A Statistical Analysis of Hitting Streaks in Baseball
Author(s): S. Christian Albright
Source: *Journal of the American Statistical Association*, Vol. 88, No. 424, (Dec., 1993), pp. 1175-1183
Published by: American Statistical Association
Stable URL: http://www.jstor.org/stable/2291254
Accessed: 04/04/2008 21:35

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://dv1www44.jstor.org:6085/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=astata.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We enable the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

http://dv1www44.jstor.org:6085

A Statistical Analysis of Hitting Streaks in Baseball

S. Christian ALBRIGHT*

This article presents a study of hitting streaks for individual batters in the game of baseball. All baseball fans know that players go through periods where they never seem to make an out and other periods where nothing will get them to base. There is no doubt that "hot streaks" and "cold streaks" do occur. The question explored here is whether these streaks occur more (or less) frequently than would be predicted by a probabilistic model of randomness. I examined the records of many "regular" Major League players through four seasons, 1987–1990 and used several statistical methods to check for streakiness. These include standard methods such as the runs test, as well as a more complex logistic regression model with several explanatory variables. Based on all of these methods, there is no doubt that a certain number of players exhibited definite streakiness in certain years. But the evidence also suggests that the behavior of all players examined, taken as a whole, does not differ significantly from what would be expected under a model of randomness. Furthermore, none of the players examined in the study exhibited unusually streaky behavior over the entire four-year period.

KEY WORDS: Logistic regression; Runs tests; Streaks.

1. INTRODUCTION

This article describes a statistical study of hitting streaks in Major League baseball. Baseball has always been a game of statistics. There are several thick books published annually, such as *The Elias Baseball Analyst* (Siwoff, Hirdt, and Hirdt 1988), that contain more statistics than even the most devoted baseball fan can fathom, and the number of obscure statistics quoted on televised games appears to be growing constantly. But it is difficult to find time series statistics on individual batters that indicate whether they hit in streaks or suffer through slumps. Of course, all baseball fans know that batters do occasionally go 1 for 27 or bat over .600 in a given week. (For non-baseball fan readers, this means 1 success in 27 trials and a 60% success rate; other baseball terms will be explained as necessary.) The statistical question is whether batters do this more often than would be expected under an assumption of "randomness" (to be defined precisely later). This is the question I will examine in this article.

The first obstacle in such a study is obtaining the required data. Despite the voluminous output of baseball data, it is difficult to find time series data on individual players. It is obviously not enough to know that a player went 2 for 5, as would be quoted in a typical box score; the *sequencing* of success and failures is required. Beyond this, it is also useful to know a variety of situational variables associated with each at-bat, such as the score, the number of runners on base, and the quality and "handedness" (left or right) of the pitcher. These situational variables can then be used as predictors of batting success in a logistic regression model, as will be explained in more detail later. The point here is that such "micro" data are not readily available, and this may explain why no such study, at least on such a wide scale, has been done to date. But a company called Project Scoresheet has painstakingly collected and computerized data on every play of every game during the last half of the 1980s, and they kindly shared their data so that the present research could be performed.

This data base permitted me to examine the records of all "regular" players in both leagues over a four-year period, 1987–1990. I arbitrarily defined a regular player as one with at least 500 at-bats in a given year, where "at-bats" include unofficial at-bats that result in walks or sacrifices. This resulted in 501 distinct records, 160 of which were for 40 players who were regulars all four years. With this abundance of data, two basic questions could be addressed. First, do a significant number of players exhibit streakiness over the course of a season? Second, are some players perennially streaky? As I will document, the statistical evidence provides essentially negative answers to both of these questions despite the fact that several of the 501 records exhibit definite streakiness.

The first question requires further explanation. One might argue that if *some* players exhibit statistically significant streakiness, then this rules out a model of randomness. But consider the analogy of 501 people flipping perfectly fair coins for a long period. Purely by chance, some of them will produce sequences of heads and tails that can be classified as streaky (many fewer runs of heads or tails than expected, for example). Obviously, this does not mean that the entire process is nonrandom. One would regard the entire process as nonrandom only if *many* of the 501 people produced unusually streaky sequences. The same is true for the baseball data. With 501 seasonal records available, some are bound to be streaky purely by chance. Therefore, I concede from the start that certain players exhibit definite streakiness, at least in certain seasons, and provide evidence of this. The more difficult question I am asking is whether more players are significantly streaky than would be predicted by a model of randomness.

2. PREVIOUS RESEARCH

Baseball has attracted the attention of a fair number of statisticians and operations researchers. A sampling of their research appears in the books edited by Machol, Ladany, and Morrison (1976) and Ladany and Machol (1977), both of which contain articles on baseball and other sports. Evidently, however, very little research has been done on streakiness in baseball hitting. The 1988 version of *The Elias Baseball Analyst* (Siwoff et al. 1988) reports its own study of streakiness for the 1984, 1985, 1986, and 1987 seasons.

* S. Christian Albright is Professor of Decision and Information Systems, Business School, Indiana University, Bloomington, IN 47405.

© 1993 American Statistical Association
Journal of the American Statistical Association
December 1993, Vol. 88, No. 424, Applications & Case Studies

Although the authors do not get into questions of formal statistical significance, their results are interesting and somewhat similar to those presented here. They tabulated the batting averages of players in games directly after a "hot streak" or a "cold streak" (defined as a series of five games over no more than seven days during which the player batted at least .400 or less than .125) and compared these batting averages to the players' overall batting averages. They found that batting averages were just about as likely to be higher following cold streaks as following hot streaks. Using several years of data, they also tried to see whether streaky hitters in a given year were likely to be streaky in other years. This would answer the question of whether certain players are perennially streaky or whether streakiness tends to be a one-year phenomenon. Unfortunately, there were not enough streaky hitters in any particular year to yield meaningful results. One other paper that discusses baseball streaks was published by Gould (1989), but this paper is more philosophical than quantitative.

