Statistics, Sports, and Some Other Things

Robert Hooke *Westinghouse Research Laboratories*

Spectator sports provide more than just observation of athletes who perform with admirable skill. There is, for example, the drama of a young quarterback trying to lead a professional football team for the first time in front of 70,000 onlookers or that of a veteran pitcher calling on his experience to augment his dwindling physical resources in a crucial game of a close pennant race. Because these dramas are truly "live" and unpredictable, they are much more fascinating to some people than the well-rehearsed performances of the stage.

Not every moment in sports is dramatic, of course, but throughout any contest between professionals, the spectator is privileged to watch a group of people carrying out their jobs almost in full public view, to see how they meet their problems and how they react to their own successes and failures. Baseball and football provide especially good opportunities for such observation because each of these games consists of a sequence of plays, as opposed to the fairly continuous action of basketball, hockey, soccer, and racing sports. Spectators

see more than strikeouts and home runs, completed passes and interceptions. They see a manager gambling on a hit-and-run play or a quarterback deciding to pass for the first down he desperately needs. Fans have opinions on what their team should do in various situations, and they watch to decide whether their manager or coach is a good strategist or a poor one.

Management in professional sports has many similarities to management in business and industry. Some managers and coaches are smarter than others, and some make use of more advanced methods than others do. This is true in sports, as it is in business, in spite of the folk wisdom of the sports pages that often maintains that all managers and coaches are pros and about equally good. Some can get a great deal out of inferior personnel, but none can overcome more than a certain amount of incompetence among the people who work for them. Some are natural gamblers, some are always conservative, and only a few are intelligent enough to be one or the other depending on what the circumstances call for. Sports managers are different from business managers mainly in that their actions are so much more visible. Because of this visibility we should all be able to learn by watching them as they make their various moves in the goldfish bowl of professional sports.

STATISTICS AND MANAGEMENT

What does all this have to do with statistics? The real concern of statistics is to obtain usable quantitative information, especially about complex situations that involve many variables and uncertainties. "Usable" means that its purpose is to help us to improve our behavior in the future, that is, to help us learn how to extract from these situations more of whatever it is that we're trying to get. Some managers make good use of statistics and some don't; this is true whether they manage factories or baseball teams.

Suppose, for example, that we are manufacturing rubber tires. An expert will no doubt be able to detect from the example that I know nothing about the tire business; in fact, I chose this example because I've never been in the tire business and hence will not implicate real people. At some point in the process, let's suppose that we have a mass of liquid rubber that will ultimately be turned into tires. Being aware that this batch may possibly have been improperly prepared, we would like to test it in some way so that, if it is defective, we can throw it away without wasting money processing it into defective tires. Unfortunately, the true test of a tire is a road test, and we can't road test a batch of liquid, so we must perform some test that we think is relevant, such as a viscosity measurement. This measurement will take time and cost money, and sooner or later someone will raise the question "Is it worth the money we're spending on it?" This is always a good question, and it usually leads to much heated debate. The debate will include arguments based on intuition, experience, laboratory tests, and scientific theory; each has its place in the process of seeking after the truth, but they are basically predictors, and the only way to be sure of what will happen in the field is to see what actually happens in the field. This means that we should collect statistics. After measuring a batch, we should follow it through the manufacturing process and see the quality of a sample of tires made from this batch. We repeat this on another batch, and so on. After a while we can establish the relationship between tire quality and viscosity, and we can use this to determine whether to continue with the test, taking into account the testing costs, the cost of the manufacturing process that follows the test, the value of a good finished tire, and so on. Now you may say, of course, that this method isn't infallible because it involves sampling and, hence, sampling error (see the essay by Neter) and because the process may change unexpectedly, and so on. But this method gets us as close to the truth as we can come, and this final objection merely says that you never have it made, even if you use statistics.

THE STRATEGY OF BUNTING

The student of management behavior can find many instances of this type of problem in sports, and such a student, if smart, can profit from the mistakes that are made visibly on the diamond or the gridiron. Take, for example, the sacrifice bunt in baseball. There are those who swear by it and there are those who seldom use it. They engage in passionate arguments as to whether it is a good strategy. As we shall see, statistics can't settle the issue once and for all, but it can shed a great deal of light on the problem, and most of the argument could be eliminated if people would look at some of the facts.

