By the Numbers

Volume 9, Number 1 The Newsletter of the SABR Statistical Analysis Committee February, 1999

News

Phil Birnbaum, Editor

We have lots of good news this issue.

First, our budget has been increased. SABR directorship has told us that we can spend up to $1000 this year, if we put out four issues of BTN. This is twice what we originally thought we had, and should be plenty enough to produce and mail four issues in 1999, of which this is the first.

That is, if we have submissions – which brings us to our next item of good news. This issue has lots of articles, all of them quite good. Instead of holding some back for next issue, I decided to take a chance and shoot the whole wad at once, in the hope that we'll keep receiving this kind of quality and quantity for the rest of the year. So far, it seems to be working out – I received two more articles while putting this issue together, and promises for more. Thanks, everyone, for your hard work and contributions, and we all hope you'll keep them coming.

In this issue

Thanks are due also to everyone who sent kind words about the last issue. Really, your compliments should be directed towards our October contributors. Thanks again to all of them. I don't know how many of them received private comments on their work, but, regardless, rest assured that my e-mail shows that their submissions were avidly consumed by our membership. Indeed, if you like what you read in this issue, I'd encourage you to drop a line to the contributor – there's nothing a writer likes more than comments from readers. Happier contributors mean more and better research, which is what we joined SABR for in the first place.

Speaking of feedback – I'm very pleased that our members are sending responses to articles that have already appeared. We have three of those this issue, and I'm hoping the array of new material this issue will give rise to even more feedback for the next. As you'll notice, our members aren't shy about criticizing.

Thanks to everyone who agreed to receive BTN by e-mail. The response was better than I expected, and, combined with the budget increase, means that we can offer more stuff in more pages. I hope to continue to put together larger issues like this one, if the articles keep rolling in. As it turns out, photocopying costs much less than postage, so it's you e-mail volunteers who've made this possible.

If you're having problems reading the e-mail BTN, or you have any other technical problems, then, please, let me know and I'll send your copy by mail. We all pay fairly hefty membership dues, and I think that entitles us to postal delivery if we want it. Some of us are already requesting both e-mail and postal mail BTNs, and there's no problem adding you to the list. Just write and let me know.

If you haven't written to me yet to tell me to keep sending BTN, you'll find a little sticker on your envelope reminding you that if you don't drop me a line, this will be your last issue. We want to help SABR save a few bucks, and we don't want to keep sending BTN to members who don't want it any more, or who dropped their memberships. So, please, a postcard or an e-mail will keep you on the BTN mailing list for the rest of 1999.

Thanks, and I hope you enjoy the issue.

You can e-mail me at birnbaum@magi.com. Or, you can write me at #608-18 Deerfield Dr., Nepean, Ontario, Canada, K2G 4L1. ♦

Good Statistics

Phil Birnbaum

If you're a subscriber to SABR-L, our organization's internet mailing list, you've no doubt been entertained by the HEQ controversy. Briefly – and Neal will have the last word on HEQ in his column, on page 5 – one of our committee members, Mike Hoban, created a new statistic, which he presented it at last year's SABR convention. He then published it in *Baseball Research Journal.* Finally, he posted it for SABR members to read, and used it to rank the best players in baseball history.

Then, the [stool] hit the fan.

Over the next couple of weeks, and beyond, HEQ was assailed by a legion of posters questioning its utility. HEQ weights offensive events arbitrarily! they argued. It doesn't take outs into account! Its defensive component doesn't make sense! And on it went, in a surprisingly entertaining debate. Some posters created their own parody stats, including one that divided some offensive agglomeration by the number of letters in a player's name (See, the stat must work – it puts Ruth and Aaron near the top!) and another that added up a players entire agate-type stat line (which, by adding in the year along with the other stats, is certain to take the player's era into account!).

When I was done marvelling at all of this, I started to wonder: instead of arguing specifically about HEQ, and whether it's useful or not, could we come up with some kind of general principle on which HEQ, and other statistics, can be judged? If one of our members sends BTN a new stat of his own devising, is there some basic standard on which we can decide whether we accept it?

In the 1987 *Baseball Abstract*, there appeared a chapter called "Meaningless and Meaningful Statistics." In it, Bill James argued that there are three categories on which the usefulness of a statistic should be judged: Importance, Reliability, and Intelligibility. To have value as an analytical tool, a statistic must measure something significant to winning, be truly representative of the ability it's trying to measure, and be easy to understand.

Bill's idea was meant mostly for everyday statistics, and it was meant to gauge how meaningful they were in the context of evaluating players or teams. Sacrifice hits ranked near the bottom on Bill's list, not because the number of bunts is extraordinarily difficult to compute, but because it doesn't have all that much to do with winning or losing ball games.

What Bill did tells us how *significant* a statistic is – but it doesn't tell us how *useful* it is. How useful is sacrifice hits? Despite its low ranking on the James scale, sacs is a useful piece of information to know. It can be used to tell you about the manager's strategy choices, how well a hitter is thought of, and it gives an indication of how fast the runner is. Even acknowledging that the SH column doesn't relate directly to the quality of the hitter, it certainly gives us some information we want to know.

There's more to baseball statistics than "how good is that player?" And so, I'd suggest that the criteria I'm about to describe can be used to evaluate any stat – HEQ, Total Average, anything -- regardless of what that stat is supposed to measure.

1. What does the stat tell you?

Got a great new stat for this newsletter? Great. What does it do?

Any good statistic tells you something useful. GDP tells you how active the economy was; temperature tells you how warm it is; standard deviation tells you how variable a population is. All are useful things to know. But if you divide GDP by temperature, what does it tell you? It tells you GDP divided by temperature, which is meaningless.

Most of the stats we're used to here in the field of Sabermetrics are designed to tell you how good a player is. As Bill James pointed out, there's nothing more meaningful than that. But good statistics can tell you other things, too. The HR column tells you how many home runs a player hit – and, while that doesn't correlate 100% with offensive ability, it's still pretty interesting in its own right, as evidenced by McGwire mania. And Speed Score, the Bill James creation, tries to tell you roughly how fast a player is – that, too, is something that's good to know.

So, test number one: describe, in English, what your statistic is designed to tell you. If it's something you're interested in knowing, great – move on to step 2. If you're not – if the statistic measures "triples per ninth-inning walk" – throw it away.

Samples: On-base percentage measures ability to get on base (good to know if the bases are loaded). Runs Created measures offensive ability. HEQ measures offensive and defensive ability. Doubles measures the number of times a player got a two-base hit. Total Average, computed as bases per plate appearance, also measures offensive ability.

All these are useful measures, so far.

What does slugging percentage measure? It doesn't measure power – singles hitters often slug for high averages (and besides, power is *extra* bases, not total bases). It isn't normally meant as a measure of offensive production, because it leaves out walks.

So what *does* it measure? Well, nothing, actually, unless you think “bases per at-bat” is something. Slugging percentage actually takes two useful measures – batting average (which measures ability to get base hits) and isolated power (which measures power) – and adds them together. Nice though it may be to add them together, the result doesn’t actually *mean* anything. If slugging percentage were a brand new stat, we’d reject it in a minute. “It doesn’t distinguish between high-average singles hitters and low-average power hitters!” we’d cry. “And why is a triple worth three times what a single is worth? It should be only twice! And how come it doesn’t include walks?”

Slugging percentage is one of those stats that doesn’t quite measure anything real, but sticks around because we’ve all got used to it over the years.

2. What are its units?

A statistic that measures X should produce a result denominated in X.

Weight is the statistic that measures how heavy I am. It can be measured in pounds, stone, kilograms, or tons. But it can’t be measured in feet, points, or percent. If you didn’t know how to rank a bunch of people by weight, you could probably come up with a substitute. You could multiply their height by their calorie intake. You’d wind up with something that correlates to weight; young children would have lower numbers than women, who in turn would have lower numbers than fat men.

But if this statistic is meant to measure weight, it fails. Weight is not measured in inches times calories. It’s measured in pounds. Your hypothetical statistic measures nothing specific – it’s an arbitrary number, not a measure or an estimate.

If you want to get serious, you have to estimate weight in units of weight. And so, you could try to get an estimate of weight, in pounds, by adding height, in inches; two times the waist size; and daily calories divided by 6. You could – but then it would be obvious that the stat didn’t work too well; it underestimates tall people, and overestimates short people, for instance.

Without your stat in the correct units, there’s no way to tell if it works or not – if it correctly measures what you claim it does.

OK, let’s try a few real-life examples.

Runs Created measures bulk offense. Bulk offense is measured in runs. Runs Created gives a number of runs. Perfect.

Power is measured in tendency to hit for extra bases. Isolated Power is denominated in extra bases per at-bat. Correct.

MVP voting success can be measured in award share – percentage of total possible votes. Rob Wood’s formula for MVP voting, explained elsewhere in this issue, predicts award share. Excellent.

Now, let’s try Total Average. TA is supposed to measure offensive ability. Offensive ability is measured in runs per something, or even wins per something. TA gives us bases per something – and so, on this criterion, TA fails.

This is not to say that TA is a bad thing. If what you really want to know is bases per plate appearance, TA is the stat for you. The problem is, first, that I don’t really care about bases per plate appearance, unless it correlates to something I *do* care about. And, second, TA doesn’t claim to be about bases – it claims to measure offensive prowess. As far as I’m concerned, if TA claims that it tells me offensive prowess, it should give me offensive prowess. Sure, bases/PA might be highly correlated to runs, even more so than Runs Created or Linear Weights. But TA stops halfway there, preventing me from evaluating it properly.

Here’s a hard case: Bill James’ Speed Score. Calculated as the average of several diverse categories, like stolen base percentage and range factor, the units are so complex that they’re called “points”. Is that a blow against Speed Score? Actually, it’s not. Speed Score is trying to measure our intuitive idea of how fast a player is. What are intuitive ideas measured in? Points seems to be as good a unit as any, and so Speed Score is fine. (If Bill is trying to estimate *actual* speed, instead of intuitive speed, he’d better give us a stat in miles per hour, or seconds to first base, or something other than points.)

I wouldn’t go so far as to say that a statistic with the wrong units can never be useful. But at best, it’s a definite sign that the stat doesn’t measure what it says it measures. At worst, it’s an attempt to obscure the subject, and to avoid having the stat checked against reality.

3. Does it work?

You’ve got a statistic that’s supposed to predict something. Does it, in fact, predict it accurately? Does it work?

One way to find out is to compare your stat against actual data. Runs Created seems to work – the number of runs it predicts for a team is usually very close to the actual number. In most cases, if you add up the RC estimates for every member of the team, they come close to both the RC estimate for the team, and the actual number of runs the team scored. Plus, RC estimates tend to be in line with runs and RBI totals for individual players.

A second way to show a statistic works is to evaluate the theoretical underpinning. Usually, it’s much easier to undermine the theory than affirm it. Take Total Average, again – it treats a walk the same as a single. Obviously, a single is worth more than a walk, since it always advances baserunners, sometimes two bases, whereas a walk only sometimes advances runners, and one base at the most. This is a valid criticism of TA – if it treats a walk and a single as having equivalent benefit, it must be wrong in at least one of the two cases. That’s not enough to conclude that it’s useless – despite this small flaw, empirical data might show it still works pretty well after all – but it certainly qualifies as a red flag.

Confirming the statistic theoretically is harder, but not impossible. Take, for instance, Earned Run Average. ERA is supposed to be a measure of a pitcher's ability to prevent runs. How do we know it's better than RA (ERA including unearned runs)? The empirical way is by showing that ERA correlation between the same pitcher's different seasons is better than RA correlation – that is, that if we assume a pitcher's ability is constant from season to season, ERA is a more consistent predictor of it. But the easier way is to argue that because unearned runs are obviously independent of the pitcher's ability – which is what we're trying to measure – leaving them out gives us a better-working stat.

The problem with the theoretical approach is that our evaluation of the efficacy of the stat depends on t he truth of the theory. If certain pitchers tend to allow more unearned runs because they give up ground balls that are more difficult to field – well, that throws the theory right out the window. In Sabermetrics, as in any science, empirical data trumps theory. If your stat works, its results should compare favorably to reality.

4. Is there a need for it?

OK, so you've come up with a new stat that predicts offensive prowess, and it seems to work pretty well. But do we need it? Why aren't Runs Created and Linear Weights and Offensive Runs Produced good enough?

There are at least a couple of reasons why we might have a use for your new stat: either it's more accurate, or it illustrates a new relationship.

Whether a statistic works is not a black or white issue. Some statistics work better than others, and some work in different situations. Even the most reliable stats don't work at all in certain cases. Linear Weights predicts that a team that hits .000 will score negative runs. And Runs Created predicts that a team that hits three home runs and makes 27 outs will score 1.2 runs, when, in reality, it will score 3.

No stat is perfect. But the fact that an existing stat doesn't always work perfectly is, itself, insufficient reason to just throw it out and substitute your own stat. Why is your stat better than Linear Weights? If your stat predicts runs within 5%, but Runs Created predicts runs within 2%, then why should we use your stat? It isn't enough to point out Runs Created's flaws – you've got to prove that your new stat is less flawed.

But if your stat shows a new relationship, it might be accepted even if it's a bit less accurate than existing stats. That's another reason why there's room for both Linear Weights and Runs Created. The former shows that runs are related linearly to basic offensive events, while the latter illustrates that runs are roughly proportional to the ability to get on base multiplied by the ability to advance runners. Each stat shows a different relationship, and each stat therefore contributes to our understanding of how baseball works.

On the other hand, Offensive Runs Produced, a statistic introduced by Paul Johnson in the *Baseball Abstract* one year, hasn't survived very well over the years. Why? Because, if you do a little algebra, you'll see that ORP is almost exactly the same stat as Linear Weights (with slightly different weightings for events). It might be slightly more accurate, or slightly less accurate, or (most likely) more accurate sometimes and less accurate other times. But it doesn't tell us anything that Linear Weights didn't already tell us, and so we have no need for more than one of the two.

And so: before you send me your article extolling your great new stat: make sure it measures something useful. Check that its units are those of what it measures. Provide evidence that it works. And, finally, check the literature to make sure there isn't anything already that works better.

If you do all that, there's a good chance your stat will be accepted. And if you don't, it's almost a sure thing it won't. There's an infinite number of ways to add, subtract, multiply and divide baseball numbers to devise a new stat. We're just interested in the useful ones.

Oops?

Well: after I wrote all this, it occurred to me that there's at least one exception (and maybe more; let me know): namely, On-base plus Slugging [OPS]. Although this stat breaks three of the four rules – it doesn't mean anything intrinsically, it has no units, and other stats like Runs Created work much better – the fact that it takes so much of an offensive line into account, and that it correlates well with runs, means that it gets used as a rough measure of batting skill in many studies (including at least one in this issue of BTN). Why? I'm not sure, but perhaps it's because it's easy for non-sabermetricians to understand, being simply the sum of two traditional stats.