In the sport of basketball, Gilovich, Vallone, and Tversky (1985) and Tversky and Gilovich (1989a) have studied streakiness in shooting baskets. Their study is similar to the current investigation but much more limited. They examined the shooting records of nine Philadelphia 76ers during an entire season and used a number of formal statistical tests to determine whether players shoot in streaks. Their basic conclusion is that the players they examined did not tend to be streaky. In particular, they were just about as likely to make a shot after making several shots in a row as after missing several in a row. But the authors acknowledge that their analysis omitted complicating factors that are probably relevant but difficult to quantify. For example, not all basketball shots are equally difficult (e.g., a layup is a much higher percentage shot than a 20-foot jump shot), and it could be that hot players who have made three consecutive long jump shots try something too ambitious on the next shot, whereas cold players refuse to shoot until they get an easy layup. Although this factor is not built into their statistical model, it could certainly be part of the reason for their results. The papers by Larkey, Smith, and Kadane (1989), Tversky and Gilovich (1989b), and Hooke (1989) debated the ideas and conclusions of the basketball papers referenced earlier.

Baseball and basketball players are very different; however, there are also situational variables in baseball that should be included in a realistic model of hitting. The conditions on successive at-bats change: A left-handed pitcher may replace a right-hander, there may now be runners in scoring position (on second or third base), the score may now be tied in the bottom of the ninth inning, and so on. Unlike the basketball study, the analysis of baseball hitting streaks formally incorporates some of these factors.

3. METHODOLOGY

The statistical analysis of streakiness might appear obvious: One should simply check whether a batter does significantly better or worse following a hot or cold streak than he does normally. However, several complicating factors make the problem more challenging, and all are related to the peculiarities of baseball.

First, there is the technical question of classifying the results of successive at-bats. Should a distinction be made between different types of hits (singles, doubles, triples, and home runs), or should they all just be labeled "successes?" And what about walks and sacrifices? Walks (free trips to first base because the pitcher doesn't throw strikes) are not counted in official batting averages, but they are counted in on-base percentages. Sacrifices (intentional outs used to move a runner along) are also not counted. But if one decides to classify any given at-bat as a "success" or a "failure," in the sense of getting a job done, it makes sense to treat both walks and sacrifices as successes. I decided *not* to distinguish between types of hits, so that singles and extra-base hits are counted equally. Concerning walks and sacrifices, however, I decided to run two separate sets of analyses: one where walks and sacrifices count as successes, the same as hits, and one where walks and sacrifices are ignored. In the former case I counted *all* at-bats, whereas in the latter case I counted only *official* at-bats. Obviously, for some hitters it could make a big difference how walks are counted, because a relatively large proportion of their at-bats result in walks. (Ricky Henderson is an example). This is less true for sacrifices, because very few batters have more than 10 sacrifices in a given year. Nevertheless, I counted sacrifices the same as walks and either treated them all like hits or ignored them entirely.

Second, a realistic analysis of hitting streaks should not look *only* at a batter's time series of successes and failures in successive at-bats. As mentioned earlier, the conditions on successive at-bats change continually. Therefore, it is certainly conceivable that a batter's probability of success on a given at-bat depends at least as much on these situational variables as on his recent batting history. This poses two questions: Which situational variables are relevant, and how should they be incorporated into a statistical model? I included the following conditions: (1) whether the game is at home or away; (2) whether the game is at night or during the day; (3) whether the game is on grass or artificial turf; (4) whether the current pitcher is right-handed or left-handed, as well as his season-long *ERA* (earned-run average—the number of earned runs allowed per nine innings); (5) the number of runs the batter's team is ahead or behind by (positive if ahead, negative if behind); (6) whether there are two outs or less than two outs; (7) whether there are runners on base, and whether there are runners in scoring position; and (8) whether the game is yet into the seventh inning. These are factors that are typically claimed to affect batting success, and I found each of them to be significant for a subset of the players (although some were significant for many more players than others). I will explain later how I incorporated these factors into a model.

Third, an important question is how "streakiness" is defined. For example, the basketball study of Gilovich et al. (1985) and Tversky and Gilovich (1989a) looked at the fraction of shots made following a streak of k consecutive successful shots or k consecutive missed shots, where $1 \le k \le 3$. The same can be done in baseball, but there is no reason to restrict the analysis to *consecutive* strings of successes or failures when defining a streak. There are any number of ways to define a hot streak or a cold streak. The approach taken by the Elias writers (in Siwoff et al. 1988) and described here in Section 2 is only one of many possibilities. I dealt with this in the logistic regression model by using the *number* of successes in the most recent k at-bats, for various values of k, as an explanatory variable for the probability of success in the current at-bat. This has two advantages. First, by varying k from 1 to 20 (where the upper limit 20 was chosen arbitrarily), short-term streakiness as well as long-term streakiness can be examined. Indeed, some players appeared to be influenced more by their long-term history than by their short-term history, and others vice versa. Second, by using the *number* of successes in the previous k at-bats as the explanatory variable, I do not have to settle on *specific* definitions of hot or cold streaks (e.g., 0 for 10 or 4 for 6), as was done in the basketball study.