The sacrifice bunt is a play that is used to advance a base runner from first base to second, or from second to third, normally sacrificing the batter, who is thrown out at first. Many managers use the sacrifice bunt routinely, and they refer to their behavior as "percentage baseball," as if they knew the percentages, which, apparently, most of them do not. The routine is that you bunt if there is a man on first or second, nobody out, and your team is only slightly ahead, tied, or not "too far" behind. One or more runs behind is considered too far for the visiting team, and two or more runs behind too far for the home team, the difference coming from the fact that the home team bats last and can afford to "play for a tie."

Why does the manager decide to bunt? Ultimately, of course, he does it to win more games. At the moment of doing it, he is trying to increase the chance of getting one additional run while giving up some of the opportunity to get several. The theory is simple. It takes at least two singles or a double to score a runner from first, while a runner on second can usually score on a single alone. In addition, if a runner on second can be moved to third with only one out, a score will result from any hit, error, wild pitch, passed ball, long fly, balk, or slow grounder. Proponents of bunting are fond of quoting this list, but it contains some fairly rare events, and this raises the real questions—when we use the bunt, by *how much* is our chance of scoring one run increased, and *how much* do we sacrifice in terms of possible additional runs? Again, the only way to get an answer to this question that is relevant to real major league players playing under the pressure of real games is to take statistics from actual games. There is no way to provide realistic conditions for an experiment, and theory (see Cook, 1966; Hooke, 1967) is of dubious value.

Although records of games played exist in the archives of organized baseball, turning these into usable data is a major task that, if it has been done, has not been made public to my knowledge, except in Lindsey (1963). Lindsey discusses records of several hundred major league games played in 1959 and 1960, and he produces some very interesting statistics, some of which are shown in Table 1.

To see how we, as armchair managers, would use this table to decide about bunting, let's look at the first two lines. (For the moment, we'll think only of average cases, but no good statistician dwells on averages alone, so we'll discuss special situations later.) We start, say, with a man on first and no outs. The table says that this situation was observed 1,728 times (occasionally, perhaps, more than once in the same inning). In a proportion of these cases equal to 0.604, no runs were scored during the remainder of the inning; that is, in 1,044 cases no runs scored, and 1,044/1,728 = 0.604. This means also that the proportion of times at least one run scored from this situation is 1 — 0.604, or 0.396. We use these proportions as estimated probabilities of the various events; thus near the end of a tight game, the number 0.396 measures the average "value" of having the situation of a man on first and no outs. For earlier parts of the game, the value is more closely related to the number of runs that are scored in an inning, on the average, starting from this situation; this is given in the fourth column as 0.813 for the situation in question.

Now if we make a sacrifice bunt that succeeds in the normal way, the runner on first will move to second and there will be one out. Is this a better situation than we had? In the sense of average number of runs scored, it is decidedly worse; the first and second lines of Table 1 show that the average number of runs scored from the man-on-first-no-out situation is 813 per thousand, but from the man-on-second-one-out situation, it is only 671 per thousand. On the average, then, a normally successful bunt loses 142 runs per 1,000 times it is tried. But what about the last inning of a tight game when we only care what has happened to the probability that at least one run will be scored? This figure

Table 1 Relation of runs scored to base(s) occupied and number of outs

Base(s) Occupied	*Number of Outs*	*Proportion of Cases No Runs Scored in Inning*	*Proportion of Cases of at Least One Run Scored in Inning*	*Average Number of Runs in Inning*	*Number of Cases*
1st	0	0.604	0.396	0.813	1,728
2nd	1	0.610	0.390	0.671	657
2nd	0	0.381	0.619	1.194	294
3rd	1	0.307	0.693	0.980	202
1st, 2nd	0	0.395	0.605	1.471	367
2nd, 3rd	1	0.27	0.73	1.56	176

Source: Lindsey (1963).

has dropped from 0.396 to 0.390; these numbers are so close that their difference is readily explained by chance fluctuation from the sample. So we conclude that the advantage of having the runner go from first to second is almost exactly canceled by the disadvantage of the additional out that typically occurs on a bunt play.