Having said that, you'll notice that it gets used where only a *relative* measure is required – to break batters into rough skill categories, for instance. In cases where you need to actually, seriously measure offense, OPS doesn't do you any good – you need something denominated in runs. To paraphrase Bill James, OPS is a stat that's almost always used as a word, rather than as a number. You'll seldom see it added, subtracted, multiplied, or divided. In that sense, it's more like Speed Score than Runs Created.

Phil Birnbaum, 18 Deerfield Dr. #608, Nepean, Ontario, Canada, K2G 4L1; birnbaum@magi.com ♦

Oh, No … Even More About HEQ

Neal Traven, Committee Co-chair

Subscribers to SABR-L, the Society's internet discussion list, have heard a great deal -- more, perhaps, than they wish to -- about the Hoban Effectiveness Quotient (HEQ) measure in the past several months. As co-chair of this committee, I've made a conscious effort to remain "above the fray" (or is it "out of the crossfire"?), in public. In commenting now, my intent is to step back from looking at the flaws in the HEQ system itself, to look instead at the SABR community's reaction to HEQ as well as Hoban's response to that reaction. I hope to identify some places where fundamental differences in research methodology and emphasis fuel the Hoban controversy.

In part, I'll illustrate the points I wish to make by contrasting Hoban's entrance into the SABR consciousness with that of another new voice, Michael Schell. After drawing that comparison, I'll pinpoint some issues of scientific methodology and research collegiality where, dismayingly, I must conclude that Michael Hoban simply doesn't "get it".

To bring committee members who aren't on SABR-L up to speed, I'll first summarize Hoban's and Schell's work (trying, to the best of my ability, to be fair and accurate in the descriptions) and disclose my own interactions with the two researchers. Following that, I'll describe what I see as some explanations for the negative reaction to Hoban in comparison to the response to Schell's work.

As Hoban describes his system, it is supposed to be a "simple, fan-friendly" measure of a player-season, combining offensive and defensive value. Seasons can be combined and averaged to produce career scores. The offensive side truly is simple -- the sum (TB + RBI + SB + 0.5*BB) -- while his defensive measure is not -- combining PO, A, DP, and E using position-specific coefficients and weightings. The method is applied to all seasons without accounting for playing era, ballpark differences, season length (Hoban uses only "total" rather than "rate" stats), or any other adjustments.

Schell's work is more narrowly focused than Hoban's. His aim is to rank-order hitters using a number of statistical adjustment techniques to account for a) late-career decline, b) rules of play and era effects, c) overall quality of talent, and d) ballpark effects. Schell has chosen to measure hitting in terms of batting average, though he fully acknowledges that BA is not the most powerful measure of offensive prowess and has included a section on OBP in his manuscript.

My first contact with Michael Hoban came about two years ago, when he sent out a short summary of his work. Since then, I've seen several longer manuscripts, at least one of which SABR Publications Director Mark Alvarez asked me to review for the *Baseball Research Journal* from a statistical perspective. I must state that my review was decidedly negative, for many of the same reasons that Hoban has been so roundly criticized on SABR-L -- failure to adjust for time and place, use of team-dependent quantities like RBIs, and the like. The SABR public became more aware of Hoban's work when he presented his findings in a research session at last year's convention. In the past several months, Hoban's HEQ has precipitated the aforementioned firestorm of discussions on SABR-L, he wrote about Joe Jackson's career in the latest *BRJ* (I didn't review that one), and I understand that McFarland & Co. will soon release a HEQ book.

It's also been nearly two years since I first heard about Michael Schell. During the 1997 Tony Gwynn lovefest (remember that *Sports Illustrated* cover story?), I was contacted by reporters from the Raleigh *News and Observer*, the Milwaukee *Sentinel*, and other newspapers, seeking comments about the contention that Gwynn would soon supplant Ty Cobb as "the best hitter ever". I learned from the reporters that Schell used era, ballpark, and other adjustment techniques ... but that the underlying measure to which he applied his adjustments was batting average. Furthermore, at that time Schell went out of his way to say that he was avoiding contact with SABR and Sabermetrics while building his models and conducting his analyses. Though he didn't make a presentation there, Schell attended the 1998 convention. We discussed his work briefly, and he offered to send me his full manuscript for review and critiquing. After I read the manuscript, we exchanged several emails discussing the strengths and weaknesses of his work. I believe that responses addressing some of my criticisms were incorporated into the final document that went to Princeton University Press, the publisher of Schell's upcoming book.

On the surface, there are many similarities between the two Michaels. They're both academics (Hoban at Monmouth University in New Jersey, Schell at the University of North Carolina), both in numbers-oriented disciplines (mathematics and biostatistics, respectively), both "outsiders" relatively new to our organization, both writing books describing their approaches, both seeking to make their work known in the public arena beyond SABR, and both drawing conclusions seemingly at odds with SABR conventional wisdom. Why, then, has Hoban raised hackles and attracted widespread controversy while Schell's work has drawn little attention aside from some wistful questions lamenting his choice of a less-than-ideal quantity to measure?

In discussing the factors that have contributed to the rocky reception of HEQ, it is difficult to overlook the issue of personality style. Although Michael Hoban is far from the only stentorian, high-energy, headstrong self-promoter in SABR, it appears to me that the manner in which he has reacted to criticism of his presentation and his SABR-L postings contributes in no small measure to HEQ's difficulties.

But it's more than that. Hoban responds to SABR-L postings critical of his work by either repeating previous statements or applying the same system to additional players. He refuses to engage his critics in any meaningful manner. To cite just one of many examples, he was asked to justify the coefficients in his measure of offense ... specifically, to explain the basis for weighting TB, RBI, and SB at 1.0 while BB has a weight of 0.5. His answer, if that's what we can call it, was (this is an exact quote) "A walk is worth half a point because a single is worth one point." He will not (cannot?) explain the empirical evidence in support of the coefficients, nor will he (can he?) explain how the specific quantities that comprise his metric were selected -- why TB, why RBI, and so forth.

I could enumerate many more issues where Hoban has failed to respond in a meaningful way to SABR researchers. but that would be a pointless and frustrating exercise. The important conclusion to be drawn is that by refusing to enter into a dialogue, Hoban's answer to all criticism sounds like it comes directly from the Queen of Hearts in *Alice in Wonderland* -- "because I say so".

This rhetorical technique ("argument from authority") is simply unacceptable in an empirical research setting. That Hoban has chosen this path demonstrates that he fails to take SABR seriously as a research forum. Frankly, I am astonished that a person holding a distinguished position in academic administration, who must surely have submitted scholarly papers to mathematics journals, would respond to peer review of a manuscript with such condescension.

In contrast, the other Michael has been more than willing to address contentious issues raised by critical readers of his work. He describes the assumptions underlying his models, explains why and how he adjusts his data, tests his results, cites relevant literature. While surely not being as methodologically rigorous as in his biostatistical research, Schell clearly approaches his analyses of batting average in the same scientific, empirical spirit. I am eager to see his methodologic rigor applied to a more useful measure of offense than BA ... and I suspect he's looking forward to it as well.

Rereading what I've written here, I see that its tone is rather more negative than I'd wanted it to be. For that I apologize. I could, perhaps, close with the Rodney King question -- "Why can't we all just get along?" -- but instead I'll just remind us all that respect is a two-way street. Those who criticize someone else's work need to show respect for the efforts of the person who presents that work, and those whose work is criticized need to respect the value of the critiques. Improving our ability to analyze and model what happens on the baseball field is an iterative process.

Neal Traven, 500 Market St. #11L, Portsmouth, NH, 03801; 603-430-8411; baseball@ttlc.net ♦

Informal Peer Review

Last issue, I put out a call for members to volunteer for occasional, informal peer review of articles to appear in BTN. I thought I got quite a few responses, but after reviewing my mail, I can find only four names.

That means that I've obviously misplaced some: and so, with apologies, if you already wrote, and you're not one of the four I've found (Rob Wood, Keith Carlson, John Matthew, or Jim Box), I ask that you write me again.

I promise that this time, I'll do it right, and a full list of volunteers will appear in the next issue of BTN, for sure.

Once the list is published, contributors will be able to contact peer reviewers on an as-needed basis – that is, if you want someone to look over your manuscript in advance, these people are willing. Of course, I'll be doing a bit of that too, but, as much as I'd like to, I don't have time to contact every contributor with comments on their work. (I will get back to you on more serious issues, like if I don't understand part of your method or results.)

So: if you're interested in being on the list, send your name, e-mail address, and areas of expertise (don't worry if you don't have any – I certainly don't), and you'll see your name in print next issue.

Commentary on "A Leadoff Hitter Rating System"

Herm Krabbenhoft

A recently published study on leadoff hitters, the author argues, has "no merit": the rating criteria are "unsupported and arbitrary", and many of the hitters ranked did not actually lead off.

There are numerous serious problems with the "Leadoff-Hitter Rating System" presented by Felber last issue of *BTN*.[1]

Over the past dozen years I have engaged in comprehensive research on leadoff batters.* So, the title of Felber's article elicited considerable interest for me. And, the article's abstract intensified that interest. Unfortunately, Felber's proposal, interpretation, and execution did not deliver. Here are some of my thoughts and comments -- with supporting documentation -- on the shortcomings of Felber's "Leadoff-Hitter Rating System."

First of all, a leadoff-hitter rating system -- by definition -- rates leadoff hitters. Thus, the fundamental requirement of Felber's research should have been to determine the identities of the (principal) leadoff hitters -- not to arbitrarily present lists of players who may have been leadoff hitters. Only after having ascertained the identities of the leadoff hitters should one proceed to rank them. However, Felber stated, "I include players who it would make sense to me to have been used as leadoff hitters; ... I have also excluded players who do very well on this system, but who would make illogical leadoff candidates."

The three measurable characteristics singled out by Felber to critically evaluate leadoff hitters have, in fact, been employed historically to rank *all* players -- on base percentage was introduced in the middle 1950s; "to second" is merely a subset of total average (introduced in the 1980s); and "runs" is a variant of run-getting (which was commonly used in 19th century statistical compilations). These three characteristics are not unique to leadoff hitters. However, each of these criteria may be utilized for the ranking of leadoff batters -- *provided that only bona fide leadoff hitters are in the domain.* Also, in order to be included in the leadoff hitter domain, some meaningful minimum qualification requirement must be met, such as a specified number of leadoff batter games. Felber apparently had no such qualification standards.

Felber's unsupported and arbitrary use of the 1.2 and 0.8 factors for relative on base average and relative run scoring ability -- i.e. his "editorial judgment" -- is totally inappropriate. Such factors should only be used after verification with sound statistical analysis (e.g. via regression analysis).

On account of these (and other) serious flaws in Felber's Leadoff-Hitter Ranking System, I conclude there is absolutely no merit to it.

What follows are some specific issues that I have with Felber's System.

For the 1997 NL and AL Leadoff Hitters lists provided by Felber, 14 players are included for each league. Felber states "one player per team" -- although no team affiliations were given. So, is one to assume that the players listed were their team's best (or principal) leadoff-hitter? The facts (presented below) reveal such an assumption to be dismally poor.

Looking at the NL first, here are some facts from my research.[2]

- Rickey Henderson of the San Diego Padres compiled a NL leadoff hitter score of 403.05 (calculated according to Felber's formulas). Henderson was San Diego's principal leadoff hitter with 76 first-up games. (Rickey was also the leadoff batter for the AL's Anaheim Angels in 29 games). Henderson's NL leadoff hitter score was significantly better than that of Padres teammate Quilvio Veras (who appeared in just 57 games as a leadoff hitter -- most of them after Henderson left the Padres in August). Felber incorrectly placed Henderson in the AL list; *Rickey belongs on the NL list.*

- Walt Weiss of the Rockies was included by Felber. Yet the fact is that he was the starting leadoff batter for Colorado in only 32 games. The principal leadoff batter for the Rockies was Eric Young (114 leadoff games). Young also appeared as the Dodgers' first-up batter in 25 games. Perhaps Felber considered Young to be the Dodgers' principal leadoff batter. *But*, ...

- The Dodgers' principal leadoff batter was, in fact, Brett Butler (76 leadoff games).

- And then there's Pokey Reese of the Reds. He was included by Felber even though he was the starting leadoff batter in only 35 games. The principal leadoff batter for Cincinnati was Deion Sanders (111 leadoff games). Sanders, who was not included by Felber, fashioned a leadoff-hitter score substantially greater than Reese.

- And what about Felber's choice of Carl Everett of the Mets? He was the starting leadoff hitter in 50 games. However, Lance Johnson batted first in 62 games for the Mets (and in 33 games for the Cubs). Since Felber included both Everett and Johnson in his NL list, perhaps Felber felt that Everett

* Editor's note: An extensive list of some 25 of Mr. Krabbenhoft's studies on leadoff hitters (most of them published in *Baseball Quarterly Reviews*) is available from Mr. Krabbenhoft or myself (Phil).

should represent the Mets and Johnson the Cubs. This is pure nonsense because ...

- It results in Brian McRae being left off the NL list by Felber. McRae batted leadoff in 70 games for the Cubs (and 26 games for the Mets).

comprised of considering three separate statistical categories: (1) leadoff batting average; (2) overall on base percentage; and (3) overall total average. Furthermore, to ensure that the rankings are meaningful, I define "Principal Leadoff Batters" as those players who appeared as leadoff batters in at least 75% of their team's games; only these players officially qualify for the leadership positions in the above three statistical departments.

Table 1. 1997 National League Principal Leadoff Batters

Team	Player	leadoff games	leadoff at bats	leadoff runs	leadoff hits	leadoff rbis	leadoff bat ave	overall games	overall bat ave	overall OB%	overall TA
ATL	Lofton	121	487	90	163	47	.335	122	.333	.409	.791
CHI	McRae *	70	290	36	68	17	.234	108	.240	.329	.656
CIN	Sanders	111	460	52	125	24	.272	115	.273	.329	.710
COL	Young *	114	465	76	127	43	.273	118	.282	.363	.750
FLA	Castillo	51	208	20	49	4	.236	75	.240	.310	.481
HOU	Biggio	156	614	145	191	80	.311	162	.309	.414	1.062
LA	Butler	76	292	45	89	17	.305	105	.283	.363	.658
MON	Grudz'lanek	107	458	51	118	33	.258	156	.273	.385	.662
NY	Johnson *	62	254	42	76	21	.299	72	.309	.307	.603
PHI	Cummings *	48	184	22	58	20	.315	63	.303	.369	.747
PIT	Womack	148	635	83	176	50	.277	155	.278	.326	.712
STL	DeShields	117	472	74	135	48	.286	150	.295	.357	.841
SD	Henderson	76	284	58	78	25	.275	88	.274	.422	.945
SF	Hamilton	110	444	74	119	40	.268	125	.270	.354	.665
total	McRae	96	396	52	93	21	.235	153	.242	.326	.653
total	Johnson	95	396	59	118	34	.298	111	.307	.370	.734
total	Young	139	569	97	156	56	.274	155	.280	.359	.746

Notes: (1) An asterisk after a player's name indicates that he played for two teams -- McRae (CHI and NY); Young (COL and LA); Johnson (NY and CHI); Cummings (PIT and PHI); the combined performance is given at the end of the table; Cummings did not lead off any games for the Pirates. (2) OB% are those reported in the 1998 edition of The Sporting News Baseball Guide; (3) Total Average (TA) was calculated according to the formula given in the 4th edition of Total Baseball.

