What's so bad about leaving runners on base?

Brendan Bingham

How often have you heard it? A baseball announcer wraps up the day's broadcast with words such as, "The home team just couldn't come through in the clutch today, leaving seven men on base in the last four innings; that was the difference in this 4-to-2 loss." As a young fan and frequent listener to game broadcasts, I took this sort of commentary to mean that leaving runners on base is a bad thing, not merely a source of frustration for the devoted fan, but a revelation of the team's weakness, its inability to seize opportunity. And narrow losses are the penalty for the sin of leaving runners on base. In Game 6 of the 1986 World Series, for example, the Red Sox left 14 runners on base. This fact is often cited as evidence that much more went wrong for the Sox in that game than one misplayed ground ball. If each runner left on base (LOB) is a run unfulfilled, then the simple-minded expectation, in statistical terms, is that there should be negative associations firstly between LOB and runs and secondly between LOB and wins.

Of course, one could offer the competing hypothesis that in the "get 'em on, get 'em over, get 'em in" world of baseball, the more you get on, the more you get in. But because every inning must end, the more you get on, the more you leave on. From this perspective, one would predict the LOB-runs and LOB-wins associations to be positive.

Relying on MLB boxscores available through www.retrosheet.org, I set out to assess the correlation between LOB and runs scored. For each game played in May of 2007, I tabulated the identities of the home and away teams, how many runs each scored, and how many runners each team left on base. I then calculated the numerical relationship between LOB and runs (the Pearson product-moment correlation) and the categorical relationship between leaving runners on base and winning.

Why May of 2007? I was interested in examining a large enough sampling of games (at least a few hundred) that meaningful patterns could emerge. However, because I was doing data entry by hand, I was not motivated to look at a sampling of thousands of games. Perhaps motivated by laziness, I convinced myself that I did not wish to chase any patterns that could only be seen upon examination of years and years worth of data. So, a month's sampling seemed sufficient. I chose the month of May not completely arbitrarily. May lacks the snowstorms of April, the payroll firesales of July, and the rookie call-ups of September. By May, teams have played enough games to know themselves, their strengths, their weaknesses, and the game strategies and tactics that best suit their talents. Moreover, no matter how well or poorly a team has done in April, come May it is still in the pennant race. I don't know whether any of these nebulous factors would have any effect on the relationship between runs and LOB; I will leave that for someone more diligent to discern.

Why track home and away teams separately? While there is no theoretical limit to how many runs a team can score in an inning, LOB is constrained. A team can leave no more than three runners on base in an inning, 27 in a nine-inning game. In most games won by

the home team, the bottom of the ninth inning is not played, giving the home team fewer opportunities to leave runners on. In contrast, in all games won by the visitors, both teams have an equal opportunity to leave runners on base. Therefore, it seemed that any inherent relationship between runs scored and LOB might more clearly emerge in games won by the visiting team. It should be noted, however, that in tabulating wins, runs, and LOB, I made no attempt to denote games won in the bottom of the ninth, games that went to extra innings, or games that were rain-shortened. I was willing to trust that any robust patterns concerning the relationship between runs and LOB would shine through whatever influences shorter and longer games could impose. Again, anyone with the motivation to look deeper is welcome to parse out such factors.

To the question of the relationship between runs and LOB, as summarized in Table 1, in May of 2007, MLB teams played 419 games, of which the home team won 230 (or 55%). Despite this home team advantage, home and away teams on average scored practically the same number of runs per game, 4.65 and 4.63, respectively. Home teams averaged 6.84 LOB per game, while visiting teams averaged 7.19, but remember that losing visiting teams often have one more inning in which to produce LOB. Of greater importance to the question at hand, there was a small but statistically significant positive correlation (r) between runs scored and LOB. This correlation was qualitatively similar for home teams and visiting teams, 0.215 and 0.143, respectively, and highly statistically significant in both cases. In many areas of inquiry in which correlation statistics are applied, correlations on the order of 0.2 are considered unimpressive or even negligible, even when the sample size is large enough for such correlations to achieve statistical significance. While I do not strictly subscribe to this point of view, the point can be illustrated if you consider that the predictive value of a correlation is dependent not on the correlation coefficient (r), but on the square of this quantity (r^2, sometimes called the coefficient of determination). For all games in May 2007, r = .178, so r^2 = .032. What this tells us is that approximately .03 (or 3%) of the variance in runs scored is "explained" by the variance in LOB. Put another way, if I am asked to deduce the number of runs a team scored in a given game in May of 2007 and I am told how many runners the team left on base, I would have only a 3% advantage in "guessing" the number of runs scored compared to the situation in which I'm provided no information at all. The importance of the observed correlation between runs and LOB in this dataset is not that it is impressive or predictive – it is neither – but that it is not a negative quantity, as the fictionalized baseball announcer above might have had us believe. Rather, the data for games in May of 2007 offer no evidence that a team is "punished" for the "weakness" of leaving runners on base.