Finally, it is necessary to define randomness as it is used here. If the situational variables are temporarily ignored, then a model of pure randomness assumes that a given player follows an independent Bernoulli process with a parameter p that is fixed throughout the season (but varies across players). If the situational variables are incorporated, then randomness still implies a Bernoulli process, but the parameter p is allowed to vary from one at-bat to another as a function of the situational variables. In either case the key is the independence assumption: The batter's probability of success on one at-bat does *not* depend on the results of previous at-bats. Moreover, except for the influence of situational variables, the value of p does not change throughout the season. Streakiness is then one form of nonrandom behavior, where a batter's probability of success is smaller after recent failures and larger after recent successes. Note, however, that a batter whose results are not random could also be the opposite of streaky—that is, he could bat better after failures and worse after successes. There are many examples of this type of behavior in the data set. For lack of a better term, I will refer to this opposite-of-streaky behavior as "stability."

Because of these considerations, numerous statistical methods could be applied to baseball data. Clearly, some of these ought to "overlap" in the sense of providing similar information; however, they might not always measure the same thing. Therefore, I found it useful to apply several types of analysis to the same data and then see whether an overall pattern emerged.

I will now discuss the particular analyses that I performed. Some of these are fairly "standard" in the sense that they are well-known and they look only at the time series of individual at-bats, whereas others incorporate the situational variables discussed earlier. Also, each was performed when all at-bats count (i.e., when hits, walks, and sacrifices are successes), and also when only official at-bats count (i.e., when only hits are successes). Clearly, these methods are not the only ones that might be useful. For example, Naus (1974) found the distribution of the maximum number of successes in any m consecutive trials in a series of M Bernoulli trials with a given number of successes, and he suggested that this might be used to examine baseball hitting streaks. But a line had to be drawn somewhere. I decided to limit the analysis to fairly well-known methods, with one exception: I also looked for serial correlations of individual at-bats with the modified Box–Pierce statistic (Farnum and Stanton 1989), but obtained no additional insights not revealed by the other methods.

3.1 Runs of Individual At-Bats

Probably the most obvious method is to count runs. For a given player, I classify each at-bat as a success or failure, and then count the season-long number of runs (not to be confused with the baseball meaning of "runs"). Fewer runs than would be expected under a model of randomness then provides an indication of streakiness. In contrast, more runs than expected indicates stability, where a batter tends to do better after failures and worse after successes.

Let N_S, N_F, N, and R be the number of successes, the number of failures, the number of at-bats, and the number of runs for a given player in a given season. Then the runs test is based on the statistic $Z = [R - E(R)]/\sigma_R$ (Canavos 1984), where

$$E(R) = \frac{N + 2N_SN_F}{N}, \qquad \sigma_R = \sqrt{\frac{2N_SN_F(2N_SN_F - N)}{N^2(N-1)}}.$$

Under randomness Z is approximately $N(0, 1)$. In this study the approximation should be good, because N is quite large.

3.2 Markov (First-Order) Dependence of Individual At-Bats

By defining the "state" of any at-bat as success or failure (1 or 0), one can check whether successive at-bats form a first-order Markov chain. (It is possible to check for higher-order dependence, but I did not do so in this study). Under randomness the estimated rows of the one-step transition matrix should be similar. A standard chi-squared test (Canavos 1984) is appropriate. The test statistic is given by

$$\chi^2 \simeq \frac{N(N_{00}N_{11} - N_{10}N_{01})^2}{N_0^2N_1^2},$$

where N_{ij} is the number of times where state i is followed by state j, N_i is the number of occurrences of state i, and N is the number of at-bats. (The standard exact formula for χ^2 is slightly different because the column sums may differ by 1 from the row sums, but this difference is negligible for large N.) Assuming randomness, this statistic has an approximate chi-squared distribution with 1 degree of freedom. For each player I also calculated the difference between the proportion of successes following a success and the proportion of successes following a failure. This difference shows how *much* better or worse a player's batting average was following a success versus a failure.

3.3 Logistic Regression Models with Explanatory Variables

The drawback to all of the preceding statistical methods is that they use only the time series of successes and failures in individual at-bats. As stated earlier, the amount of streakiness could certainly be affected by situational variables. Therefore, I also used a probabilistic model that incorporates these variables, as follows.

Let X_n be 1 or 0, depending on whether a player is or is not successful on at-bat n, and let $p_n = P(X_n = 1)$. I used a logistic regression model (see, for example, Hosmer and Lemeshow 1989) of the form

$$\ln\left(\frac{p_n}{1 - p_n}\right) = \alpha + \beta Y_n + \sum_{k=1}^{K} \gamma_k Z_{kn},$$

where Y_n is related to the batter's recent history of success and Z_{kn} is an expanatory variable relating to the situational variables in effect during at-bat n. For example, Z_{1n} might be 1 or 0, depending on whether the current pitcher is right-handed or left-handed, Z_{2n} might be the pitcher's ERA, Z_{3n} might be the number of runs the batter's team is currently ahead or behind by, and so on. These Z's capture the information discussed earlier that might affect a player's success probability.

In contrast, Y_n indicates the player's success in recent at-bats. I tried eight possibilities. I first let Y_n be the number of successes in the most recent k at-bats, for $k = 1, 2, 3, 6, 10$, and 20. I also let Y_n be an exponentially weighted sum of the batter's last 20 X's, specifically,

$$Y_n = \sum_{i=1}^{20} \delta^{i-1} X_{n-i}$$

for $\delta = .80$ and $\delta = .95$. The purpose here was to allow the probability of success to depend on the last 20 at-bats, but to give more weight to the more recent at-bats.

For each definition of Y_n, I estimated two logistic regression equations that included Y_n and a constant term. The first equation also included all situational variables, whereas the second entered these situational variables in a stepwise manner, with entry criterion set at the .10 level. The results in the aggregate from these two approaches were similar; I will concentrate only on the former in the following section. These equations allowed me to see whether the probability of success depends on recent history, *after* taking into account current game conditions. Note that for each equation, a positive value of β indicates streakiness, a negative value indicates stability, and a zero value is consistent with randomness.