Another serious flaw with Felber's research is that he did not utilize correct plate appearance statistics. Thus, ALL of Felber's calculated "To 2nd," "Runs," and the derived final "Score" values are inaccurate. For example, Craig Biggio's 1997 plate appearances are 744 (not the 703 used by Felber); similarly, the NL's plate appearances are 87,353 (not the 84,907 used by Felber). Thus, Biggio's "To 2nd" and "Runs" values are 164.49 and 129.93, respectively (not 169.26 and 134.63 as reported by Felber); Biggio's final "Score" is 443.97 (not 453.44 given by Felber).

In marked contrast to Felber's system, the approach that I have used to evaluate leadoff batters consists of first ascertaining the principal leadoff batter for each team for each season. This fundamental objective is achieved by examining the boxscore of each major league game for every season -- every leadoff batter is recorded along with his at bats, runs, hits (and runs batted in). With this solid data base of leadoff batter identity and performance, I have been able to reliably rank these players. The particular modus operandi I employ to rank leadoff batters is

Concurrently, I also believe it is instructive to compare the performances of each team's principal leadoff hitter. To illustrate this methodology, the Principal Leadoff Batters for the 1997 NL season are listed in Table 1.

Inspection of the table reveals that the leadoff batting champion was Kenny Lofton. The overall on base percentage leader among the principal leadoff batters was Rickey Henderson. And the overall total average leader among the principal leadoff batters was Craig Biggio. Only three of the fourteen principal batters amassed 122 or more leadoff games (i.e. 75% of each team's 162-game schedule) -- Biggio, Womack, and Young.
Comparison of the players listed in Table 1 with Felber's NL list reveals that Felber's NL list is 50% wrong in terms of leadoff batter makeup. Clearly, Felber's "System" fails to provide complete, accurate, and meaningful conclusions for the NL.

Turning now to the 1997 AL leadoff batters, Felber's "System" also fails miserably. Without going into the details here, it is pointed out that Felber's "System" omitted/excluded the following 6 principal leadoff hitters -- Tony Phillips (ANA), Otis Nixon (TOR), Derek Jeter (NY), Damon Mashore (OAK), Mark McLemore (TEX), and Jose Offerman (KC).

As poor as Felber's "System" is for the single 1997 season, it is perhaps even worse for his "Top 50 Leadoff Hitters of All Time." In addition to the egregious flaws pointed out above, Felber chooses to establish a ranking based on a player's three-year score. This is fundamentally unsound. A few examples will suffice to completely invalidate Felber's "Top 50 Leadoff Hitters of All Time."

- Rod Carew is Felber's #14 based on his 1973-1975 performance. The facts are that Carew was a leadoff batter in just 62 games in those 3 seasons -- less than 13% of the games played by his team.

- Max Carey is Felber's #24 based on his 1921-1923 performance. The facts are that Carey did not lead off in even one game during this entire 3-year period.

- Craig Biggio is Felber's #31 based on his 1995-1997 performance. The facts are that during the 1995 and 1996 seasons Biggio was a leadoff batter in only 26 games -- only 8% of the games played by the Astros.

Altogether, there are at least 13 players on Felber's "Top 50" list who do not merit inclusion because they were not principal leadoff hitters during the 3-year periods arbitrarily chosen by Felber. Moreover, among the actual all-time leadoff batters not included by Felber are: Pete Rose, Eddie Yost, Richie Ashburn, Max Bishop, Lloyd Waner, Brett Butler, and Charlie Jamieson.

In summary, I reiterate that Felber's "Leadoff-Hitter Rating System" (a) is not based on "*specific*, measurable characteristics" unique to leadoff hitters; (b) does not identify leadoff hitters (i.e. distinguish leadoff batters from players in other batting slots); (c) does not viably evaluate leadoff batters.

Persons interested in factual treatment of leadoff batters throughout the history of major league baseball are encouraged to request an extended reference list from the author.

Footnotes

1. "A Leadoff-Hitter Rating System," B. Felber, *By The Numbers,* **8** (1) 5 (1998).
2. "Leading Leadoff Hitters (1997)," H. Krabbenhoft, *The 1998 BQR Yearbook,* in press (1998).

Herm Krabbenhoft, Baseball Quarterly Reviews, P.O. Box 9343, Schenectady, NY, 12309, bqr9343@aol.com. ♦

Is This Your Last Issue?

If your envelope has a little label on the back saying this is your last issue, you need to subscribe. There's no charge – your membership in the Statistical Analysis Committee and the subscription to BTN are all included in your SABR membership fees. However, because of member turnover and budget squeezes, we want to make sure we're not sending BTN to anyone who doesn't want to receive it.

So: if you got the label, but you want to keep receiving BTN, just drop me a card or e-mail, saying, "Yes, Phil, I'm still a SABR member and I'd like to keep receiving BTN." (If you want to receive BTN by e-mail, just say so. Details on e-mail subscriptions are elsewhere in this issue.)

If you do *not* have that label on your envelope, it means you've already contacted me, and your subscription is just fine.

Re: "A Leadoff Hitter Rating System"

Clifford Blau

The author makes two points regarding Bill Felber's leadoff rating system: first, that it may not make sense to evaluate leadoff hitters differently than batters in other batting-order positions; and, second, that including runs scored in the formula is redundant.

I have two main points to make regarding Bill Felber's article in the October 1998 issue of *By The Numbers*, "A Leadoff-Hitter Rating System."

First: does it make sense to rate leadoff hitters differently than other batters? Most studies of lineups I have seen, such as those by Mark Pankin and Bill James, suggest that the batting order does not make a large difference in the number of runs that a team scores. In these studies, the differential between the best possible and worst possible lineups is on the order of 20-30 runs per season, while the difference between the best lineup and a conventional one is much smaller, perhaps 5 to 10 runs per year. These figures assume a normal distribution of talent in the lineup. But suppose one had a team with nine hitters whose on base plus slugging were identical, say .750. It would be interesting to see if such a team would score significantly more runs with a leadoff hitter who has a .400 on base average and a .350 slugging average versus one with figures of .300/.450. It seems to me based on the above-referenced studies that there would be very little difference. If this is so, the conclusion I would draw is that the specific skills thought to be desirable in a leadoff hitter are much less valuable than are general offensive skills. That is, a Juan Gonzalez type would be a better leadoff hitter than a Brett Butler type, despite a low on base average. Therefore, leadoff hitters should be evaluated in the same way as any other. (After writing the above, I obtained the STAR Baseball Simulator at http://www.geocities.com/ResearchTriangle/Thinktank/4884 and conducted the above experiment. Through 20,000 simulated seasons, there was no difference in the scoring of the two lineups.)

Second: if I am wrong and leadoff hitters should be evaluated differently than others, I do not think Mr. Felber's system is as good as it could be. The three parts of his formula are the ability to reach base, the ability to reach scoring position without aid, and the ability to score runs. It seems to me that the third part does not belong, since it is largely a product of the first two abilities, as well as the performance of the following batters. Bill James, in his 1983 *Baseball Abstract*, introduced a formula for estimating the number of runs scored by a leadoff hitter based on the number of times he reached each base on his own. This shows that runs scored is generally predictable based on the first two factors in Mr. Felber's system. It seems to me that the third factor should be the player's ability to score runs given the first two skills. This would be some measure of baserunning, rather than basestealing skill. Since data are not normally available, perhaps a surrogate such as basestealing or speed score could be used. In addition, there are several charts available which show the probability of scoring and average runs produced from each base-out combination. These data could be used to come up with the proper weighting of the three factors.

Clifford Blau, 16 Lake Street, #5D, White Plains, NY, 10603. cliffordblau@geocities.com ♦

Canada's Postal Service Not As Bad As It Looks

Although the postmark on the last BTN said October 9, it was actually mailed on *November* 9. Canada Post is sometimes slow, but not *that* slow. My understanding is that most of you received your BTN within 7 to 10 days of mailing.

This message was brought to you by national pride.

"Big Bad" Flaw Wasn't

G. Jay Walker

An article in last issue's BTN, by Clem Comly, argued that the "Big Bad Baseball Annual" contained a flaw in its method of solving a problem with the "Offensive Wins Above Replacement" statistic. Here, author G. Jay Walker argues that the article was not intended as a solution, but as a jumping-off point to discussion. He acknowledges that Mr.Comly's analysis is preferable to the original, and promises an even better solution in a yet-to-be-published book.

Feedback and criticism are an important part of analysis, and the Big Bad Baseball Annual (BBBA) is glad to hear from our readers, even when they point out things that we haven't gotten quite right. However, Clem Comly detracts from an otherwise interesting analysis by his chosen theme in a recent article ("A Flaw in the Big Bad Baseball Annual" (BTN—October 1998)). In fact, what he cites as a flaw is not a flaw at all.

In the "Beauty With a Blemish" article, we presented some hypothetical data for purposes of introducing the reader to a set of problems in the current method for calculating Offensive Wins Above Replacement (OWAR). Appearing on the second page of a ten-page article, the data example was in no way presented as a solution to the OWAR problem, but merely served as a jumping-off point for additional analysis. Our proposed solution to the problem actually came later in the article.

It appears to us that Clem misinterpreted the intent of that passage as if it was, in fact, our solution. As it turns out, his "correction" to BBBA's "flaw" is actually his own solution to the OWAR problem. Along with several other correspondents who wrote to us directly, Clem persuasively argues that a solution to OWAR should be based on the reallocation of outs to the team. After reflection and discussion, we have come around to this idea as being preferable to our proposed solution in "Beauty With A Blemish," which was to use plate appearances instead of outs.

In part because of dialogues with those readers who made an effort to contact us with their ideas, BBBA has move beyond the OWAR issue and has developed a new and improved offensive evaluation system, called Extrapolated Runs/ Extrapolated Wins (XR/XW). The new statistic will debut in the 1999 BBBA in late February.

G. Jay Walker, 3731½ Villa Terrace, San Diego, CA, 92104, walker@nhrc.navy.mil. ♦

Submissions

Submissions to *By the Numbers* are, of course, encouraged and drooled over. 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 (but no death threats, please) are all welcome.

Articles should be submitted in electronic form, either by e-mail or on PC-readable floppy disk. I can read most word processor formats. If you send charts, please send them in word processor form rather than in spreadsheet form; this will make it easier for me to format your charts for publication. Unless you specify otherwise, I may send your work to others for comment (ie, informal peer review).

I usually edit for spelling and grammar. (But if you want to make my life a bit easier: please, use two spaces after the period in a sentence. Everything else is pretty easy to fix.)

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

What Drives MVP Voting?

Rob Wood

Quesionable MVP awards, at least by Sabermetric standards, are not an uncommon occurrence. Why do the voters cast their ballots the way they do? In this article, the author examines past patterns of MVP voting, and comes up with a rough formula to show how a player's batting line can predict how many votes he'll get.

Introduction

The 1998 league MVP awards went to players that had very good credentials, but still caused some consternation in several quarters. Sammy Sosa beat out Mark McGwire by a wide margin (30 to 2 first place votes) in the NL, and Juan Gonzalez beat out a good crop of candidates in the AL.

As you know, Sosa led the league in total bases, runs scored, and runs batted in. McGwire led the league in home runs (no kidding), walks, on base average, and slugging percentage, not to mention most of the sabermetric measures such as runs created and total player rating. Of course, the Cubs made the playoffs via the wild card while the Cardinals did not.

Do not think that controversial MVP awards are a modern phenomenon. Alas, the history of baseball is sprinkled with surprising MVP awards. To name but a few: Frank McCormick in 1940, DiMaggio over Williams in 1947, Hank Sauer in 1952, Yogi Berra in 1954, Elston Howard in 1963, Zoilo Versalles in 1965, Roberto Clemente in 1966, Boog Powell in 1970, Don Baylor in 1979, Andre Dawson in 1987, Robin Yount in 1989, Barry Larkin in 1995.

In this article, I do not want to focus on any specific MVP injustices, though clearly McGwire would have received my vote. Instead, I hope to analyze past voting to see if we can determine what actually drives MVP balloting. In particular we can test whether leading the league in RBI, as Sosa and Gonzalez did, actually has the largest impact as many people believe. By so doing, we will be able to better understand what the voters think "Most Valuable Player" really means.

I have compiled a database of each of the 67 seasons' complete MVP balloting by the BBWAA from 1931-1998. Following tradition I use a player's MVP Award Share as the data we are attempting to understand. Award share is the fraction of all possible points that a player receives, where the maximum is 1.00 if he is a unanimous MVP (i.e., received all the first place votes). I am not sure how the votes were cast and tallied over the years, but currently a player receives 14 points for a first place vote, 9 points for a second place vote, 8 for 3rd, 7 for 4th, …, down to 1 point for a 10th place vote. Using award shares is necessary for historical analysis as the number of voters, and thus the raw number of MVP points, has changed over the years.

Award shares range from 0 to 1. The average award share across all the 134 league MVPs is .85. The highest award share of 1.00, of course, was received by the 13 unanimous MVP winners (Greenberg 1935, Rosen 1953, Mantle 1956, Robinson 1966, Cepeda 1967, McLain 1968, Jackson 1973, Schmidt 1980, Canseco 1988, Thomas 1993, Bagwell 1994, Caminiti 1996, Griffey 1997). The lowest award share received by an AL MVP was .55 by Yogi Berra in 1951, and the lowest by an NL MVP was .57 by Marty Marion in 1944.

The "explanatory" variables that I will use to try to explain MVP award shares are the league leaders in offensive categories, as well as where the team finished in the standings. The analysis will be purposefully coarse as I am not attempting to determine who actually deserved each MVP, but the underlying reasons why the player won. For example, I will not use any measure other than leading the league, so finishing second in a category gives you no credit in my model. This is admittedly crude, but I want to first get an idea of what is driving things. Subsequent research could expand and/or refine my model.

