

By the Numbers

The Newsletter of the Statistical Analysis Committee of the Society for American Baseball Research

Volume 5, Number 1

February, 1993

Greetings!

Welcome to the first 1993 edition of *By the Numbers*, the newsletter of SABR's Statistical Analysis Committee. As announced in the previous newsletter, I am the new editor of the newsletter and have assumed the chairmanship of the Committee from Don Coffin.

Don has taken on additional duties in his professorial guise, and solicited interest from the Committee in helping out or "taking over" the role of chairman. The baton has officially been passed with the issuance of this edition of the newsletter.

For all committee members, I would like to express my thanks and appreciation for the great job that Don did in resurrecting the Statistical Analysis Committee, and keeping it active over the past several years. On a personal note, I would like to thank Don for the assistance he has provided me in assuming the chairmanship of the Committee. I hope I can maintain the high standards Don has set for Committee activities and quality of the Committee newsletter.

I also want to take this opportunity to welcome new committee members. The committee is open to all SABR members who are interested in the application of analytical methods to the sport of baseball. As a general rule, anything

related to "sabermetrics" is germane to this committee. Committee members range from the ardent researcher to the person who simply likes to read others' research. Of course, all types are welcome, but the survival of the committee relies largely on its active researchers. Thus, I encourage you all to actively engage in research and pass the results along to the rest of the committee members.

Along with individual members' research, the committee sponsors larger efforts as well. These include providing assistance or service to SABR members as a whole or to other SABR committees. An on-going effort of the committee is to compile a complete bibliography of publications incorporating or pertaining to statistical analysis of baseball. The bibliography will include several classification categories (e.g., team vs. individual performance, offense vs. defense, etc.) If you are interested in assisting in this effort, or in learning more about it, please contact **Dave Nichols** at 2555 Bell Avenue, Mountain View, CA 94043.

I plan on continuing the schedule of a quarterly newsletter. The next edition of the newsletter is planned for May/June, in time for the national convention in San Diego. In order to ensure timely delivery

and accurate committee records, please send me any address changes and/or corrections.

Of course, the preamble to the newsletter would not be complete without a plea for new material. The intent of the newsletter is mainly to be a journal of the research of the committee members. For years many researchers worked in isolation and without benefit of communication and knowledge from others (the proverbial "isolated caves" analogy).

The committee and its newsletter hope to be avenues to bring us all closer together. As such, I hope to hear from any of you who are engaged in statistically-based baseball research. To appear in the newsletter, the research need not be finalized or polished (work-in-progress is welcome). Send me your research in any format (electronic DOS/Mac 3.5"/5.25", or hard-copy). Another feature I hope to include in the newsletter besides "research articles" is communication of ideas/feedback from committee members.

Following this greeting, we have a section on news and views from committee members. An announcement on the Retrosheet effort follows. We then have one article by **Barry Codell**. Barry introduces a new statistic called the "Diamond Weight", which is designed to measure a hitter's offensive performance in a more accurate and interpretable manner than does on-base average plus slugging average.

My address has appeared in a couple different places recently with an incorrect zip code. My correct address is: **Rob Wood**, 2101 California St. #224, Mountain View, CA 94040. My home number is (415) 961-6574, and my

daytime number is (415) 854-7101. I look forward to hearing from all of you.

Notes from committee members

In this section of the newsletter, I will pass along news/ideas/information I receive from committee members.

Don Coffin forwards an article by M.R. Lanoue and J.J. Revetta entitled "An Analytic Hierarchy Approach to Major League Baseball Offensive Performance Ratings". The methodology and the results are quite interesting. If you would like a copy of the article, please contact me.

Pete DeCoursey writes a weekly baseball column for the Reading Eagle during the baseball season. Many of his columns deal with the value and use/misuse of baseball statistics. One of Pete's columns that I particularly enjoyed presented the case for Dave Winfield as one of the all-time greats.

Bill French has written a provocative newsletter on the relative importance of various baseball statistics. If you are interested, contact Bill at 1221 Stanford, Oakland, CA, 94608; (510) 653-8074.