Conclusion. On the average, bunting with a man on first loses a lot of runs. On the average, it doesn't increase the probability of scoring at least one run in the inning. Here we've assumed that the batter is always out at first, but, of course, he is sometimes safe, thereby increasing the efficacy of bunting. It is probably more often true, however, that the front runner is thrown out at second, a disaster to the team that chose to bunt. It would appear that bunting with a man on first early in the game should be done only when it so takes the defense by surprise that the chance of the batter's being safe is substantial. Even late in a tight game there is no visible advantage to such bunting unless special circumstances prevail.

Now let's think of the problem of the man on second with nobody out. The table tells us that he will score (or at least somebody will score) in all but 381 cases out of 1,000, that is, in 619 cases out of 1,000. *If* we can move him to third by sacrificing the batter, we can raise the 619 to 693. (Note that we lose 214 runs per 1,000 tries doing this, but let's again consider the case where it is late in the game and we need only one run.) Here there is indeed something to be gained by a successful bunt play, but it's time to face reality: the bunt play doesn't always work. How often it works depends on a lot of things, and we don't have statistics for an average result, but let's see how we would use them if we did.

If the batter bunts the ball a little too hard, the defending team happily fires the ball to third base and the lead runner is put out, leaving the offensive team with a man on first and one out, their probability of scoring at least one run having gone from 0.619 to 0.266, the latter figure coming from the complete table in Lindsey's paper. The typical manager does not admit the possibility of such an event. After it happens he dismisses it with the remark "These young fellows don't know how to bunt like we used to." I know this remark was being made before any of today's managers were making their first appearance as professional players, and it probably originated in the nineteenth century by the first nonplaying manager. The remark is merely an excuse for not studying the problem, but let's not be too hard on baseball managers; we have pointed out already that the moves they make in plain sight are duplicated by other kinds of supervisors in less visible circumstances.

As I said above, we don't have statistics for the results of a bunt try with a man on second, so I'll make up some, trying to be as realistic as possible from unrecorded personal observations over the years. Here they are:

1. 65% of the time the runner moves to third, and the batter is out (normal case).
2. 12% of the time the runner is put out at third, and the batter is safe at first.

3. 10% of the time the runner must stay at second, and the batter is out; for example, when the batter bunts a pop fly or strikes out.
4. 8% of the time the batter gets on first safely, that is, he gets a hit, and the runner also advances.
5. 5% of the time the bunter hits into a double play, that is, he and the runner are both thrown out.

Now to compute the overall probability of scoring at least one run, we simply multiply and add according to the rules of probability. If you don't know these rules, do it this way: start with 1,000 cases. In 650 (that is, 65%) of these we have result 1, namely, a man on third and one out. The table says that he will score 69.3% of the time, so we take 69.3% of 650, and get 450. That is, in 450 cases the bunt succeeds as in 1, and a run ultimately scores. Now in 120 cases the outcome is as in 2, and Lindsey's complete table says that a score then occurs 26.6% of the time. So we take 26.6% of 120 and get 32. Add this to the 450 and keep going. What we get for all five cases, using Lindsey's complete table where necessary, is

$$0.693(650) + 0.266(120) + 0.390(100) + 0.87(80) + 0.067(50)$$
$$= 450 + 32 + 39 + 70 + 3$$
$$= 594.$$

In other words, we will get at least one run in only 594 cases out of 1,000. Before the bunt our chances were 619 out of 1,000, so we have shot ourselves down. Of course, if our hypothetical data in 1 to 5 above are too pessimistic, the correct result will be a little more favorable to bunting, but it would appear that any realistic estimates will lead to the conclusion that bunting is not profitable on the average.