League Leaders

Let's first take a look at the raw data. Table 1 reports how the league leader in the offensive categories fared in that season's MVP balloting. You can see that I include all of the traditional offensive categories and several of the new sabermetric categories. The table is ordered by how often the MVP winner led in each offensive category.

The first row of the table indicates that the player who led the league in RBI won the MVP 38% of the time, finished in the top 3 in MVP balloting 64% of the time, finished in the top five 77% of the time, and garnered an average MVP award share of .58.

The only variable that may need explanation is Total Player Rating. This is Pete Palmer's Total Player Rating (TPR) among hitters only, denoting the player who contributed the most wins to his team using Pete's linear weights formulas for hitting, base running, and fielding. For the purposes of Table 1, if there was a tie for the league leadership in an offensive category, among those players tied the player who received the highest award share was used.

The table of league leaders' MVP showing within each league looks remarkably similar. Indeed, I have not identified any significant differences between the two leagues' MVP balloting. Throughout the article, all data and results pertain to both the AL and NL (i.e., the combined dataset of 134 league-seasons).

For comparison to the percentages listed in the table, the MVP winner played on a team who finished first 68% of the time. Thus, about twice as often as the MVP was the RBI league leader, the MVP played on a first place team. Of course, many players play on a first place team, and divisional play has spawned multiple first place teams. Thus, it is not obvious who on the first place team will receive all the "credit" for finishing first. We will come back to this issue below.

All data used in this article is from the Bill James/Stats Inc. All-Time Baseball Sourcebook published in 1998. In this great new book, James & Stats have slightly re-worked the runs created formulas, and thus have some new league leaders in runs created than previously published . I use their designated league leader for all of the offensive categories.

Table 1: League Leaders and the MVP (1931-1998 AL & NL)

	Won MVP (%)	Top 3 in MVP (%)	Top 5 in MVP (%)	Average Award Share
RBI	38	64	77	.58
Total Player Rating	35	52	66	.52
Runs Created	33	63	80	.58
Slugging Pct	32	61	76	.57
OBA + SLG	32	61	73	.55
Total Bases	29	65	84	.58
Runs Created/27 Outs	29	59	75	.55
Home Runs	26	57	73	.53
Runs	23	46	65	.45
Batting Avg	18	47	58	.43
On Base Avg	18	35	48	.38
Hits	11	42	57	.40
Doubles	10	27	41	.30
Walks	8	21	35	.25
Triples	7	18	25	.19
Stolen Bases	4	9	15	.12

The table above shows that RBI does indeed lead the parade. But "only" 38% of the time does the league leader in RBI win the MVP. RBI is followed fairly closely by Total Player Rating, Runs Created, Slugging Percentage, and On Base Average Plus Slugging Percentage. We next try to make sense of the data.

MVP Model Data

Table 1 merely listed the number of times that the MVP led the league in various offensive categories. While this is a first step, it does not tell us what really drives MVP balloting. Of course, in many years one player leads the league in multiple offensive categories. We want to identify which categories are the true drivers and which are merely along for the ride.

I use the entire history of BBWAA MVP balloting in order to ensure that the data is sufficiently rich. We need rich and varied data to help us sort out true effects from spurious effects. Several characteristics of the dataset are desirable. First, the data should exhibit a wide range of quality of players to lead the league in each offensive category. If so, a true measure of the impact of leading the league can be estimated. The alternative is a case in which a "Babe Ruth" leads the league in a category every year and wins the MVP; did Babe win the MVP because he led the league or because he is the Babe?

While most league leaders in most offensive categories merit some degree of MVP consideration, over the years some fairly nondescript players have been league leaders. The following litany probably deserved and received only scant MVP voting for leading their league: Ivan DeJesus (runs 1978), Garry Templeton (hits 1979), Mike Hargrove (walks 1978), Matty Alou (doubles 1969), Mariano Duncan (triples, 1990), Dave Kingman (home runs 1982), Reggie Smith (total bases 1971), Larry Hisle (RBI 1977), Willie Wilson (batting average 1982), Vince Coleman (stolen bases, many years), Darrell Evans (runs created 1973), Alvin Davis (runs created per 27 outs), Kal Daniels (on base average 1988), Harold Baines (slugging percentage 1984), and Roy Smalley (total player rating 1979). The point is that over the entire history of BBWAA MVP voting, we have seen a wide variety of the quality of league leaders in virtually every offensive category. The analysis that follows will attempt to sift through all the data to uncover accurate estimates of the impact of leading the league on MVP voters.

Second, the impact of a few players should be minimized. Suppose a hypothetical player, call him Paul Bunyan, won the MVP many times. Then the offensive characteristics of Paul Bunyan could dominate the MVP impact estimates. For example, if Bunyan led the league each year in triples, as well as home runs, RBIs, etc., then our model could tell us that leading the league in triples was important to MVP voters. Over the course of 67 years, the impact of any individual is minimized. And no players have ever won more than three MVP awards (those with three are Foxx, DiMaggio, Berra, Mantle, Musial, Campanella, Schmidt, and Bonds).

Third, the data should be sufficiently rich and varied so as to not exhibit multicollinearity. This means that over the course of the history, the league leaders in two or more categories do not overlap too much. For example, suppose that the league leader in home runs is virtually always also the league leader in total bases. Then, the model would not be able to separate out these two impacts. Fortunately, the 134 league-seasons of data is sufficient to mitigate against multicollinearity. In the database, the offensive categories paired by league leaders the most were RC/27outs & OBA+SLG (107 times), SLG & OBA+SLG (100 times), and RC & RC/27outs (93 times). Among the traditional categories, the same player led the league in both home runs and slugging percentage 71 times and both home runs and RBI 70 times.

MVP Model Estimation

Now that we are satisfied that the database is sufficiently rich and varied to permit analysis, let's turn to our model and its estimation. Since award shares are necessarily between 0 and 1, a standard linear regression is not appropriate. This is a reflection of the fact that the award share relationship is non-linear. Leading the league in multiple categories (in the extreme, all categories) exhibits diminishing returns in terms of award share. If a player leads the league in HR, RBI, and batting average, also leading the league in hits or runs will not add much to his award share. But a player who only leads the league in one category will get the full benefit.

I used a transformation to ensure that the model predictions are between 0 and 1. In particular, I used a log-linear transformation based upon the cumulative distribution function of the standard exponential distribution: $F(x) = 1 – \exp(-x)$. This function starts at 0 and climbs gradually up to 1 as x increases from 0 to infinity.

The model we estimate is

```
Share = 1 – exp[-(C1*L1 + C2*L2 + … + Cn*Ln)]
```

where the Li's are the explanatory variables, described below, and the Ci's are the coefficients we are attempting to estimate. All of the explanatory variables are dummy variables, taking on value 1 if true and 0 otherwise, and are listed in the table below.

Besides the league-leading variables, I have included variables for team finishes: first place (either division or league), making the playoffs but not finishing first (i.e., wild card or losing a tie-breaking series or game), and second place (either division or league). In order to make the model manageable, I parcelled out these team finishes' credits to three or fewer players on each team. In most years only one or two players is held responsible for a team's success, either rightly or wrongly. I have attempted to mimic this in these variables.

A regression analysis can be used to estimate the C's since the model can be transformed to a standard linear model via a logarithmic transformation yielding:

```
Log[1/(1-Share)] = C1*L1 + C2*L2 + … + Cn*Ln.
```

Of course, our goal is to estimate how much leading the league in each offensive category contributes to MVP award share. Clearly the C's are not directly that answer. Since we are using a log-linear model, we need to transform our coefficients (C's) into award share weights (W's). The transformation, as indicated by the original model, is given by $W_i = 1 – \exp(-C_i)$.

Table 2 presents two sets of estimates. The first pertains to a "traditional" model in which the sabermetric variables are excluded. The second pertains to the comprehensive model which includes all traditional and sabermetric variables. For both models, the "raw" coefficient (Ci) is listed along with the transformed award share weight (Wi). The analysis was constrained to produce non-negative coefficients and weights.

In presenting the two model estimates, I am not suggesting that voters actually look at the new sabermetric league leaders, but rather that the sabermetric variables may reflect performances not captured in other variables. For the remainder of the article I will use the "traditional" model.

The table indicates that leading the league in RBI is the most important MVP award share variable. The raw coefficient of .84 translates into receiving an award share of .57. Coming in first place, and being one of the recipients for the credit, has a raw coefficient of .75 which translates into an award share of .53.

Table 2: Model Coefficients and Weights

	Coefficient (w/o sabermetric variables)	Award share weight (w/o sabermetric variables)	Coefficient (with sabermetric variables)	Award share weight (with sabermetric variables)
RBI	.84	.57	.83	.56
Team finished in 1st*	.75	.53	.74	.52
Slugging Pct	.56	.43	.36	.30
Team finished in "1.5st"*	.42	.34	.43	.35
Runs Created			.31	.27
Runs	.37	.31	.30	.26
Total Player Rating			.30	.26
Total Bases	.30	.26	.22	.20
Batting Avg	.26	.23	.22	.19
Team finished in 2nd*	.18	.17	.17	.16
Home Runs	.12	.12	.10	.09
On Base Avg	.11	.10	.08	.00
Hits	.10	.10	.09	.09
Runs Created per 27 Outs			.04	.04
OBA + SLG			.03	.03
Stolen Bases	.02	.01	.00	.00
Doubles	.00	.00	.00	.00
Triples	.00	.00	.00	.00
Walks	.00	.00	.00	.00

*at most 3 players receive this credit per team; "1.5" refers to tieing for first place and then losing a playoff or winning a wild card.

Let's use these two examples to demonstrate how to combine variables. Of course, we cannot simply add their award share weights. For one thing, .57 plus .53 exceeds 1.00, the maximum possible. And secondly, we explicitly developed a non-linear model precisely since we knew that the underlying relationship is non-linear. To find the estimated award share of a player who leads the league in RBI and whose team finished first is to add the raw coefficients, and then convert the sum via the exponential transformation. This method is used no matter how many offensive categories the player leads the league in.

Table 3 presents a table of conversions for reference. To complete the example of a RBI leader who finished first, add the raw coefficients .84 plus .75 to yield 1.59. Either by looking at Table 3 or doing the exponential transformation (1-exp(-x)), we get an estimated award share of .80. Historically, an award share of .80 would win the MVP most years.

Model Predictions

In this section, I want to use the model to identify the most under- and over-predicted MVP shares in history. Remember, as this is a crude model relying mainly on leading the league in offensive categories, being under-predicted is not necessarily the same thing as receiving "too much" MVP consideration. Similarly, being over-predicted is not the same as receiving "too little" MVP consideration. But such questions can be addressed with the model's estimates.

Table 3: Conversion of Coefficients to Shares
Share = 1-exp(-Sum of Coefficients)

Sum of Coefficients	Predicted MVP Share
.00	.00
.20	.18
.40	.33
.60	.45
.80	.55
1.00	.63
1.50	.78
2.00	.86
2.50	.92
3.00	.95
3.50	.97
4.00	.98

Table 4 shows the most under-predicted MVP shares using the traditional model. That is, these are the players who received more MVP shares than the traditional model predicted for them based upon the league leaders and team finishes.

Table 4: Top 5 Under-Predicted MVP Shares

	Year	League	Actual MVP Rank	Actual MVP Award Share	Predicted MVP Award Share	Error in Prediction
Roberto Clemente	1966	NL	1	.78	.00	.78
Robin Yount	1989	AL	1	.65	.00	.65
Charlie Gehringer	1937	AL	1	.98	.36	.62
Bob Elliott	1947	NL	1	.61	.00	.61
Luke Appling	1936	AL	2	.81	.23	.58

Leading the list is Robert Clemente in 1966. Clemente won the NL MVP in a close vote over Sandy Koufax in what turned out to be Sandy's final season. Clemente had a fine year with 202 hits (3rd in league), a career high 105 runs (4th), 31 doubles (5th), 11 triples (3rd), a career high 29 home runs, 119 RBI (2nd), 342 total bases (2nd), and a league-leading 17 outfield assists.

This case, like the others on the list, reflect MVP voters choosing a worthy candidate who did not happen to lead the league in many offensive categories. Each of the MVP winners on the list was not an undeserving recipient. Indeed, it is in years in which there is no obvious winner that we may learn the most about what voters look for in an MVP. In short, these were years in which whoever won the MVP would be deemed "under-predicted" by the model.

Table 5: Top 5 Over-Predicted MVP Shares

	Year	League	Actual MVP Rank	Actual MVP Award Share	Predicted MVP Award Share	Error in Prediction
Duke Snider	1956	NL	10	.16	.79	.63
Dick Stuart	1963	AL	13	.09	.68	.59
Al Rosen	1952	AL	10	.15	.73	.58
Gus Zernial	1951	AL	20	.04	.62	.58
Hank Aaron	1960	NL	11	.16	.73	.57

Table 5 presents the top 5 most over-predicted MVP shares using the traditional model. That is, these are the players who received less MVP shares than the traditional model predicted for them based upon the league leaders and team finishes.

Leading the list is Duke Snider in 1956. Duke led the league in home runs, on base average, slugging percentage, and walks. Of course, the Dodgers won the pennant too. Snider finished far down the list of MVP finishers. It seems that Duke did not receive credit for leading Brooklyn to another pennant, even with his great season, as teammates Newcombe, Maglie, Gilliam, and Reese all finished ahead of Snider for the MVP.

Dick Stuart led the league in RBI and total bases for the seventh-place 1963 Red Sox. Teammates Radatz and Yastrzemski finished fifth and sixth for the MVP, so voters were willing to vote for players on a 7th place club. However, Stuart was not anybody's idea of an MVP so wound up far down the list. The model does not know that some players will simply never do well when it comes to the MVP.

Al Rosen led the league in RBI and total bases for the second-place 1952 Indians, the season before he won the unanimous MVP. Rosen finished 10th in the MVP, behind teammates Wynn, Lemon, and Garcia. Of course, any slight Rosen might have received in 1952 was more than made up for the next year.

Gus Zernial was another hard luck case. Zernial led the league in home runs and RBI for the 1951 Athletics (also playing 4 games for the White Sox). He finished far down the MVP voting with a share of .04, one of the all-time lowest shares for any home run or RBI leader. The Athletics finished in sixth place, but teammates Ferris Fain, who won the batting title, and Eddie Joost, both garnered decent votes and both fared better than Zernial.

Hank Aaron led the league in RBI and total bases for the second-place 1960 Milwaukee Braves. However, Aaron's season was seen as a disappointment following his stellar 1959 season. In addition, the Braves finished far behind the Pirates, following their previous three seasons in which they won 2 and a half pennants.