I also sought to know the stability and robustness of the pattern described above. That is, does LOB correlate with scoring in other years, both within the same modern era of baseball and across eras? To address this question, I chose three additional years, from the three most recent recognized eras, from which to tabulate runs and LOB: 2000 (Long Ball Era), 1985 (Free Agency Era), and 1968 (Expansion Era). Again, only games from the month of May were included in the analysis. The year 2000 was chosen as a second example of the current era of the game, but sufficiently removed in time from 2007 that there would be only modest similarity in the population of players making up the major

leagues (about 25 or 30%, I'm guestimating). Table 1 also summarizes the LOB-to-runs correlation for these three additional months of game data. While differences among these eras can clearly be seen in the number of runs scored, ramping from approximately three per team per game in 1968 to four in 1985 to five in 2000, the differences in LOB across eras are minimal (ranging from 6.5 to 7.4 per team per game), but again, LOB has bounds (no more than three per inning). As in the May 2007 games, all correlations between LOB and runs for games played in May of 2000, 1985 and 1968 are positive and small (ranging from .101 to .271), and all but one (visiting teams in 1968) are statistically significant. All in all, the pattern of correlation between LOB and runs is remarkably similar in these four sets of games sampled from across a four-decade period.

The same dataset can also be analyzed to explore the correspondence between LOB and winning (Figure 1). Of the 419 games played in May of 2007, in slightly more than half (213), the team that left the greater number of runners on base was the winning team. In 13% of the games (55), the winning and losing team left the same number of runners on, and in 36% of the games (151), the losing team left more runners on. These numbers can be further broken down according to which team, home or visitor, won the game. Examining the 230 home team wins and the 189 visiting team wins, in both cases a greater number of games was won by the team that left more runners on base; this pattern was more striking in the subset of games won by the visiting team. The patterns for games in May of 2000, 1985 and 1968 are generally similar (Figure 1). In all three cases, producing more LOB is associated with a greater probability of winning. In May 2000, the numbers were close, 187 games won by the team leaving more runners on, 180 by the team with fewer LOB. Broken down by home victories vs. road victories, there were an equal number of road victories (80 each) by the team with more LOB and the team with fewer LOB. Again, no evidence emerges for a cost paid for leaving runners on base.

In short, leaving runners on base appears, if anything, to be a good thing. Remember that correlation does not imply a causal relationship. So while teams never set out to leave runners on base (nor should they), the runners they leave on can be seen as a residue of having successfully put runners on, and putting runners on base is the establishment upon which scoring (and winning) is built.

In an effort to vindicate the baseball announcers who are so quick to point up the negative side of leaving runners on base, I then asked whether the patterns depicted in Table 1 and Figure 1 are being unduly driven by lopsided games. As any baseball fan knows, a team's fortunes include four broad categories of game outcomes: narrow wins, narrow losses, one-sided wins, and one-sided losses. Perhaps the pattern is being dominated by one-sided games, such as the 8-0 or 10-2 affairs in which the winning team puts runners on base every inning and the losing team repeatedly goes three-up-and-three-down. Perhaps in closer games there might be no correspondence (or even a reversed pattern) between LOB and winning. To test this proposition, I broke down the analysis to two categories: games decided by two runs or less, and games decided by three runs or more (Figure 2). Remarkably, the dichotomy between close games and not-so-close games, as defined here, is not at all dramatic. In two years, 2007 and 2000, the close games show a shift in the direction of reversing the overall trend; in 2000, this shift results in a