4. RESULTS

Because of the abundance of the data and the many ways they can be analyzed, it is possible to present an overwhelming number of graphs and statistics. To reduce this to manageable proportions, I will focus on a subset of the analyses, limiting the discussion to the case where walks and sacrifices *are* counted. The other case—where walks and sacrifices are ignored—did not produce exactly the same results, but they were similar in most respects.

I begin by focusing on two of the more extreme of the 501 records, simply to show how streakiness and stability manifested themselves. Dwight Evans, right fielder for the Boston Red Sox, had an unusually streaky season in 1988. He had 628 at-bats and was successful 241 times, for a .384 on-base average. Assuming a model of randomness, his expected number of runs (streaks of 0's or 1's) was 298, whereas his actual number was only 261. The corresponding Z statistic was -3.12, one of the most negative in the data set. The relatively low number of runs was due to many long runs, as indicated in Figure 1, a time series graph of run lengths, where runs of 0s are shown as negative lengths and runs of 1s are shown as positive lengths. The Markov analysis for Dwight Evans produced a highly significant chi-squared statistic; in fact, he batted 124 points higher following a success than following a failure.

Finally, in the logistic regression model, all of the coefficients of his history variables (the Y_n's) were positive, and all but one were significant at the .10 level. For example, if Y_{1n} is the number of successes in the previous at-bat, then the estimated equation (after dropping insignificant situational variables) is

$$\ln\left(\frac{p_n}{1 - p_n}\right) = -.901 + .377 R23 - .415 T + .098 ERA + .489 Y_{1n},$$

where $R23$ is 1 or 0 depending on whether any runners are in scoring position, T is 1 or 0 depending on whether the pitcher throws right- or left-handed (Evans bats right-handed), and ERA is the pitcher's ERA. Exponentiating the coefficient of Y_{1n} produces an odds ratio of 1.631, so that the odds of success, $p_n/(1 - p_n)$, increase by a factor of 1.631 when Y_{1n} goes from 0 to 1 (and the situational variables are held constant). This was one of the largest odds ratios obtained for any player.

In contrast, consider the unusually stable 1989 season for Harold Reynolds, the second baseman for the Seattle Mariners. He batted 667 times, with 245 successes, for a .367 on-base average. His expected and actual number of runs were 311 and 346, and the associated Z value was 2.92. On close study, his time series of run lengths, shown in Figure 2, exhibits many more (and shorter) runs than the graphs for Dwight Evans. This is confirmed by the Markov analysis. Again, the chi-squared statistic is highly significant, but this time it is because Reynolds batted 115 points *lower* following a success than following a failure.

His pattern continues in the logistic regression analysis. The coefficients of all eight Y_n's are negative, and five of these are significant at the .10 level. For example, the estimated equation when Y_{1n} (one of the significant Y_n's) is included is

$$\ln\left(\frac{p_n}{1 - p_n}\right) = -.911 + .358 I7 + .114 ERA - .524 Y_{1n},$$

where $I7$ is 1 or 0 depending on whether the game is yet into the seventh inning and ERA is as above. For Reynolds the odds ratio for Y_{1n} is .592; that is, the odds of a success decrease considerably when Y_{1n} goes from 0 to 1.

The preceding two records are among the most extreme of the 501 studied. Interestingly, however, neither points to a recurring pattern for these two players. Although Dwight Evans was slightly streaky in 1989 and 1990, his 1987 record was very similar to Harold Reynold's 1989 record; that is, extremely stable. Similarly, Harold Reynolds was very stable in 1987 and somewhat stable in 1990, but he was extremely streaky in 1988. Such reversals for individual players were not at all unusual, as will be seen.

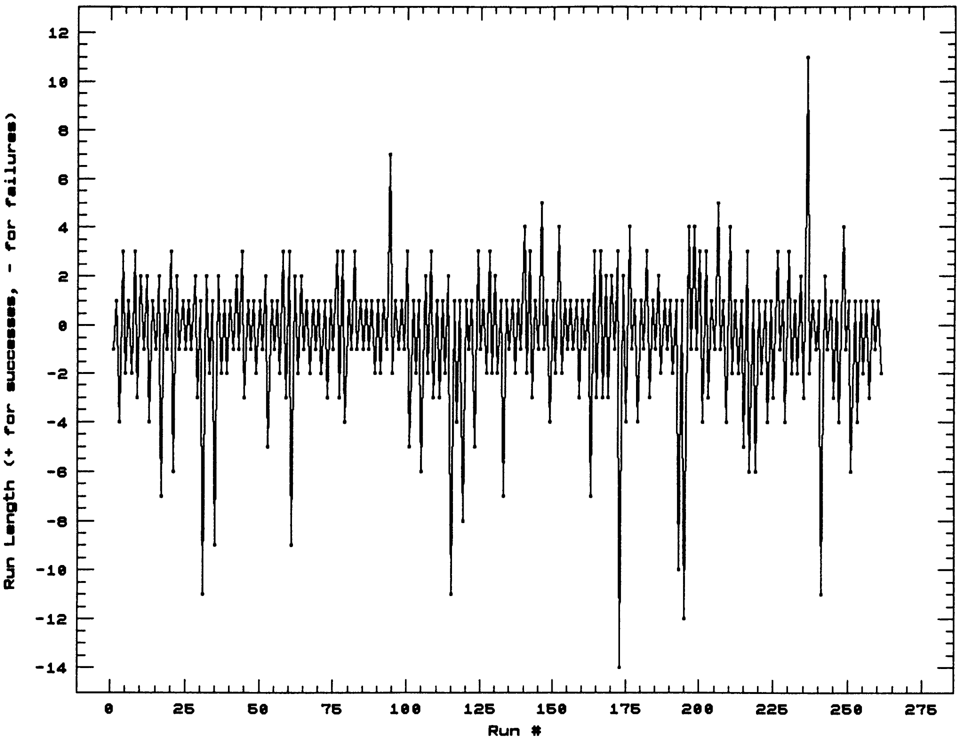

Figure 1. Time Series of Run Lengths for Dwight Evans in 1988.