1998 MVPs

What does the model predict for the 1998 MVPs? Sosa (.83) easily over McGwire (.57) in the NL due largely to leading the league in RBI and making the playoffs. The phenomenon of the wild card really throws a wild card into the MVP deck as well. It seems that many voters are giving a lot of credit to winning the wild card, a sort of dumbing down of the voting. Since it is now considered "harder" to win the wild card (more teams are vying for it and it often goes down to the wire), the MVP may come from many wild card teams in the future.

In the AL, the model predicts Gonzalez (.79) over Belle (.68), Jeter (.65), and Williams (.61) due to leading the league in RBI.

Conclusions

In this article I have attempted to determine what drives MVP voting. By analyzing the complete and entire history of BBWAA MVP balloting from 1931-1998, I have confirmed the generally held view that leading the league in RBI gives a player a leg up on the MVP award. Playing on a league or division champion is by far the second most important thing a player can do according to MVP voters. Moderately important are leading the league in slugging percentage, runs scored, total bases, and batting average.

The model predicted both Sammy Sosa and Juan Gonzalez for the 1998 MVP awards, based largely upon leading the league in RBI. This crude model can shed light on both past and future MVP races. I hope to expand and/or refine the model as appropriate over time.

Rob Wood is a management consultant. He can be reached at 2101 California St. #224, Mountain View, CA, 94040-1686, rob.wood@us.pwcglobal.com. ♦

The Batter/Pitcher Matchup

Dan Levitt

If a .300 hitter faces a pitcher off whom the .260-hitting league hits .280, what is that hitter's expected average against that pitcher? The Log5 method, introduced in the 1980s by Bill James and Dallas Adams, attempts to answer that question. Here, the author tests the Log5 method, using real-life data, and finds that it stands up to the challenge.

One of the key questions facing baseball modelers revolves around determining of the outcome of the batter/pitcher matchup. When a .310 hitter faces a pitcher with a batting average against (OAV) of .290 what should the resulting batting average be? At first glance it may appear that the result should be the average of the two, i.e. .300. Upon reflection, however, this solution is flawed. Assuming a .260 league average, the .290 pitcher is worse than average. Therefore, the batter should hit for a higher average against this pitcher than his overall average.

Bill James in his *1983 Baseball Abstract* introduces the log5 method for addressing this calculation. He credits the formula below for evaluating the batter/pitcher matchup to Dallas Adams.

$$\frac{(BatAvgxPitAvg)/LgAvg}{(BatAvgxPitAvg)/LgAvg + ((1-BatAvg)x(1-PitAvg)/(1-LgAvg))}$$

Where PitAvg equals batting average against the pitcher. In the above example this formula would be:

$$\frac{(.310x.290)/.260}{(.310x.290)/.260 + ((1-.310)x(1-.290)/(1-.260))}$$

which evaluates to .343.

Does James' theoretical calculation hold true in actual play? The *Stats Baseball Scoreboard 1996* contains a study that allows us to empirically test the validity of the above formula. Stats, in an essay titled "Who Hits Who?", looked at how batters hit in 1995 against three pitcher classes: Good (top third), Average (middle third), and Poor (bottom third). Stats' sample included all batters with at least 446 PA and pitchers who faced at least 100 batters.

The appendix to the Stats essay presented the batter data for all 136 batters with at least 446 PA. By breaking the batters into the same three groups (top, middle, and bottom third), we can create nine different comparisons in each league. That is, how well good hitters do against good pitchers, average hitters against poor pitchers, etc.

As an example of the batter calculation methodology, below I have listed the top 24 AL hitters by batting average (i.e. the top third) including how they did against the three classes of pitchers as reported by Stats. In total there were 72 AL hitters with 446 or more plate appearances and 64 such hitters in the NL.

Top Third of AL Batters

Player	Team	AB	PA	Avg	VsGood	VsAve	VsPoor
Martinez, Edgar	Sea	511	639	.356	.285	.274	.308
Knoblauch, Chuck	Min	538	629	.333	.291	.382	.347
Salmon, Tim	Cal	537	638	.330	.310	.301	.390
Boggs, Wade	NYA	460	541	.324	.276	.360	.306
Murray, Eddie	Cle	436	480	.323	.301	.312	.336
Surhoff, B.J.	Mil	415	462	.320	.277	.341	.361
Davis, Chili	Cal	424	522	.318	.248	.365	.273

Belle, Albert	Cle	546	629	.317	.264	.332	.357
Baerga, Carlos	Cle	557	600	.314	.280	.307	.342
Puckett, Kirby	Min	538	602	.314	.264	.324	.371
Thome, Jim	Cle	452	557	.314	.261	.290	.403
Seitzer, Kevin	Mil	492	570	.311	.291	.309	.363
Joyner, Wally	KC	465	550	.310	.265	.302	.353
Lofton, Kenny	Cle	481	529	.310	.286	.235	.412
Palmeiro, Rafael	Bal	554	624	.310	.273	.310	.338
Ramirez, Manny	Cle	484	571	.308	.233	.340	.333
Thomas, Frank	ChA	493	647	.308	.296	.297	.321
Naehring, Tim	Bos	433	520	.307	.263	.374	.283
Williams, Bernie	NYA	563	648	.307	.266	.347	.293
Canseco, Jose	Bos	396	450	.306	.299	.308	.302
Johnson, Lance	ChA	607	645	.306	.247	.302	.380
Rodriguez, Ivan	Tex	492	517	.303	.271	.315	.331
Clark, Will	Tex	454	537	.302	.318	.300	.289
Alomar, Roberto	Tor	517	577	.300	.258	.288	.361
Sub-Total Good Batters				**.315**	**.276**	**.317**	**.340**

Note that the above subtotals are simple averages; they are not weighted by at bats.

In the next table below I calculated the opponents batting average for each of the three pitching categories.

OAV (Opponents Batting Average) Calculation			
206 AL Pitchers with 100+ BFP			
	AB	H	OAV
OAV Top Third (69)	22,101	5,146	.233
OAV Mid Third (68)	26,917	7,340	.273
OAV Bot Third (69)	18,054	5,533	.306
Overall for 206	67,072	18,019	.269

192 NL Pitchers with 100+ BFP			
	AB	H	OAV
OAV Top Third (64)	20,691	4,712	.228
OAV Mid Third (64)	26,225	6,885	.263
OAV Bot Third (64)	18,768	5,609	.299
Overall for 192	65,684	17,206	.262

For example, the above tables indicate that when batters with an overall average of .315 face good pitchers (overall average against of .233), they will hit approximately .276. This ties out almost perfectly with James' formula which predicts a batting average of .275 when a .315 hitter faces a .233 pitcher given a league average of .269.

In fact, as the final table indicates, the actual batting averages in the eighteen situations (nine comparisons in each league) correlate extremely well with the formula. In other words, at least by this 1995 data, James' formula for predicting the outcome of the batter versus pitcher matchup holds up extremely well.

Summary Table

Summary of Batter versus Pitcher Matchups

Vs Pitcher Type	Batters ⇨		All Actual	All Formula	Good Actual	Good Formula	Average Actual	Average Formula	Poor Actual	Poor Formula
AL	All with 100+ BFP	OAV	.286		.315		.289		.255	
	Good	.233	.250	.249	.276	.275	.251	.251	.223	.220
	Ave	.273	.290	.290	.317	.319	.292	.293	.262	.259
	Poor	.303	.317	.325	.340	.356	.323	.328	.287	.291
NL	All with 100+ BFP	OAV	.283		.312		.283		.253	
	Good	.228	.245	.247	.266	.274	.251	.247	.218	.219
	Ave	.263	.284	.283	.314	.313	.280	.283	.259	.253
	Poor	.299	.317	.321	.351	.353	.322	.321	.279	.289

Dan Levitt, 4401 Morningside Road, Minneapolis, MN, 55416, danrl@ibm.net. ♦

Book Reviews Wanted

Every year, a number of books and magazines are published with a Sabermetric slant. Many of our members have never heard of them. Our committee members would like very much to hear when this kind of stuff comes out.

If you own a copy of any baseball book of interest, we'd welcome a summary or a full-length review. Since we've hardly published for the last couple of years, even reviews of older books – say, 1997 or later – would be welcome. The only restriction, please: the book should have, or claim to have, some Sabermetric content.

Send reviews to the usual place (see "Submissions" elsewhere in this issue). Drop me a line if you want to make sure no other member is reviewing the same publication, although multiple reviews of the same book are welcome, particularly for major works.

As of the moment, I have a volunteer for the Bill James manager book. Any other book you want to review is yours, on a first-come-first-served basis. Let me know which book you're doing, so I don't assign the same book twice.

And if you're an author, and you'd like to offer a review copy, let me know – I'll find you a willing reviewer. This is especially appreciated if you've already plugged your book in the pages of BTN. (hint, hint. ☺)

Measuring Offensive Productivity by the Standings Value Method

David Shiner

How important is a specific offensive event – the stolen base, for instance -- to winning a pennant? In 1983, Bill James attempted to answer this question by looking at the standings position of teams that led the league in a specific stat, and comparing it to the finish of the team that trailed the league in that category. In this study, the author updates the Bill James study, decade by decade.

In the 1983 edition of the *Baseball Abstract*, Bill James constructed and carried out a study measuring the relationship between team totals in a variety of offensive categories -- doubles, triples, home runs, walks, stolen bases, and so on -- and position in the final standings. Examining the period 1969 through 1982, he demonstrated that stealing bases was far inferior to other offensive weapons when it came to winning ball games.

It is hardly a revelation that home runs are more valuable than stolen bases in today's game. According to James, "You cannot win a pennant by stealing bases. Nobody ever has, nobody ever will" (*1983 Baseball Abstract*, page 96). But is this statement necessarily true? Would it hold for the dead-ball era, for example?

To answer this question, I expanded James' study to include every major league season since the turn of the century. The method according to which the study was conducted (which, as James points out, is only one of many which could be devised for this purpose) is simple enough. It begins with the following steps:

1. Find the team that led its league (or division, since 1969) in a given category during a given year.

2. Record the final standing of that team. In case of ties, record the team that finished higher in the standings.

3. Compute the average of the final standings for all teams in each category over the selected period.

The results James found for the five previously-mentioned categories for the period 1969-82, which represented the entire history of divisional play at the time he was writing, are shown in chart #1.

Chart #1 -- Average Finish in Standings of Team Leading Division in:

Doubles	2.46
Triples	2.32
Home Runs	2.43
Walks	2.69
Stolen Bases	2.75

As you can see from the chart, the average team leading its division in triples during this period finished between second and third place in the standings, but closer to second. Teams that led their division in stolen bases also averaged a finish between second and third place, but much closer to third. The other offensive weapons -- doubles, homers, and walks -- fell somewhere in between. All in all, there was not much of a difference among the five categories.

The method continues as follows:

4. Find the team that finished last in its league (division) in a given category during a given year.

5. Record the final standing of that team. In case of ties, record the team that finished higher in the standings.

6. Compute the average of the final standings for all teams in each category over the selected period.

For the period 1969-82, James presented the figures in chart #2.

The range here is much greater than in Chart #1. Teams hitting the fewest triples in their division averaged a finish midway between third and fourth places in those years, while teams with the lowest doubles totals finished around fifth place on average. A relatively ow number of doubles was thus much more damaging to a team's final destiny than a low number of triples.

James then completed the study with the following step:

7. For each category, subtract the average finish in the standings of teams which were last from the average finish in the standings of teams which were first.

Chart #2 – 1969-1982: Average Finish in Standings of Team Trailing Division in:

Doubles	5.11
Triples	3.50
Home Runs	4.39
Walks	4.50
Stolen Bases	3.54

The purpose of this final step is, in James' words, to determine "the average distance between the teams finishing first in the league and those finishing last." The greater this "distance," he reasoned, the more valuable the offensive category. His findings for the years 1969-82 are in chart #3.

This chart shows the final results of the method for the period James studied. Stolen bases was the least efficacious category for this period, with the division leader finishing on average less than one place ahead of the tailender. Home runs were much more valuable: teams which led their division in homers finished on average almost two places in the standings ahead of teams finishing last in that category. We can therefore say that homers had a "standings value" of 1.96 for this period, while stolen bases had a "standings value" of only 0.79.

Chart #3 -- Average Distance Between Leader and Tailender in:

Doubles	(5.11 - 2.50)	2.68
Triples	(3.50 - 2.32)	1.18
Home Runs	(4.39 - 2.43)	1.96
Walks	(4.50 - 2.69)	1.81
Stolen Bases	(3.54 - 2.75)	0.79

In order to see whether the same results would obtain during other eras, I studied the data for the entire 20th century to date. The standings values of these categories for the first two decades of this century are shown in chart #4.

According to the standings-value approach, home runs were the least important major offensive category during the dead-ball era. This result is not surprising, since so few home runs were hit in those days. During the first decade of the century the average team hit 20 home runs per season, or less than one per week. During the teens home runs increased to about 26 per team per season, still a very low number by current standards. As the chart indicates, their standings value also improved in the second decade of the century, although not enough to surpass any of the other categories.

Chart #4 -- Standings Values During the Dead-Ball Era

Category	1900s	1910s	Overall
Doubles	3.94	2.63	3.24
Triples	2.74	2.64	2.69
Home Runs	1.68	2.32	2.02
Walks	3.21	3.04	3.12
Stolen Bases	2.69	2.32	2.49

Results begin with the 1900 National League season and the 1901 American League season. The 1914-1915 Federal League seasons are also included in the data for the ‘teens.

This is not to say that home runs were worthless in those days. As the chart demonstrates, the team leading the league in home runs finished an average of about two places ahead of the team which hit the fewest over this period, and the category value increased as homers became more frequent. Nevertheless, compared with other offensive weapons, home runs were of relatively minor importance.

Home run totals more than doubled from the teens to the twenties. Their efficacy according to this method, though, was surprisingly small; in fact, they had less impact on the standings than during the previous decade. Home run totals increased again, this time by more than a third, in the 30s, and their correlation with a high finish also improved significantly. Home run totals have been among the most important of the major offensive categories ever since. (See chart #5.)

Chart #5 -- Standings Values During the Babe Ruth Era

Category	1920s	1930s	Overall
Doubles	1.80	2.70	2.25
Triples	4.75	2.35	3.55
Home Runs	2.05	3.35	2.70
Walks	3.75	3.30	3.53
Stolen Bases	1.70	1.75	1.73

The totals for every other major offensive weapon except stolen bases also increased during this period. Stolen base totals were going in the opposite direction from home run totals: in the twenties there were about half as many stolen bases as there had been during the dead-ball era, and steals decreased by another third during the 30s. Their standings value also declined from the dead-ball era. Stolen bases had become the least important offensive category in this group, and with the exception of the 1950s it would remain so to the present day.