marginally greater winning percentage by the team with fewer LOB. On balance, however, this is a small effect, and there is no evidence of such a shift in the datasets from May 1985 or May 1968. Again, these data offer no evidence for the existence of a cost (in terms of winning) associated with leaving runners on. There is little more than a hint that one-sided games contribute more heavily to the overall pattern than close games. Even in close games, the team with more LOB is more likely to win.

Here are two final points that emerge from this analysis of LOB statistics. First, games in which a team leaves no runners on base are extremely rare. Of the more than 1400 games considered herein, in only five (about one third of one percent) did one of the teams fail to leave a runner on base. One of these games was Jim "Catfish" Hunter's (Oakland) perfect game against Minnesota (May 8, 1968); all perfect games result in no LOB for the losing team, but perfect games are rare. In two other games the team allowing no LOB was the winner: a Chien-Ming Wang (NYY) performance (May 5, 2007) in which he allowed no Seattle baserunners through seven innings, followed by a solo HR, a single, and a double play in the eighth inning; and a May 17, 1985, Bret Sabehagan (KC) 2-hitter against the Brewers, in which both Milwaukee players who hit safely were eliminated in double plays. Interestingly, in two other games, the team leaving no runners on base went on to win. On May 31, 2007, Toronto beat the White Sox 2-0 on the strength of two solo homers. Finally, in the most involved affair, Milwaukee beat Minnesota 5-3 on May 26, 1985. Milwaukee scored in two innings, and put a runner on base in another. In all, they had six hits and a walk with five runners scoring, one runner erased on a double play, and one caught stealing. While there are many ways to leave the bases empty – scoring on sacrifice flies, ground-outs and wild pitches can be added to the list – it's apparently very rare to do so nine innings in a row. Second, in assembling and analyzing this dataset, I had the nagging fear that there was some artificial force at work constraining the correlations in Table 1 to be positive. To address this issue, I also examined the correlation between the visiting team's runs scored and home team's LOB (and vice-versa), correlations expected to be zero, or close to it. If such correlations were found to be of the same order of magnitude (0.2) as same-team LOB vs. runs scored, then the interpretation of the correlation coefficients in Table 1 could be questioned. To my comfort, most of these correlations were near-zero and not statistically significant. For example, for the May 2007 dataset, the correlation between LOB (away team) and runs (home team) is .0542 ($p = .2685$), and between LOB (home team) and runs (away team) is -.0128 ($p = .7945$).

Table 1. Summary of games played in May 2007, 2000, 1985, and 1968, including wins, winning percentage, runs per game, and LOB per game for the competing teams. The Pearson product-moment correlation (r) is calculated between a team's runs scored and LOB, and the probability (p) associated with that correlation is also reported.

Year	Team	Games	W	Win Pct	Runs/game	LOB/game	r	p
2007	Home	419	230	.549	4.65	6.84	0.215	<0.0001
	Away	419	189	.451	4.63	7.19	0.143	0.0033
	Total	838	419	.500	9.28	14.02	0.178	<0.0001
2000	Home	405	227	.560	5.44	7.19	0.116	0.0200
	Away	405	178	.440	5.06	7.37	0.268	<0.0001
	Total	810	405	.500	10.50	14.56	0.187	<0.0001
1985	Home	336	184	.548	4.21	6.74	0.114	0.0364
	Away	336	152	.452	4.22	6.99	0.271	<0.0001
	Total	672	336	.500	8.43	13.73	0.197	<0.0001
1968	Home	282	141	.505	3.11	6.52	0.220	0.0002
	Away	282	138	.495	3.32	6.67	0.101	0.0911
	Total	564	279*	.500	6.43	13.19	0.152	0.0003