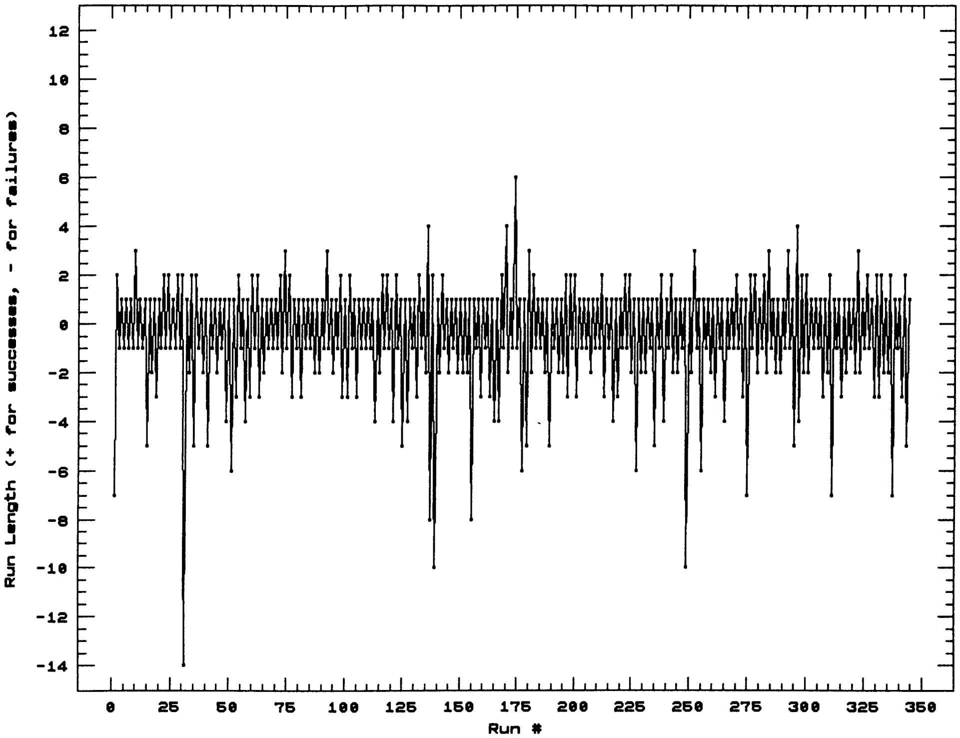

Figure 2. Time Series of Run Lengths for Harold Reynolds in 1989.

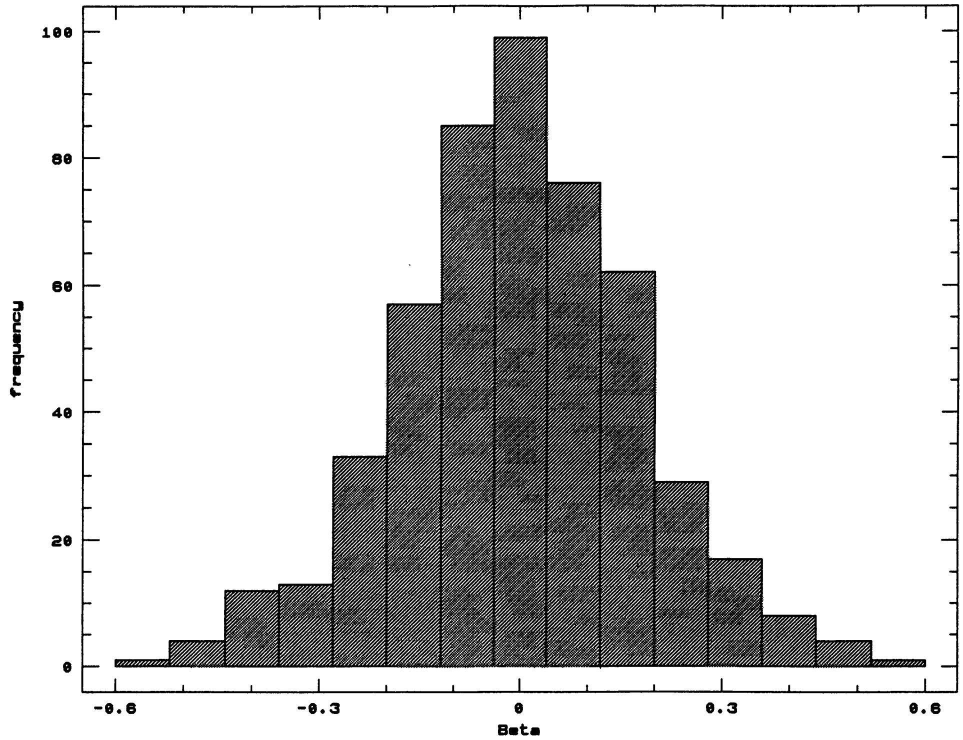

Figure 3. Histogram of Short-Term History Beta in Logistic Regression Model.