In the 40s (chart #6), the speed categories, represented here by steals and triples, were now in serious decline. The dominant offensive strategy during this period consisted of getting people on base and relying on the three-run homer. The number of walks increased about 13% from the totals of the thirties. Home run totals declined slightly in the forties, then increased by more than 60% in the fifties. Their contribution to a high place in the standings mirrored these trends, as the chart demonstrates.

Chart #6 -- Standings Values During the Ted Williams Era

Category	1940s	1950s	Overall
Doubles	4.25	1.65	2.86
Triples	3.10	0.61	1.77
Home Runs	3.20	3.91	3.58
Walks	4.10	2.48	3.23
Stolen Bases	1.15	2.17	1.70

For the sake of creating historically appropriate categories wherever possible, I have included the entire pre-expansion period in the "1950's" category. This means that 1960 is included for both leagues and 1961 for the National League.

Stolen base totals declined steadily during this period. Still, their standings value increased significantly in the 50s, a result due entirely to the great success of the fleet Dodgers in the NL and the more qualified success of the "Go-Go" White Sox in the AL. This recovery would be temporary.

Pitching was in ascendance during the 60s, dominating as it had not since the dead-ball era. Stolen bases were up more than a third from the previous decade, while totals for every other offensive category in our study declined somewhat. Nevertheless home-run-hitting teams continued to dominate the standings (see chart #7), while stolen bases had less of a positive effect than ever before. In general, expansion periods have produced a sharper distinction than normal between the "haves," who hit homers and win games, and the "have nots," who steal bases and lose. Last year (1998), for example, stolen bases did not correlate positively with wins at all.

Chart #7 -- Standings Values After Expansion and Before Divisional Play

Category	1960s
Doubles	2.20
Triples	1.27
Home Runs	3.67
Walks	1.53
Stolen Bases	0.13

Again for reasons of historical appropriateness, this era begins with the 1961 American League season and the 1962 National League season. It concludes with the 1968 season for both leagues, just before the introduction of divisional play and rule changes favoring offense.

Reintroducing James' figures from the 1969-82 period and adding more recent totals gives chart #8.

Despite the general increase in offensive production, we see the same basic trends as in the 60's. Home runs have had considerably lower standings value than during the previous "era," but given the fact that each division consists of fewer teams than ever before the overall value of team homers is probably as great as it's ever been. Stolen base totals per team were higher in the 80s and early 90s than at any time since the dead-ball era, but with minimal effect on their standings value. With stolen bases in decline over the past few years, their standings value has declined as well.

Chart #8 -- Standings Values During the Period of divisional Play

Category	1969-82	1983-98	Overall
Doubles	2.68	1.24	1.86
Triples	1.18	0.67	0.89
Home Runs	1.96	1.91	1.93
Walks	1.81	2.22	2.04
Stolen Bases	0.79	0.40	0.56

We began by asking whether Bill James was right in claiming that stolen bases can never compete with home runs. It turns out that while he was not literally correct, he was correct in the sense he intended -- that is, with respect to the conditions of major-league baseball as it is currently played. In the early days of the century, when home runs were rare, stealing bases was a relatively valuable offensive stratagem, better than counting on home runs. This has never been the case since that time. Given trends in baseball since Babe Ruth's "rise to power," it's fair to say that stolen bases are no threat to replace home-run hitting as the dominant offensive weapon in the major leagues any time soon.

A much more extensive and detailed version of this paper is available from the author. David Shiner, 706 Washington, Knollwood, IL, 60044, cunegonde@prodigy.net. ♦

Stolen Base Strategies Revisited

Tom Ruane

How often must a runner successfully steal second base to make the risk worthwhile? Pete Palmer studied this issue many years ago, but did not take into account variables like quality of the subsequent hitter, or speed of the base thief. Here, via a decade's worth of real-life play-by-play data from Retrosheet, the author addresses some of those issues.

In *The Hidden Game of Baseball*, Pete Palmer determined the expected future runs for each of the 24 game situations (with outs going from 0 to 2 and bases ranging from empty to full) and used that to evaluate, among other things, break-even points for stolen base attempts. Pete used the results of computer-simulated baseball games to develop his charts. Thanks to the folks at Retrosheet and Project Scoresheet, I've been able to use actual play-by-play data to generate similar information for National Leagues games from 1980 to 1989:

MenOn	Number of Outs			MenOn	Number of Outs		
FST	0	1	2	FST	0	1	2
---	.452	.239	.091	---	.261	.148	.061
x--	.815	.486	.210	x--	.424	.267	.124
-x-	1.051	.653	.313	-x-	.608	.400	.216
xx-	1.384	.852	.404	xx-	.617	.413	.219
--x	1.278	.912	.358	--x	.819	.651	.269
x-x	1.638	1.131	.471	x-x	.843	.646	.275
-xx	1.884	1.313	.576	-xx	.841	.667	.274
xxx	2.176	1.481	.718	xxx	.855	.664	.316

The figures on the left are the average number of runs scored during the rest of the inning starting from the given situation. For example, in innings where the team at bat has a man on second with no one out, they add an average of 1.051 runs before making the third out. The figures on the right contain the percentage of times the team will score at least one run.

To paraphrase part of Pete's argument, since an attempt to steal second with no outs risks a loss of .576 runs (.815 - .239) for a gain of .236 runs (1.051 - .815), it must be successful more than 70.9% of the time to increase a team's expected runs. His formula for determining this was:

```
             ( start EFR - fail EFR )
------------------------------------------------------
( start EFR - fail EFR ) + ( success EFR - start EFR )
```

or more simply:

```
( start EFR - fail EFR ) / ( success EFR - fail EFR )
```

Palmer's study was more complicated than I've suggested here, using tables like these in conjunction with others containing win probabilities for various game situations (bottom of the seventh, home team down by a run), and like much of that book, it represented a major advancement in the analysis of the game. But while it has been used as a foundation for several other studies (Gary R. Skoog, for example, based his article "Measuring Runs Created: The Value Added Approach" in the 1987 *Bill James Baseball Abstract* on this work), it has also come in for its share of criticism. One of the major criticisms is that it ignores the varying abilities of the players involved. How different would the chart above be if the hitters due up were Ruth and Gehrig? Or Rey Ordonez and Bobby Jones? What if Rickey Henderson were the lead runner? Or Ernie Lombardi? Another criticism is that Palmer's study treats the stolen base as a binary event, when in fact, it has several possible outcomes.

Despite the general title, in this article we'll be looking at the wisdom of attempting to steal second with only first base occupied. The main reason for this is that it's far and away the most common base-running situation. In 1987, for example, it accounted for 77.7% of all stolen base activity. An attempted steal of second base with men on first and third was a distant runner-up with 9.1%. But while I'll be limiting my discussion to one class of stolen base, I will be including the data that will allow the reader to apply the same method to the other scenarios.

Here are the break-even points for a steal of second using Pete Palmer's formula:

Number of outs:	0	1	2
Break-even Pct:	70.9	70.4	67.1

In general, Palmer tended to down-play the dangers of an aggressive running game, which is ironic in light what he said in his book:

"The stolen base... is an overrated play, with even the best base stealers contributing few extra runs or wins to their teams. The reason for this is that the break-even point is so high, roughly two steals in three attempts."

One of the problems with this kind of analysis is that there's already some amount of stealing and getting caught already embedded in the data, perhaps skewing the results. This is perhaps more easily seen by example. I arrived at the 70.9 in the table above by plugging the following numbers into the break-even formula:

```
( .815 - .239 ) / ( 1.051 - .239 )
```

But what if every time a runner was on first in this situation, he successfully stole second? In that case, the expected runs from the starting state (man on first, no one out) would equal the expected runs from the "success" state (man on second, no one out), since whenever we hit the first we made a successful transition to the second. This would've modified the above formula to:

```
( 1.051 - .239 ) / ( 1.051 - .239 )
```

making it appear as if there were nothing to be gained by attempting to steal second. Conversely, if the runner took off and was caught each time, the formula would've been transformed into:

```
( .239 - .239 ) / ( 1.051 - .239 )
```

making it appear as if there were no risk involved. One way to attempt to correct for this is to remove the type of plays being evaluated from the charts. How is this done? I generated the original charts by examining every play, determining the starting situation (outs and base-runners) and calculating the number of runs scored from that play until the end of the inning. To remove stolen base events (and here I'm including stolen bases, caught stealing, pick-offs, errors attempting pick-offs, and balks), I simply ignore any of these plays when generating the charts. Consider the following inning:

Outs	MenOn	Runs	Play
0	---	2	strikeout
1	---	2	walk
1	F--	2	steal of second
1	-S-	2	walk
1	FS-	2	caught stealing third, trailing running to second
2	-S-	2	two-run home run
2	---	0	strikeout

When determining the expected future runs at the start of this section, I included the data for all of these plays. To remove stolen base events from the chart, I would not factor in the two running plays above. Note that this doesn't prevent these plays from affecting the data. For example, in the inning above I would record that two runs were scored from the no one on/none out situation. But clearly that result was affected by both the steal of second (which may have caused the subsequent walk) and the caught stealing.

Here's what the adjusted chart looks like:

Outcomes when a steal was not attempted on the next play

MenOn FST	Number of Outs 0	1	2	MenOn FST	Number of Outs 0	1	2
---	.452	.239	.091	---	.261	.148	.061
x--	.816	.489	.206	x--	.415	.262	.118
-x-	1.052	.650	.313	-x-	.608	.397	.216
xx-	1.395	.855	.404	xx-	.618	.414	.219
--x	1.281	.922	.357	--x	.821	.658	.267
x-x	1.643	1.142	.464	x-x	.848	.658	.271
-xx	1.886	1.315	.574	-xx	.841	.668	.271
xxx	2.177	1.484	.716	xxx	.856	.666	.314

The before and after break-even points:

	-- With Running --			--- Without ----		
Number of outs:	0	1	2	0	1	2
Break-even Pct:	70.9	70.4	67.1	71.0	71.1	65.8

So those earlier percentages were slightly lower with none and one out and a little too high with two outs. But it doesn't appear as if the running plays affected the data much at all.

Given these break-even points, has all that base-running helped or hindered the team at bat? One way to look at this is to generate charts using only those events removed earlier and then compare the two sets.

Outcomes when a steal was attempted on the next play

MenOn FST	Number of Outs 0	1	2	MenOn FST	Number of Outs 0	1	2
x--	.811	.475	.237	x--	.468	.293	.161
-x-	1.031	.723	.330	-x-	.628	.481	.239
xx-	1.114	.776	.437	xx-	.589	.404	.230
--x	.533	.340	.588	--x	.467	.284	.559
x-x	1.584	1.033	.548	x-x	.794	.546	.322
-xx	1.444	.929	.911	-xx	.889	.500	.667
xxx	2.000	1.042	1.222	xxx	.727	.458	.815

The differences with a man on first:

Stealing - Without Stealing

Number of Outs 0	1	2	Number of Outs 0	1	2
-.005	-.014	+.031	+.053	+.031	+.043

The -.014 in the second column means that teams scored an average of .014 fewer runs when, with a man on first and one out, the runner attempted to steal second than they did when the man on first didn't run. These differences aren't much, but they do seem to indicate that running is a much better strategy either with two outs or as part of a one-run strategy than it is otherwise.

But how does the speed of the lead runner change this? No one disputes that Will Clark (who was caught on 17 of his 22 attempts in 1987) should have run less than he did, but what about Tim Raines and Vince Coleman?

Using the Speed Score statistic devised by Bill James in his 1987 Baseball Abstract, I grouped the lead runners into 3 categories, using scores of 3.0 and 5.5 as the dividing lines*. When I initially looked at this, I found something very surprising: the speed of the lead runner was almost as important as the ability of the subsequent batters in determining the number of runs scored during the rest of the inning. And that brings us to the second trap inherent in this type of study: always be sure to isolate what you want to examine. What I had been doing originally was compounding the effects of skilled batters *and* fast runners. In other words, when fast runners are on base (for example, a typical lead-off hitter), the batters coming up are generally better than those at the plate when slow runners are aboard. So while I thought I was examining the speed of the lead runner, I was also letting the effects of a better class of hitter leach into my study. So I

* When runners were on first and third, the speed factor of the man on first was used.

changed the test to ignore all cases where the weighted OPS (on-base plus slugging percentage) of the batters coming to the plate was less than .700 or greater than .800. This cut the sample size roughly in half, but allowed me to more closely examine the influence of speed on an offense.*

The results with a man on first base:

Speed Score	Number of Outs 0	1	2	Number of Outs 0	1	2	Avg OPS
< 3.0	.845	.484	.213	.409	.242	.120	.742
> 3.0 AND < 5.5	.837	.522	.231	.427	.283	.135	.746
> 5.5	.962	.590	.267	.497	.328	.163	.751

Despite what I said earlier about the effects of the running game, there seems to be a substantial benefit to having speed on the base-paths. The number of expected runs increased by an average of 14% from the slow to the fast group and the chance of scoring a single run rose by 22%. Part of this is due to slightly better hitters coming to the plate. Even though we eliminated all situations where the weighted OPS of the hitters due up was less than 700 or over 800, the hitters up with fast lead runners were still slightly better.

But how much of this is due to stolen bases? What net gain or loss do we see from base-running events when we look at situations involving the fastest runners?

	Stealing - Without Stealing: Number of Outs 0	1	2	Number of Outs 0	1	2
Fastest Runners	-.014	-.045	+.030	+.060	+.018	+.047
All Runners	-.005	-.014	+.031	+.053	+.031	+.043

In this light, attempting to steal seems an even more costly strategy with the fastest runners in the lead. But we just showed that teams score more runs with speed on the base-paths -- if the fast runners aren't helping their teams by stealing bases, where are those extra runs coming from? Here's what happens with a batter hits a single with a runner on first and no outs:

No Outs	Bat->1st Run->2nd	Bat->1st Run->3rd	Bat->2nd Run->3rd	Bat->1st Run->Out	Bat->2nd Run->Out	Other
Slow	76.9	18.3	3.0	.4	.4	1.0
Average	66.3	28.4	2.7	.9	.7	.9
Fast	55.8	39.1	2.1	.6	.6	1.7

This is a significant improvement, especially in light of the low risk involved. The results with one and two outs are very similar.

When there's a groundout with no out, here's what happens:

	GDP	Force at 2nd	Play at 1st	Other
Slow	43.5	39.4	16.7	.5
Average	43.9	32.6	22.5	.9
Fast	36.0	31.2	31.3	1.5

With one out:

	GDP	Force at 2nd	Play at 1st	Other
Slow	45.2	33.6	20.5	.7
Average	45.3	28.4	25.5	.7
Fast	42.3	27.4	29.4	.8

Again, another improvement without the extra risk of losing a baserunner. By the way, one thing I did not find with a fast runner on first was an increased number of wild pitches, passed balls or errors.

