years were his age 26-27 years, when he had a Net Steals per TRF rate of .274. These also happen to be the same years; 1985-86! Henderson had other years where he stole more bases, but was caught stealing more often as well (when he set the record with 130 SB in 1982, he was caught 42 times).

Figure 3 shows the comparative rates each man's 1 through 12 best consecutive years. Twelve years covers Coleman's entire career. These years are *not* in chronological order; a player's best four-year set may be completely different than his best single season.

The chart shows that no matter how many years are used, Coleman was the better base stealer, when you account for how often each player reached first base. We obviously cannot extrapolate to how good Coleman *would* have been in his late 30s and early 40s, had he been a talented enough player to have been employed at those ages; we can only say how good he was when he played.

By fortunate coincidence, Henderson's 12 best years of thievery were his age 23-34 seasons, which are also the ages Coleman was in the majors. Table 4 gives their statistics for those years, first as totals, and then per 650 plate appearances.

Table 4 – Rickey Henderson and Vince Coleman batting and base stealing statistics, ages 23 through 34

	AB	H	2B- 3B- HR	BB	SB - CS	AVG	OBP	SLG
Henderson (totals)	5995	1729	298-42-204	1191	906-189	.288	.410	.454
Coleman (totals)	5392	1424	176-89- 28	476	752-177	.264	.326	.345
Henderson (per 650 PA)	542	156	27- 4- 18	108	82- 17	.288	.410	.454
Coleman (per 650 PA)	597	158	19-10- 3	53	83- 20	.264	.326	.345

Henderson played more in these twelve seasons; he was healthier and/or a better quality player. He also reached base far more often, and had a large power advantage over Coleman. He scored 57% more runs during this time, and drove in 94% more. Per plate appearance, Rickey's main advantage was drawing walks while Coleman was making outs, and at times jogging home with a round-tripper instead of reaching on an infield single. But Vince was just as liable to pilfer second base, even though he was on first far less often.

There is no doubt as to who was the better player. But the greatest base thief ever? Not even close. While it's true that he may not have even been scouted as a potential major league hitter in 1930, and while I personally believe he was over-rated in terms of his value to his teams, one has to acknowledge that he had a 12-year MLB career, in which time Vince Coleman showed he was the best there ever was at pilfering a base. Vincent Van Go, king of the base thieves.

Tom Hanrahan, Han60Man@aol.com ♦