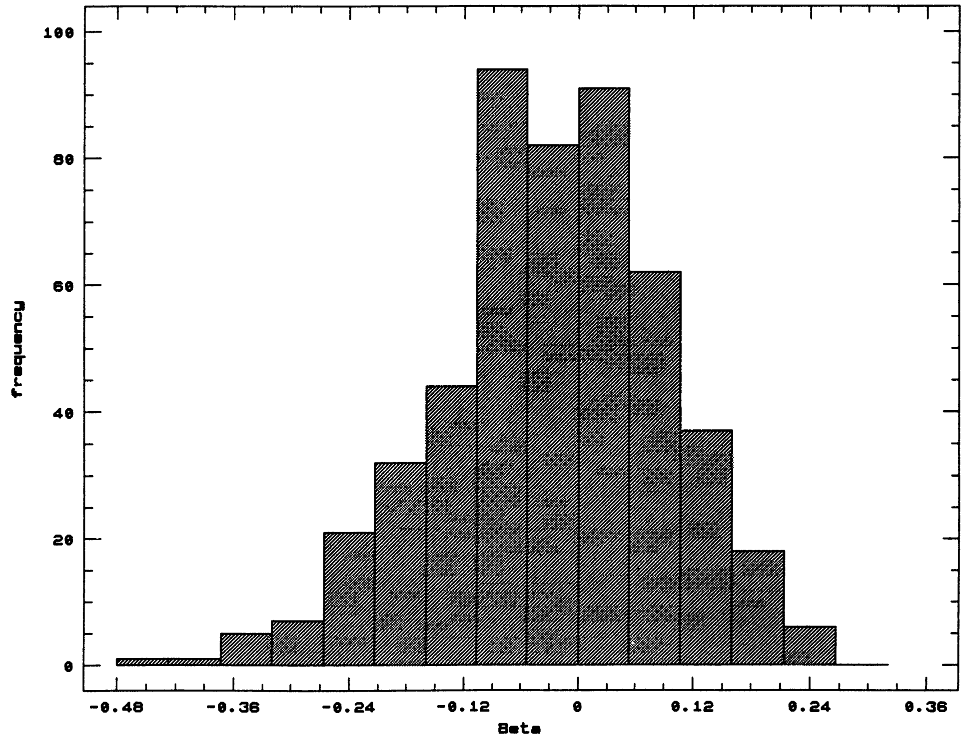

Figure 4. Histogram of Long-Term History Beta in Logistic Regression Model.

Extreme records are interesting to study, but they do not reflect the behavior of the majority of the players. First, consider the runs Z values. According to a model of randomness, these Z values should be distributed normally with mean 0 and standard deviation 1. The actual 501 Z values had a sample mean $-.256$ and variance .993. Their histogram is normally shaped, although the mean is significantly to the left of 0, indicating some streakiness in the aggregate. But there are nearly as many batters with stable records (positive Z's) as those with streaky records (negative Z's).

These results are reinforced by the Markov analysis. The 501 differences (on-base average following success minus on-base average following failure) are approximately normal, with mean 8.90 and standard deviation 40.43. This mean is significantly positive, just as the mean of the Z values was significantly negative, again indicating that more players are streaky than stable (although there are many of each).

The logistic regression models, however, produce a somewhat different, and possibly truer, picture. For the sake of brevity, I will limit the discussion to two of the history variables: Y_{1n} (defined earlier) and Y_{7n} (the exponentially weighted average of the last 20 at-bats, with weight $\delta = .8$). These two variables represent short-term and long-term history; the other Y_n's offer no additional insights. Figures 3 and 4 show histograms of the estimated β's for these two history variables. Note that the histogram for the short-term variable is almost exactly symmetric around 0 (mean $-.0041$), where the histogram for the long-term variable is