Bob Grove writes from Western Australia where he is a sport scientist and baseball coach. He has written two interesting articles on his research. One is on the components of team performance and game outcome in international competition, and the other is on the differences in pitching from a windup versus a set position. If you are interested, contact Bob at The University of Western Australia, Human Movement and Recreation Studies, Nedlands, Perth, Western Australia 6009; (09) 380-2361.

Dan Heisman publishes a lively and informative quarterly newsletter on baseball's active statistical leaders. If you are interested, contact Dan at 1263 Paso Fino Dr., Warrington, PA, 18976; (215) 343-6033 evenings.

Doug Pappas forwards an interesting historical tidbit. In Kings of the Diamond (1965), Lee Allen and Tom Meany quote from Touching Second (1910) by Johnny Evers and Hugh Fullerton. Evers arranged for scorers to track the percentage of ground balls passing through infields of various classifications during the 1909 season. He found a primitive "Zone Rating" of .880 for college infields, .906 for the Central League, .917 for the American Association, and .935 for the majors. Doug estimates that the modern day average major league zone rating is approximately .870 (below what Evers found for a college infield of eighty years ago!). Possible reasons include today's livelier ball and artificial turf.

Joe Slap writes "In comparing players' statistics for careers and for individual seasons, people seem not to take into account the fact that each player did not have to compete against his own team. The batters didn't have to face their teams' pitchers, and vice versa." Joe mentions Lefty Grove not having to face Cochrane, Simmons, Foxx, and Williams. This is intriguing. Does anyone have any thoughts on this issue (e.g., how significant the adjustment would be)?

Dave Smith passes along information on Retrosheet, an effort established to collect, standardize and computerize play by play accounts of "pre-Project Scoresheet" major league baseball games. As most of you know, Project Scoresheet has been (and continues to be) instrumental in providing researchers with the raw data they need to investigate

many important statistical issues in baseball over the past several years. Dave's announcement follows in the next section.

I have a short comment on Dave Smith's excellent article on the usefulness of the quality start statistic which appeared in the previous *By the Numbers*. Remember, one of the criticisms of the stat is that a pitcher can get credit for a quality start even if he pitches only six innings and gives up three runs. Dave showed that by several measures, the quality start is a valid and useful statistic; it does a good job of distinguishing between efforts that lead to a win versus efforts that lead to a loss. I would find it interesting to know the winning percentage of a "Moss Klein" game (six innings, three runs allowed). That is, what is the winning percentage of a hypothetical "Klein pitcher" who started every major league game over the past eight seasons? An estimate of this winning percentage could be obtained by "substituting" the Klein pitcher for the team's starting pitcher, one team at a time, for every game over the past eight seasons. I imagine that the winning percentage would be higher than many (including Moss) believe.

Retrosheet

by Dave Smith

Retrosheet was founded in 1989 for the purpose of computerizing play by play accounts of as many pre-1984 major league games as possible (games since 1984 are available through the Project Scoresheet data base). Retrosheet's work has three distinct aspects.

First is the collection of the game accounts, which have been obtained from several sources. Individual fans have donated copies of programs they scored at the park or on their own scoresheets at home, retired sportswriters have allowed copies to be made of their daily scorebooks, and some major league teams have allowed copies of their play by play accounts. The second activity is the translation of these accounts to a unified, modern system, which is essential since there is an extraordinary variety of scoring systems which have been used. The final activity is the entry of the translated accounts into the computer.

Once data entry has been accomplished, the game accounts will be available to baseball researchers. Details have not been completed, but it is likely that the accounts will be deposited in the Hall of Fame library in Cooperstown.

Retrosheet has been very successful in the collection of game accounts, with some 45,000 pre-1984 games in hand. We have a large group of volunteers who are actively involved in the translation effort and the computer entry is proceeding as well. However, there is still much to be done and any offers of help would be greatly appreciated. Therefore, we ask SABR members to make available copies of any game accounts they might have as well as to volunteer their assistance in the translation and inputting efforts.

Please contact Dave Smith, Retrosheet, 6 Penncross Circle, Newark, DE, 19702; (302) 731-1570.

DW: A Way to Truly Weigh the Diamond!

©1992 by Barry F. Codell

The Diamond Weight (DW) is the third and final number in a series of statistics I have been pleased to share with SABR members since 1977. To recount:

The Base-Out Percentage or BOP (1979 Research Journal) sought to show individual offensive accomplishment in the simplest and most comprehensive of ratios: bases attained to outs accumulated.