The intelligent use of statistics requires more than just a look at the averages. The above data and accompanying arguments show that bunting, used indiscriminately as many managers do, is not a winning strategy. This doesn't mean that one should never bunt, however. The man at bat may be a weak hitter who is an excellent bunter, and the man following him may be a good hitter; the batter may be a pitcher whose hitting ability is nil, but who can occasionally put down a good bunt; or the other team may clearly not be expecting a bunt, so that the element of surprise is on our side to help the bunt become a base hit. In any of these cases the bunt can be a profitable action. The role of statistics is to show us what our average behavior should be. In general, if the average result of a strategy is very good, we should use it pretty often. If the average result is poor, we should use it sparingly, that is, the special circumstances that lead to it should be very, very special. There are those who say that statistics are irrelevant and that they treat every case as a special case. This is probably impossible, and if such people would examine their behavior over a long period of time, they would probably find it quite statistically predictable. Incidentally, if one takes the point of view that surprise is the whole thing, that is, that the objective is to be unpredictable, then a randomized strategy is indicated; this is elaborated in any book on mathematical game theory.

THE STRATEGY OF THE INTENTIONAL WALK

Another strategic move in baseball is the intentional base on balls. The opposing team has a man on second, say, with one out, and we decide to put a man on first intentionally, either to try to get a double play or to have a force play available at all three bases. Is it worth it? Lindsey's table shows that before the intentional walk the probability of scoring at least one run is 0.390, but afterward it is 0.429. Clearly, on the average, the intentional walk is a losing move; followed by a double play it's great, but followed by an unintentional walk it can lead to a calamity, and the latter possibility is part of the reason for the numerical results just quoted. Widespread use of the intentional walk seems to be based on sheer optimism, as the statistics appear to show that the bad effects, from the point of view of the team in the field, definitely outweigh the good ones, on the average. What about special cases? If the batter is a good one, to be followed by a poor one, then the data don't necessarily apply, and the intentional walk may be a good thing. It probably should seldom be used early in the game, though, unless the following batter is a weak-hitting pitcher because it causes the average number of runs to go up from 671 per 1,000 to 939 per 1,000 owing to the additional base runner, and it is doubtful that there are many special cases that are so special as to outweigh this fact. At the end of the game, the data of Table 1 may be too optimistic in favor of the intentional walk for this reason: great additional pressure is placed on a pitcher who gives an intentional walk to fill the bases with the potential game-ending run on third, and this pressure seems to produce an increase in the frequency of wild pitches, hit batsmen, and unintentional walks.

COLLECTION AND USE OF DATA

Figures such as those in Table 1 are obviously of little value unless they are based on a rather large number of cases. It isn't at all obvious, though, how large the number of cases should be. Mathematical statistics answers questions about how large sample sizes should be, but the questions must be specific. We can't, as we are sometimes asked to do, say that 100 (or 1,000 or 2,000) is a good all-around sample size. If, however, we are asked to find the probability of at least one run resulting from a man on first with no outs, we can, with certain reasonable simplifying assumptions, determine how large the sample must be so that we can be 90% sure, for example, of being within 0.005 of the correct answer. Table 1 shows in the last column the sample size that was used to produce the data of the earlier columns. For an individual keeping records as a pastime, this represents a major effort. We would think that baseball people, engaged in a competition in which a few extra victories can make a difference of a great deal of money, would go to the trouble to collect even larger samples. They wouldn't want to go too far in this direction, however, because information tends to become obsolete. Changing rules, playing fields, and personnel cause the game to change slightly from year to year. Sometimes scoring is relatively low for a few years, and then it increases for a few years. Data gathered in one of these periods of time may not be altogether valid as a basis for decisions in another.

Data of the sort we have been talking about here are sometimes called *historical* as opposed to *experimental,* or *controlled.* The distinction is important in many areas. For example, if statistics are produced showing that smokers have lung cancer with much higher frequency than nonsmokers, this *historical* fact *in itself* does not demonstrate that smoking increases the lung cancer rate. (After all, children drink more milk than adults, but this is not why they are children.) The problem is that there may be other variables that, for example, help cause lung cancer and also influence people to become smokers. Nevertheless, the historical statistics on cancer were very suggestive and led to various experiments in laboratories that have strengthened most people's belief in a causal relationship. We can make good use of historical data, in other words, but we must be careful about inferring cause-and-effect relationships from them.

No doubt because of frustrations in trying to draw conclusions from historical data, statisticians developed the science and art called the *design of experiments.* If we can do a properly designed experiment, we are in a much better position to draw valid conclusions about what causes what, but the possibility of a designed experiment is not always open to us. When we can't experiment, we must do what we can with available data, but this doesn't mean that we shouldn't keep our eyes open to the faults that such data have.