* This same problem could've skewed my previous comparison between the expected runs of stolen base events and all others. But since fast runners tend to be on ahead of better hitters, this skewing would make the conclusions even more unfavorable to the running game.

But does speed on the basepaths help the batter as well? You often hear that pitchers are distracted by the runner or will throw an inordinate proportion of fastballs, improving the performance of the batter. Is there any truth to this? Listed below are the OPSs of hitters with a man on first along with their overall OPS:*

	0 out		1 out		2 outs	
Runner Speed	First	Overall	First	Overall	First	Overall
Slow	.765	.745	.768	.745	.727	.746
Average	.795	.746	.787	.745	.738	.746
Fast	.824	.747	.811	.749	.787	.749
Total	.796	.746	.789	.747	.729	.747

So there does seem to be something tangible to this rumored benefit after all.

Finally, what should the effect of the batters' ability be on stolen base strategies? To determine this, I took a weighted average of the OPS of the next three hitters and separated them into three categories, using .700 and .800 as the dividing lines.# Since I wanted to examine stolen bases, I removed them from the study (using the method described earlier), and computed the break-even points with each class of hitter.

	Number of Outs			Number of Outs		
	0	1	2	0	1	2
OPS < .700	.720	.753	.627	.569	.626	.492
.700 < OPS < .800	.762	.744	.709	.604	.621	.584
OPS > .800	.792	.793	.710	.645	.661	.598

In other words, the better the hitter at the plate, the less you should run.

In The Hidden Game of Baseball, John Thorn and Pete Palmer entitled the chapter that discussed these strategies "The Book... and the Computer." Back in 1985 when their book was first published, the computer was primarily used for generating thousands of simulated games and analyzing the results. This article has been an attempt to extend their research using a computer and ten seasons of actual play-by-play data. There are several potential traps in using the methods I've employed above. Some I've pointed out to the reader and avoided; others, no doubt, I've stumbled into unwittingly. It was not my hope to have the last word on this area of baseball strategy. Rather, I've attempted to move the discussion forward and to suggest new ways to take advantage of the enormous amount of information now available to researchers in this field.

Tom Ruane, 92 College Ave., Poughkeepsie, NY, 12603, truane@bestweb.net. ♦

* Once again, hitters with OPSs less than .700 or greater than .800 are removed from the chart.

The weighted OPS average was computed as follows, where OPS1, OPS2 and OPS3 represent the OPS of the man up, on deck and the man following him, respectively:

0 out : (OPS1 + OPS2 + OPS3) / 3
1 out : (OPS1 + OPS2 + (OPS3 * .67)) / 2.67
2 out : (OPS1 + (OPS2 * .33)) / 1.33

A New Way of Platooning – Ground Ball/Fly Ball

Thomas A. Hanrahan

The platoon effect for left- or right- swinging batters vs. lefty or righty pitchers is well-established. But there is a similar effect, the author shows, for batters' and pitchers' fly ball or ground ball tendencies, and it is almost as significant.

Over the past century, platooning hitters has become an accepted practice. Experience has shown that right-handed batters hit better against left-handed pitchers, and vice-versa. In fact, when you saw the word "platooning" you likely immediately thought of right- vs. left-handed. But is the handedness of the batter and pitcher the only way to maximize performance, or even the best? What other criteria might be important?

I researched one potential factor: how do batters who hit mostly ground balls (or flies) fare against pitchers who allow mostly flies (or grounders)? As I will show, there is a definite platoon effect at work, and the results are significant. Specifically:

- Ground ball hitters do better against fly ball pitchers;
- Fly ball hitters do better against ground ball pitchers;
- The effects are larger for batters who strike out often;
- The effects are larger than the standard left/right platoon difference in a few cases.

Data

I used the book Player Profiles (by STATS, Inc.) which gave hitting breakouts for the years 1992-96. Each pitcher has a statistic called GB/FLY ratio. STATS defined fly ball pitchers (FLY-P) as those whose GB/FLY ratio (ground balls allowed to fly balls allowed) was greater than 1.50. Ground ball pitchers (GB-P) have GB/FLY ratios of less than 1.00. Anyone in between is classified as neutral, and not used in this study. The MLB average GB/FLY ratio was 1.30, and about 60% of all pitchers are "neutral".

Every hitter has a breakout of his performance against FLY-P and GB-P for these 5 years. I chose batters who had at least 300 plate appearances against both types of pitchers for this study. This gave me a set of 161 batters. Their total numbers are given in table 1.

Table 1

	AB	H	AVG	BB	KO	HR%	OBA[A]	SLG	KO%	RC/G[B]
vs. GB-P	93113	26427	.284	9687	14529	2.5	.351	.431	14.1	5.49
vs. FLY-P	67857	18412	.271	7584	11849	3.4	.345	.450	15.7	5.53

[A]OBA here is (hits+walks) / (at bats+walks), which is slightly different than the official statistic that includes things like HBP and SF.
[B]RC/G is runs created per game, defined by OBA * SLG / (1-AVG) * 26, per Bill James

FLY-P allowed more home runs but fewer singles, and walked and struck out a few more. Overall effectiveness was about even. None of these findings surprised me.

I measured an individual batter's performance by calculating RC/G and took the difference of RC/G vs. FLY-P minus RC/G vs. GB-P. When I plotted this FLY/GB DIFF against the batters' percent of time they hit ground balls, I found a definite correlation; the more a batter tended to hit mostly flies, the better he was likely to hit against a GB-P. The data was pretty widely scattered, but much of this is just natural variation dealing with the sample sizes (300 plus PA) for each batter. A stronger correlation was seen when I ran a regression using the batter's KO%. Batters who struck out often have a more pronounced effect than those who didn't whiff much. A plot of this correlation is shown in chart 1.

A linear regression was run just for the batters who struck out more than the average (KO%>15.0). The regression equation is

```
FLY/GB DIFF = -6.08 + (11.024) * (GB%)
```

This means that a batter who hit groundballs only 40% of the time (a true FLY hitter) would be expected to be –6.08+11.024*.40 = 1.58 RC/G better against a GB-P than a FLY-P. A batter who hit grounders 65% of the time (a real slap hitter) would be 1.08 RC/G *worse* against a GB-P. Again, this is only for batters who strike out often.

Mark McGwire, the most excellent current home run hitter, is the most extreme fly / high KO hitter in the study. He hit GB only 30% of his balls in play, by far the lowest of all 161 batters (the next lowest was 36%). He also struck out in 21% of his plate appearances. He in fact hit convincingly better against GB-P; a slugging of .721 (“only” .624 vs. FLY-P), and an OBA of .468 as opposed to .392. As Mel Allen would have said, “how aBOUT that!”.

Table 2 shows the batters in the study who had the largest difference in RC/G (GB-P vs. FLY-P). The top 4 who hit GB-P better are all fly hitting / high KO / big power guys. The 5 who hit FLY-P better are a mixed group; it surprised me that some real power hitters also hit lots of ground balls.

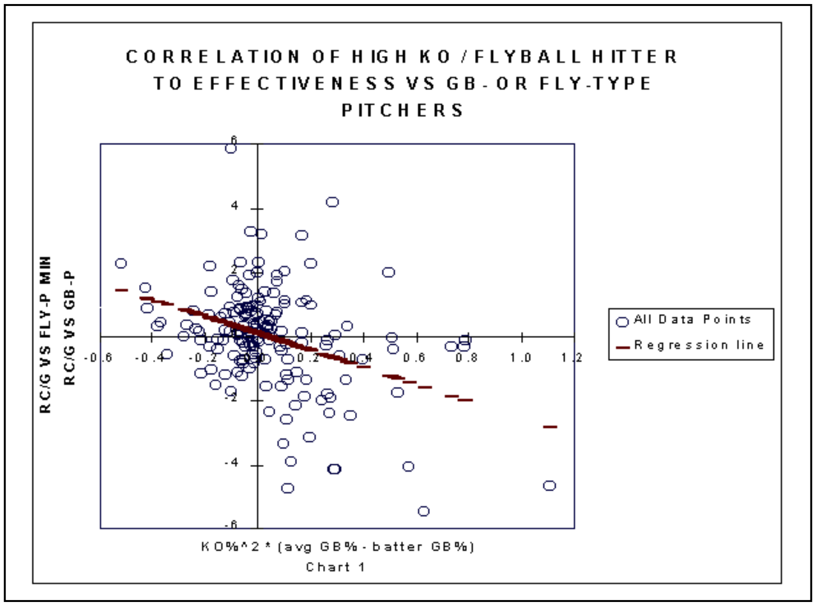


Chart 1

BATTER	vs. GB-P OBA	SLG	vs. FLY-P OBA	SLG	GB%	KO%	RC/G DIFF FLY-GB
C Hoiles	420	613	341	395	38	19.3	-5.48
G Hill	381	578	275	424	53	22.4	-4.74
M McGwire	468	721	392	624	30	20.7	-4.65
J Conine	396	498	300	361	48	20.3	-4.16
R Palmeiro	355	499	411	620	44	12.0	3.17
E Martinez	432	539	447	671	54	12.1	3.17
E Young	318	330	404	446	60	7.5	3.28
P Kelly	241	271	351	495	46	17.6	4.16
M Piazza	356	438	424	681	59	16.4	5.84

After I found that the correlation was stronger for high KO batters, I constructed a regression involving the 2 variables: GB% and KO%. The previous regression had only used GB% and merely eliminated the low KO% batters from the database. By far I saw the results most clearly when I combined the 2 variables into one parameter. I believe this was so because the two were related, in that many high KO% hitters are also fly hitters. What I did was use the deviation from the average GB% which was 55.3 (for the batters in this study; the actual major league average GB% was 56.5, but I think that you would get very similar results using either number) to define how far a hitter fell on the GB or FLY end of the spectrum. I multiplied this by $KO\%^2$ (that’s strikeout percentage squared) to account for how often the batter might swing and miss. By putting these together, I attempted to define how much any given batter was swinging over or under the ball. I actually used KO% first, but found the fit was much better using $KO\%^2$. So, the best equation came out to be

$$\text{FLY/GB DIFF (RC/G)} = .078 - 264 * (.553 - \text{GB\%}) * (\text{KO\%})^2.$$

This predicts that McGwire would hit $.078 - 264 * (.553 - .300) * (.207)^2 = -2.79$ RC/G worse against FLY-P. This is significantly <u>greater</u> than the expected difference when facing a right- vs. left-handed pitcher, which would be about 1.2 RC/G* (about 30 points in batting average, plus a difference in power stats). So, if you’re trying to decide to bring either your left-handed-flame-throwing-fly-ball pitcher or your right-handed sinkerballer in to face big Mac, I say go for the southpaw without a moment’s thought. Of course, Mark is the extreme case.

* Pete Palmer in The Hidden Game, p165, found that batters had an OBA+SLG of .779 when hitting with the platoon advantage, and .694 without, for a difference of 85 points, using data from 1974-77. Bill James came up with a difference of 87 points using the same methods for batters in the years 1984-87 in his 1988 Abstract. These would both equate to about 1.2 runs created per game.

Chart 2 plots the actual FLY-GB difference along with the regression formula line. You can see that the 9 batters (represented by circles) furthest to the right (which are fly ball high KO% batters) *all* hit better against GB-P, although some were very close to being even. McGwire is the circle waaaaay off to the bottom right.

You might say that there is a lot of "scatter" in the data (the r-squared value was .13, which isn't high), but if I had plotted a chart showing left/right platoon differences for 161 batters, you would see much scatter there also; it's the nature of the beast when working with 500 plate appearances or so per hitter. The student t statistic for this regression, a measure of whether this correlation might have occurred just due to plain old luck, is 5.08, putting the odds of this happening by chance at 1 in a million! I would say there is a definite relationship.

Interpretation

Let's step away for a minute from the data and ask *why*? I had not expected these results, because I had always heard announcers talk about how you have to keep the ball down to a big slugger, and how a pitcher ought to be aware of throwing a high fastball or a hanging curve to a big fly hitter, lest he deposit it in the bleachers. So, what's going on here?

My explanation: obviously, every batter tries to hit the ball squarely. However, power hitters tend to be fly ball hitters, because they can be rewarded for hitting under the ball if it goes 400 ft. Conversely, slap hitters naturally ought to err on the side of being over the ball rather than hitting a meek fly. So, when a fly hitter swings and misses, it's often because he is UNDER the ball (other times, of course, he is ahead or behind, or takes a called strike). Counteracting this is the GB-P, who throws some kind of sinking pitch that the batter tops or misses So, a GB-P sinkerballer meets a fly hitter, and presto, he lines it squarely for a base hit. In order for a fly hitter, especially one who strikes out often, to hit a FLY-P with a rising fastball, he would have to adjust and say to himself, "you know, fans normally cheer for me when I get under the ball and put it in the upper deck, but this guy makes a living blowing the ball by (or over) people's bats. Maybe I had better aim for the top of the ball against him". The results of this study indicate that this mental conversation does not take place. Fly hitters do poorly against FLY-P.

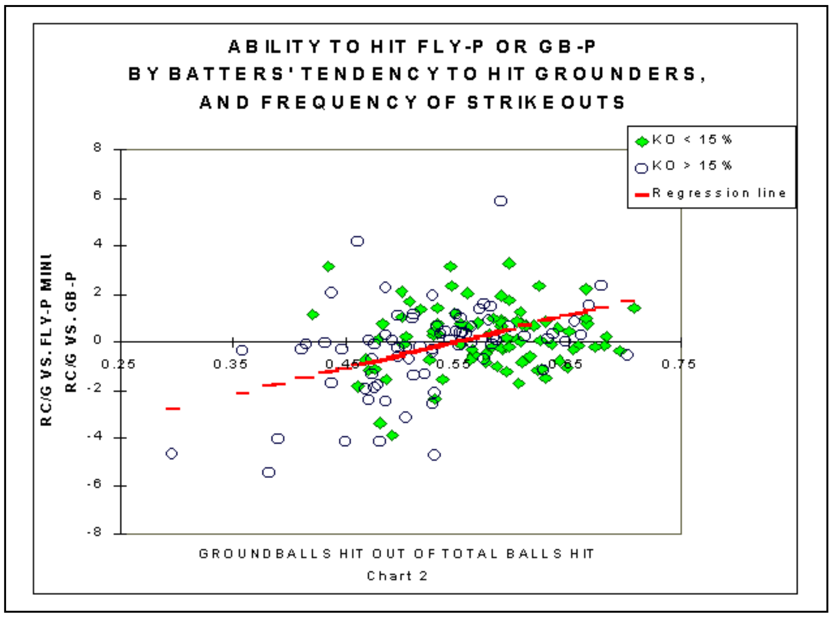


Chart 2

Other Investigations

I tried to find other factors that may have been important, but could not. No correlation was found when comparing power hitters versus singles hitters, or hitters who walk often or seldom.