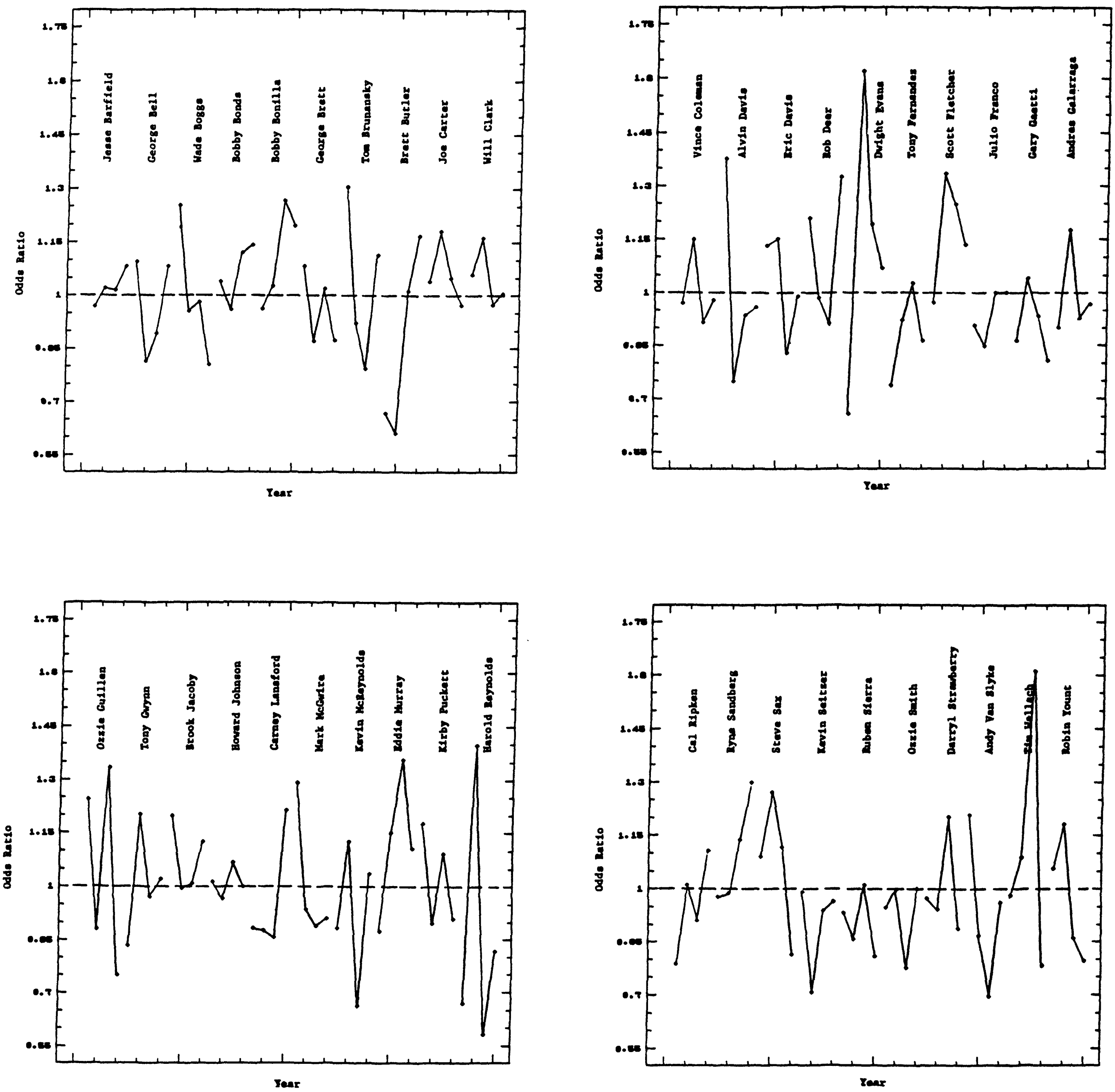

Figure 5. Time Series Graphs of Odds Ratios for Short-Term Variable for 4-Year Regulars.

somewhat skewed to the left and has a negative mean. So the tendency toward streakiness seen in the runs and Markov analyses is now either eliminated or even turned around. Even if one considers only those season-long records with statistically significant β's, the picture is about the same. Of the 501 equations with Y_{1n}, 46 of the β's were significant at the .10 level. Of these 46 β's, 25 were negative and 21 were positive. For the equations with Y_{7n}, there were 57 significant β's, only 20 of which were positive.

Based on the preceding evidence (and on similar evidence from equations with the other Y_n's), three generalizations are possible. First, less streakiness and more stability tend to surface in the logistic regression analysis than in the runs or Markov analysis. Second, there are generally about as many statistically significant β's as one would expect from a randomness model. For example, there were 46 and 57 significant β's; one would expect about 50 out of 501. Third, there seems to be a trend toward stability as we go from short-term to long-term history. For example, there were many more significantly *negative* β's in the equations with Y_{7n} than in those with Y_{1n}. Many (about half) of the batters are more likely to be successful if their previous one or two at-bats were successful. But the probability of success for the majority of batters is a *decreasing* function of their cumulative success in the last 10 to 20 at-bats.

Because many of the batters were regulars in one or more consecutive years, it is also possible to examine whether streakiness or stability carries over from one year to the next.