The Runs Tallied or RT (1989 Research Journal) addressed that area (team scoring contribution) untouched by the BOP, via the most concise method possible: the averaging of Runs and RBI.

The 1992 Diamond Weight statistically conveys the essential action stemming from the batter's box while critiquing that latest of sacred cows, On Base Average + Slugging Average (OBS).

OBS seemingly displays an "equal partnership" between two numbers, each complementing the other's strength, while compensating the reciprocal weakness. However, closer scrutiny shows this to be illusory.

Let's analyze two hypothetical hitters from the outset. Player A has doubled and walked, creating a 3.000 OBS, while Player B has doubled and singled, compiling a 2.500 OBS! While a walk can be valuable, who can argue it is worth more than a single?

The problem lies with the combining of two numerators without a common denominator. On-Base Average is

calculated by Reached Base total divided by Plate Appearances; Slugging Average is calculated by Total Bases per At Bats. The percentages added have different limits (1.000 for OBA, 4.000 for SA), so what can their addition mean? What is, for example, a .750 OBS? The scales, it appears, have been erroneously put together.

The Diamond Weight is a more meaningful and logically proportioned formula:

$$\frac{(RB+TB)}{PA} = DW$$

The DW reminds us instantly of what constitutes a "plate appearance perfection": the home run. A homer accounts for the maximum weight (5.000) of diamond successes (1 grain reached base, 4 grains total bases) in any venture into the batter's box. (It is only at this 5.000 that the OBS agrees, coincidentally. When Bob Nieman of the Browns uniquely opened his career with consecutive home runs, he was maintaining his 5.000 DW and 5.000 OBS!)

Of course, all players attain mortality, and none circles the diamond each time he steps up to the plate. Given this reality, the DW takes over, better serving the obvious (Player A, 2.000; Player B, 2.500 DW) with its precise measuring of plate appearances.

Let's now look at baseball's two premier batsmen the past two seasons, Barry Bonds and Frank Thomas. Using the DW formula for 1991 and 1992, we

see Bonds with 539 reached bases and 577 total bases in 1,246 plate appearances. Thomas with 629 reached bases and 616 total bases in 1,411 plate appearances, leading to their .880 and .882 Diamond Weights, respectively, over that span.

What is exemplified is the truer balance of On-Base and Slugging considerations. Within his .880 DW, Bonds' Reached Bases led to 49.2% of his DW, and his Total Bases led to 50.8% (in his OBS, the percentage was 43.3% to 56.7%).

Within Thomas's .882 DW, Reached Bases comprised 50.4% of his DW and Total Bases comprised 49.6% (OBS showed a 45.1% to 54.9% comparison). League averages reflect this tendency.

The Diamond Weight therefore says that Thomas has a diamond success (a reached or total base) of .882 per plate appearance (also translatable to 17.6% of diamond perfection through the ratio of actual to possible, i.e., 1,245 Diamond Successes (RB+TB) in 7,055 opportunities (PA x 5)).

Since the BOP and RT have become known, many of their offshoots have been published, from pitchers' BOP allowed (1979 Baseball Graphics) to RT "one-man lineup" (1992 Baseball Black Book). Personally, I currently try Base-Out Scoring Translation (BOST) and Runs Tallied/Outs Projection (RTOP) to keep tracking the elusive diamond figure.

Here's hoping the DW and its further perusal provide Society members food for imagination and nourishment for action -- a supplanting of the OBS!

Here are the 1992 DW leaders:

	Reached	Total	Plate	Diamond
	Base	Bases	Appearances	Weight
	<u>(RB)</u>	<u>(TB)</u>	<u>(PA)</u>	<u>(DW)</u>
1. B. Bonds	279	295	612	.938
2. G. Sheffield	238	323	618	.908
3. E. Martinez	230	287	591	.890
4. F. Thomas	312	307	711	.871
5. K. Griffey	233	302	617	.867
6. M. McGwire	220	273	571	.863
7. F. McGriff	249	295	632	.861
8. A. Van Slyke	261	310	685	.834
9. R. Sandberg	258	312	691	.825
10. K. Puckett	260	313	696	.823