CONCLUSIONS

So what have we learned from our look at sports statistics? We have learned these do's and don'ts:

1. Don't waste time arguing about the merits or demerits of something if you can gather some statistics that will answer the question realistically.
2. If you're trying to establish cause-and-effect relationships, do try to do so with a properly designed experiment.
3. If you can't have an experiment, do the best you can with whatever data you can gather, but do be very skeptical of historical data and subject them to all the logical tests you can think of.
4. Do remember that your personal experience is merely a hodgepodge of statistics, consisting of those cases that you happen to remember. Because these are necessarily small in number and because your memory may be biased toward one result or another, your experience may be far less dependable than a good set of statistics. (The bias mentioned here can come, for

instance, from the fact that people who believe in the bunt tend to remember the cases when it works, and vice versa.)
5. Do keep in mind, though, that the statistics of the kind discussed here are averages, and special cases may demand special action. This is not an excuse for following your hunches at all times, but it does mean that 100% application of what is best on the average may not be a productive strategy. The good manager has a policy, perhaps based on statistics, that takes care of most decisions. The excellent manager has learned to recognize occasional situations in which the policy needs to be varied for maximum effectiveness.

Since this article was written, computers have invaded sports just as they have many other fields of activity, and greatly increased use of statistics has followed. Consequently, some of my remarks about the use of statistics, particularly in baseball, are no longer true as stated. Also, because of changes in the game ranging from new ballparks to rule changes such as the introduction of the designated hitter, the 1959–1960 data of Table 1 may be obsolete for one of the major leagues. All this means that the examples are out-of-date while the general principles remain true. In fact, a new similarity between sports management and business management has been added, namely, that the increased use of statistics has not necessarily resulted in more intelligent use of statistics. Since my retirement in 1979 I have not followed major league baseball at all closely, but from watching postseason games I've concluded that few, if any, managers have used the increase in available knowledge to improve their strategies.

PROBLEMS

1. Refer to Table 1. In how many cases with a man on third and one out, did no runs score?
2. Suppose second base is occupied and there are either no outs or one out. In how many of such cases in Table 1 are no runs scored in the inning?
3. Suppose there are runners on first and second, no outs, and it is early in the game. Assuming the batter will be out and the runners advance one base, do the figures in Table 1 suggest a bunt? Explain your answer.
4. Suppose there are runners on first and second, no outs, and it is the last inning of a tight game. Assuming the batter will be out and the runners advance one base, do the figures in Table 1 suggest a bunt? Explain your answer.
5. Suppose the statistics for the results of a bunt try with a man on second are 70%, 13%, 9%, and 3%, respectively, instead of 65%, 12%, 10%, 8%, and 5% assumed by the author. Would bunting then be profitable on the average in this situation? Explain your answer.

6. When might a sacrifice bunt be a wise move in a situation where, on the average, it is not?
7. a. Distinguish between *historical* and *experimental* data.
   b. Why didn't Lindsey conduct a controlled experiment?
8. Use the following additional statistics from Lindsey and the outcome percentages given in the text. Assume there is a man on second and one out. The batter attempts a bunt.

Base Occupied	*No. of Outs*	*Probability That No Runs Score in the Inning*
1	2	.886
2	2	.788
3	2	.738
1, 3	1	.367

   a. How many times (out of 1,000 cases) will at least one run score?
   b. How does possibility 5 (bunter hits into double play) enter your calculation?

REFERENCES

E. Cook. 1966. *Percentage Baseball*. Cambridge, Mass.: MIT Press.

R. Hooke. 1967. Review of Cook (1966). *Journal of the American Statistical Association* 62:688–690.

R. Hooke. 1983. *How to Tell the Liars from the Statisticians*. New York: Marcel Dekker. (See especially sections 18, 25, 28, and 61.)

G. R. Lindsey. 1963. "An Investigation of Strategies in Baseball." *Operations Research* 11:477–501.

J. Thom and P. Palmer, with D. Reuther. 1985. *The Hidden Game of Baseball*. Garden City, N.Y.: Doubleday.