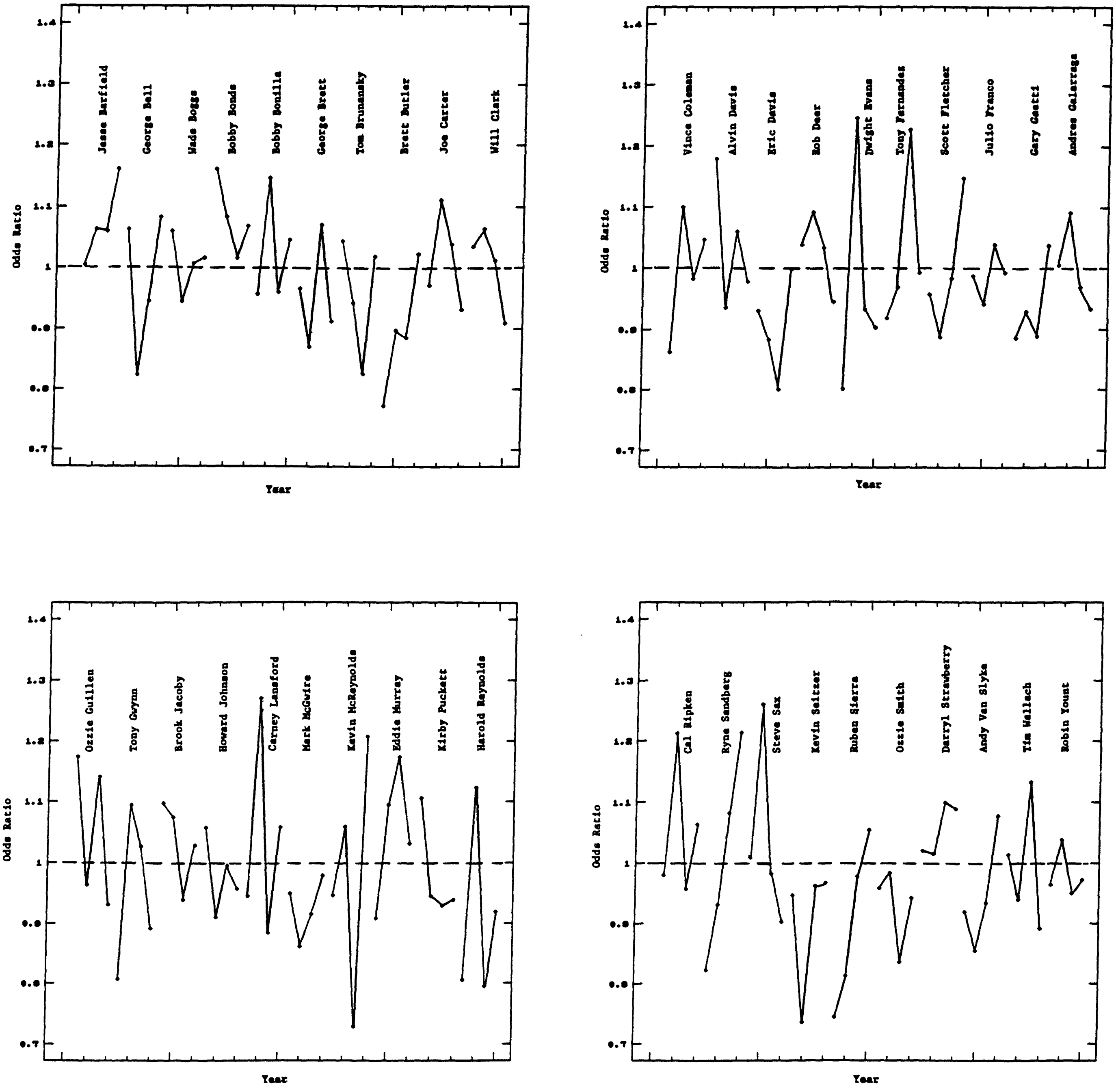

Figure 6. Time Series Graphs of Odds Ratios for Long-Term Variable for 4-Year Regulars.

There were 237 (out of 501) records where a batter was a regular in the following year. For these 237 records, a scatterplot (not shown) of the β's for one year versus the β's from the previous year for the short-term variable Y_{1n} exhibits no structure, with associated correlation only $-.080$. From this evidence it is clear that streaky or stable behavior in one year does not guarantee similar behavior in the succeeding year. Similar scattergrams and correlations exist for the other history variables.

Other evidence supporting the lack of year-to-year consistency can be seen by examining the 40 batters who were regulars for all four years. Figures 5 and 6 give 40 plots of the odds ratios $\exp(\beta)$ over time, one for each player. The graphs in Figure 5 are based on the short-term variable Y_{1n}; those in Figure 6 are based on the long-term variable Y_{7n}. The lack of consistent behavior across years is clear. A few of these batters showed some consistently streaky (or stable) behavior over the four years according to one of the history variables but not the other, but even in these cases, the odds ratios were usually not very far from 1. Even more telling is the fact that those players with extremely large or small odds ratios in one year did not sustain them over the other years. Again refer to the graphs of Dwight Evans and Harold Reynolds for good examples of this. It appears that if there are any extremely streaky batters over a multi-year period, this analysis has failed to find them.

5. CONCLUSIONS

The data analysis performed here, on 501 plus-500-at-bat records over the four seasons from 1987 to 1990, has failed to find convincing evidence in support of wide-scale streakiness. In fact, the evidence is more in line with a model of randomness. It is certainly true that some players exhibit significant streakiness during a given season, but this would be expected under a model of randomness, just as one would expect a certain proportion of people flipping fair coins to produce streaky sequences of heads and tails. The proportion of batters who exhibited nonrandom behavior was reasonably close to that predicted by a random model. Furthermore, in the logistic regression model with situational variables included, more of the significantly nonindependent records indicated the *opposite* of streakiness than streakiness. Finally, an examination of 40 players who were regulars for all four years showed a variety of year-to-year patterns. But not a single one of these players exhibited significantly streaky behavior over the entire four-year period.

I have made no attempt in this article to delve into the psychology of hot or cold streaks. Undoubtedly, most players in the midst of a slump lose confidence, swing badly, or simply experience bad luck, and this affects their performance until they can break out of the slump. The same type of comment can be made for batters who are currently hot. But the point of view I have taken is that of an objective outside observer, who simply records current game conditions and sequences of successes and failures. To this outside observer, it appears that actual performance is being generated in a manner reasonably consistent with a model of randomness.

[Received April 1989. Revised February 1993.]

REFERENCES

Canavos, G. C. (1984), *Applied Probability and Statistical Methods,* Boston: Little, Brown.

Farnum, N. R., and Stanton, L. W. (1989), *Quantitative Forecasting Methods,* Boston: PWS-Kent.

Gilovich, T., Vallone, R., and Tversky, A. (1985), "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology,* 17, 295–314.

Gould, S. J. (1989), "The Streak of Streaks," *Chance,* 2(2), 10–16.

Hooke, R. (1989), "Basketball, Baseball, and the Null Hypothesis," *Chance,* 2(4), 35–37.

Hosmer, D. W., and Lemeshow, S. (1989), *Applied Logistic Regression,* New York: John Wiley.

Ladany, P. L., and Machol, R. E. (1977), *Optimal Strategies in Sports,* New York: North-Holland.

Larkey, P. D., Smith, R. A., and Kadane, J. B. (1989), "It's Okay to Believe in the 'Hot Hand'," *Chance,* 2(4), 22–30.

Machol, R. E., Ladany, S. P., and Morrison, D. G. (1976), *Management Science in Sports,* New York: North-Holland.

Naus, J. (1974), "Probabilities for a Generalized Birthday Problem," *Journal of the American Statistical Association,* 69, 810–815.

Siwoff, S., Hirdt, S., and Hirdt, P. (1988), *The 1988 Elias Baseball Analyst,* New York: Macmillan.

Tversky, A., and Gilovich, T. (1989a), "The Cold Facts About the 'Hot Hand' in Basketball," *Chance,* 2(1), 16–21.

——— (1989b), "The 'Hot Hand': Statistical Reality or Cognitive Illusion?," *Chance,* 2(4), 31–34.