Qualifiers

While I think these findings are very significant, I should take care not to overstate their magnitude. Many pitchers are neither FLY-P nor GB-P; they are neutral when it comes to inducing grounders or flies. So, even though the effects of platooning some batters vs. GB-P and FLY-P are substantially as great as platooning vs. LHP and RHP, they only apply a portion of the time.

This study was done without looking at extremes among pitchers. The Player Profiles book did not have the data available from the pitchers' point of view. It may be that the effects are even greater for the extreme sinkerballer; or, maybe not. A further study could be made using individual batter vs. pitcher matchups to test for this.

Practical Conclusions

How best to use this information? I would suggest:

1. Choosing a pinch hitter against a particular pitcher.
2. Choosing a good day to give your utility infielder/outfielder or backup catcher a start.
3. Choosing a reliever in a short (one batter or two) situation, especially when your bullpen may not have any lefthanders available.
4. Choosing what day to sit down your high KO/fly hitter. If McGwire ever needed a day off, I would suggest sitting him against a RH fastballer like Kerry Wood (I haven't seen any data on him, but I have to guess he gets his share of pop-ups!).

In Closing

Earl Weaver seemed to me to pioneer the practice of using batters based on their success against a particular pitcher. This type of data is now widely quoted ("…he is 6 for 23 lifetime against Clemens..."), although most statisticians wold laugh at the ridiculously low sample sizes that are used. However, analyzing what *types* of pitcher a batter does well against would seem to yield more reliable results. Batters are platooned RH vs. LH because a whole class of batters is known to well against a class of pitchers. This study shows that a different class of batters hits better against a different class of pitchers. The data do not show that the effects in most cases are as important as LHP / RHP platooning. However, the effects can be large for many batter / pitcher matchups. Large enough that a MLB manager ought to know about them and how to use them. Call up your favorite team and let them know.

Thomas A. Hanrahan, 21700 Galatea St., Lexington Park, MD, 20653, HanrahanTJ@navair.navy.mil. ♦

SABR-L Update

Compiled by Clifford Blau

A summary of the sabermetric research and discussion posted recently on SABR-L, the internet mailing list.

In April, 1998, **Tom Ruane** looked at how often a runner on second will advance to third on a groundball, dividing runners into slow, medium, and fast using Bill James' Speed Score. The results, based on three seasons of data are shown in the accompanying chart.

Slow Runner:

	No Outs--Ball Hit To						One Out--Ball Hit To					
	1	2	3	4	5	6	1	2	3	4	5	6
Holds	28	2	2	7	37	49	35	4	7	24	99	110
Advances	5	0	42	79	4	28	11	1	73	141	6	67
Lead Out	0	0	0	0	0	0	0	0	0	0	0	1
DP	0	0	0	0	1	0	0	0	0	0	1	0
Scores	0	0	0	0	0	0	1	0	0	0	0	0

Med. Runner:

	No Outs--Ball Hit To						One Out--Ball Hit To					
	1	2	3	4	5	6	1	2	3	4	5	6
Holds	60	9	12	7	103	141	76	2	28	56	294	239
Advances	32	1	131	279	18	77	63	11	203	353	36	179
Lead Out	1	0	0	0	0	0	0	0	0	0	0	1
DP	0	0	0	1	1	0	2	0	0	1	2	1
Scores	0	0	0	0	0	0	0	0	0	0	0	0

Fast Runner:

	No Outs--Ball Hit To						One Out--Ball Hit To					
	1	2	3	4	5	6	1	2	3	4	5	6
Holds	40	2	5	10	59	64	54	1	15	32	182	137
Advances	23	1	137	235	17	63	41	0	131	311	31	131
Lead Out	1	0	0	0	0	1	0	0	0	0	0	0
DP	1	0	0	0	1	1	0	0	0	1	2	1
Scores	0	0	0	1	1	0	0	0	1	1	0	0

In March, **Tom Ruane** conducted a significant study in which he examined the relative cost of strikeouts, groundouts, and flyouts. He found that the most beneficial type of out was the groundout; the extra doubleplays which result are outweighed by the opportunity to advance runners and the additional errors. (In this study, he included times reached on error as outs by the batter.) In comparing two hypothetical hitters, one of whom is an extreme fly ball hitter and the other an extreme ground ball hitter, but otherwise equal, he concluded that the ground ball hitter would produce between two and five more runs per season. This despite the fact that most models we use to evaluate an individual's offense would favor the fly ball hitter, due to the additional GIDP for the other. Also, the relative cost of a fly ball versus a grounder is dependent on the speed of the batter. There is little or no difference for a slow runner, while for a fast runner the difference is almost a quarter of a run.

The last issue of BTN mentioned **Dan Levitt's** finding that there was no more variability in batting averages in expansion years than in nonexpansion years. In late March and early April, **Tom Ruane** and **Mike Emeigh** also reported on similar studies. Mr. Ruane looked at all expansion years from 1961 to 1993. While some of those seasons showed an increase in variability in batting average, others did not. Even for the years that did, in all but one instance there was a higher variability in that league within two years. He concluded that there was no evidence that expansion causes an increase in variability.

Mr. Emeigh examined batting, slugging, and on base averages for the 1961, 1962, and 1977 expansions. For the two American League years, there were small increases in variability as well as in offense, while for the National League season, there were decreases in both. He speculated that the changes were due to the new ballparks in use rather than to expansion.

Moving into May, **Tom Ruane** looked at some influences on infielders' range factors. The first was the number of chances they have on balls they do not field. Using 1981 NL statistics, he found that anywhere from 28 to 40% of second basemen's chances came on balls they did not field. For shortstops, it ranged from 23 to 32%. He next looked at these figures for entire infields, together with data on the number of groundballs hit against each team and the location of the batted balls. One finding was that a fielder's range factor components are not affected by those of the other infielders on this team. Also, the corner infielders show a much larger difference in the number of balls they field depending on the average percentage of hits to the left or right side than do the middle infielders. The number of ground outs per game varied by 2.6. He showed this variability can last for years; thus, opportunities for making plays do not even out over time. In a related note, **Rich Hansen** found that in the 1967 AL, 16 % of second basemen's assists and 60% of their putouts came on balls they did not field.

Also in May, **Bob Allen** noted that Total Baseball's Fielding Wins is not the sum of its players' Fielding Runs. Rather, it seems to be directly related to team errors.

Tom Ruane showed us that average plate appearances per game by pitchers has dropped pretty steadily from about 3.3 in 1940 to fewer than 2.5 now. He also presented a chart of the best hitting pitching staffs, which was headed by the 1915 Red Sox.

Beginning in May, and moving into June, **Tom Ruane** used play-by-play data and charts which show expected runs from each base/out situation to look at the value of stolen bases. He presented several charts comparing these values to Stolen Base Runs (SBR) as calculated in Total Baseball. He found that due to the differing situations when players steal, and including the effect of errors and pickoffs, some players with high SBR's actually cost their teams runs, such as Brady Anderson in 1992. On the other hand, several players who were caught stealing five times in eight attempts actually benefited their teams. Over the 13 years ending 1993, Tim Raines was actually a more successful basestealer than Rickey Henderson. Mr. Ruane also applied this method to batting statistics, finding some large differences with Linear Weights. This demonstrates that caution is needed in using conventional sabermetric tools to evaluate individual player statistics. In a separate post, he noted that slow runners advance on wild pitches and passed balls as often as fast runners do. He followed up by applying the same method previously used for base stealers and hitters to all pitchers, and to relief pitchers specifically.

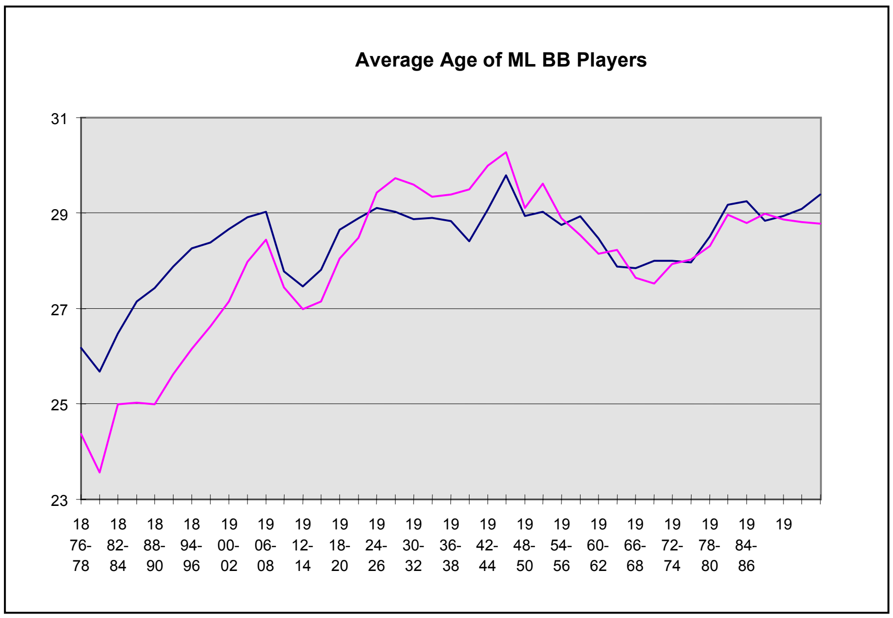
In June, **Tom Ruane** supplied some charts showing how hitters perform in different base/out situations, as well as the average overall performance of hitters who bat in each situation (as opposed to how they do in just that situation). Also, how often each position in the batting order hits in each situation, with the latter summarized by league. While the charts are too voluminous to reproduce here, there are a few interesting conclusions. Hitters do worst with the bases loaded, and the difference is not due to weaker hitters being up in that situation. Errors occur least frequently with the bases empty, and with two outs.

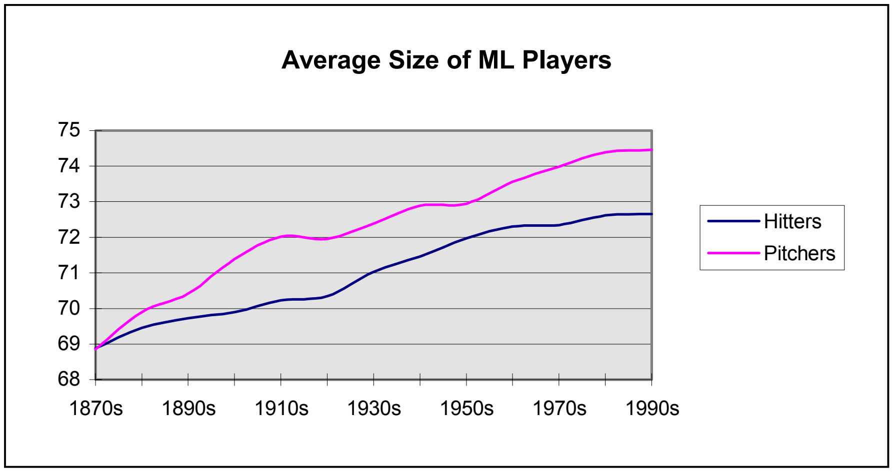
Moving into July, the ubiquitous **Tom Ruane** turned his attention to the average age of players throughout major league history. The graph below summarizes the data. He also prepared a chart that gave the percentage of players in different age groups over the years.

He also looked at the ages of teams, listing the youngest and oldest teams both on an absolute scale and relative to their leagues. In addition, he considered whether young teams tend to improve while old teams tend to decline. While the youngest teams did show remarkable improvement over the next few years, increasing their winning averages by .059 over 4 years, the oldest teams declined somewhat less, .040 over 4 years. In another article, he listed the average size of players over the years, as seen in the following graph:

Along with this, in a related series of posts, he compared the hitting and pitching results of various sized players. He found that the tallest group tends to be the best hitters, mainly due to higher slugging averages. On the other hand, tall pitchers showed no tendency to be more successful than others.

In August, moonlighting from his SABR-L duties, **Tom Ruane** contributed an article to the ESPN Web site exploring whether the first run of the game is more significant than later runs. He found that while the team scoring first wins 66.5% of the time, the team which scores the second run of the game will win 71% of those games. Runs three through nine also lead to wins more often than the first one does. This is due in part to the fact that the visiting team is more likely to score the first run, but less likely to win than the home team. That same month, **Dan Levitt** discussed **Stephen Jay Gould**'s study of variability in batting average. Prof. Gould found (in his book *Full House*) that such variability decreased steadily through major league history until the 1940's and has been fairly steady since then. Prof. Gould interpreted this to mean that hitting skill had neared the limits of human ability. Mr. Levitt questioned his conclusions, since his own study of pitcher's Opponents' Batting Average showed almost no change in variability from the 1890's to today. He also felt that increases in batting skill since World War II may be masked in Prof. Gould's study by the effects of expansion. (On the other hand, studies cited above found no increase in variability due to expansion.)





Looking at September studies now, we find **Tom Ruane** examining whether designated hitters perform worse than expected. For the years 1980 to 1997, he compared how players hit at each position and at DH versus how they hit overall. He found that while for each defensive position, players hit about as well when in the lineup at that position as they did overall, they hit (in terms of on base and slugging averages) .013 worse than when they played in the field. Pinchhitters did twice as poorly. This helps to explain why DH's do not hit as well as left fielders or first basemen. **Larry Grasso** speculated that the reason for DH's poor performance is that many times players are tired or injured while used as DH. Thus, it is not inherently harder to hit as a DH, but physical condition is to blame. Mr. Ruane did not feel this was an adequate explanation, pointing out that people who spend a lot of time both at designated hitter and playing in the field showed a similar decline in hitting.

A discussion on unearned runs led **Tom Ruane** to present some charts showing the pitchers with the most and fewest unearned runs allowed and their team's fielding average when they pitched, the pitchers whose catchers were charged with the most passed balls (one of the six was not a knuckleballer) as well those with the fewest, how pitchers performed after an error or passed ball was committed, and a comparison of fielding averages behind extreme ground ball and fly ball pitchers. No surprise, fielding averages when the fly ball pitchers were on the mound tended to be higher, although some ground ball pitchers had better luck.

In October, **Tom Ruane** responded to a question as to whether 1-0 games are more common than 3-0 games. He supplied a chart that showed that for all losing scores from 0 to 15, the most common losing margin in 1 run. Thus, if a losing team scores 7 runs, the most common number of runs it will allow is 8. However, if a winning team scores 8 runs, the most common number of runs it allows is 3, not 7. Still, it is apparent from these data, which cover the years 1901-97, runs scored and allowed are not independent of each other.

Clifford Blau, 16 Lake Street, #5D, White Plains, NY, 10603. cliffordblau@geocities.com ♦