*In May 1968, three tie games were suspended and replayed in their entirety

Figure 1

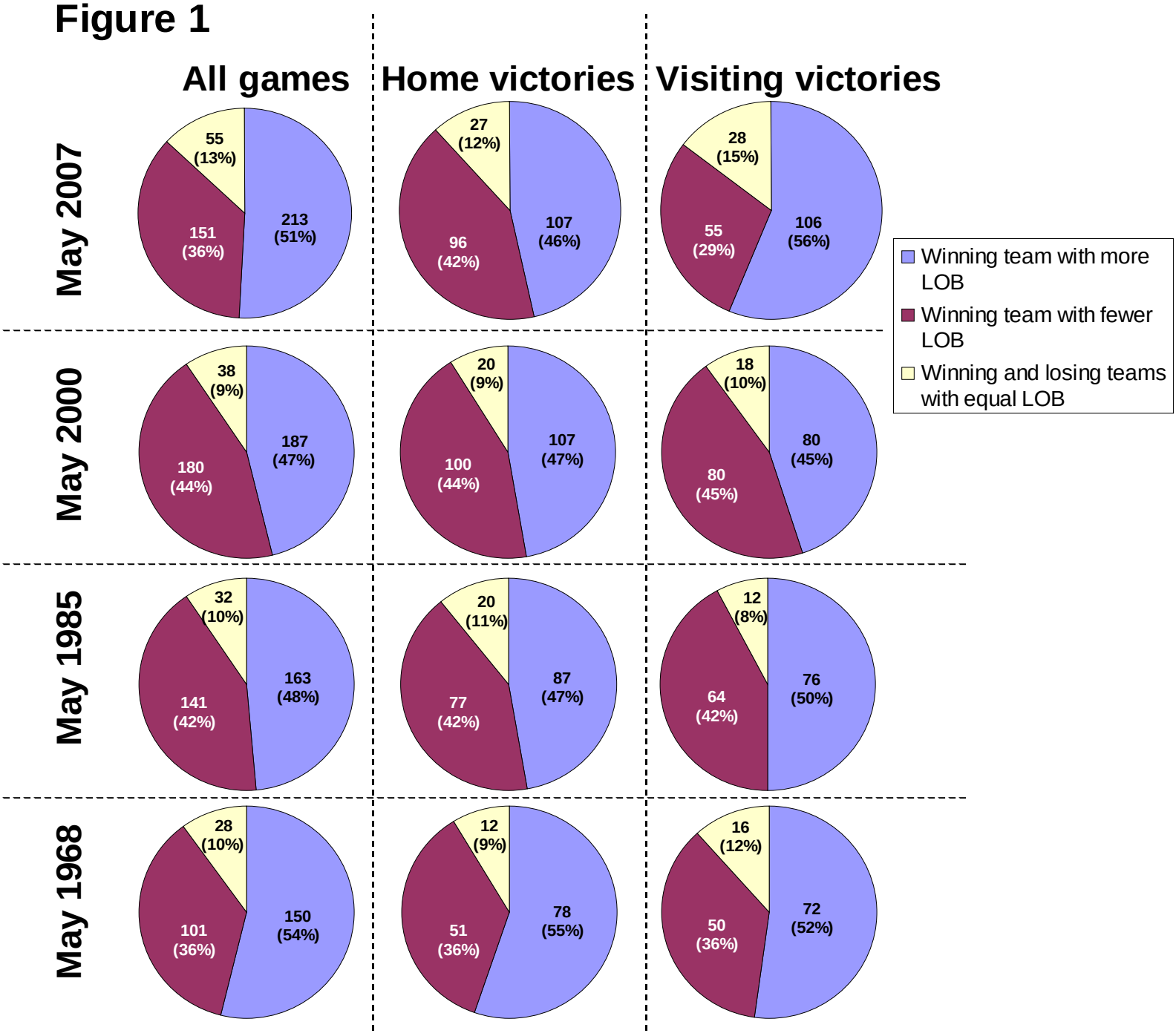

Figure 1. The correspondence between winning and leaving the greater number of runners on base for games played in May of 2007, 2000, 1985 and 1968. All games (first column) are subdivided into games that were won by the home team (second column) and games won by the visiting team (third column).

Figure 2

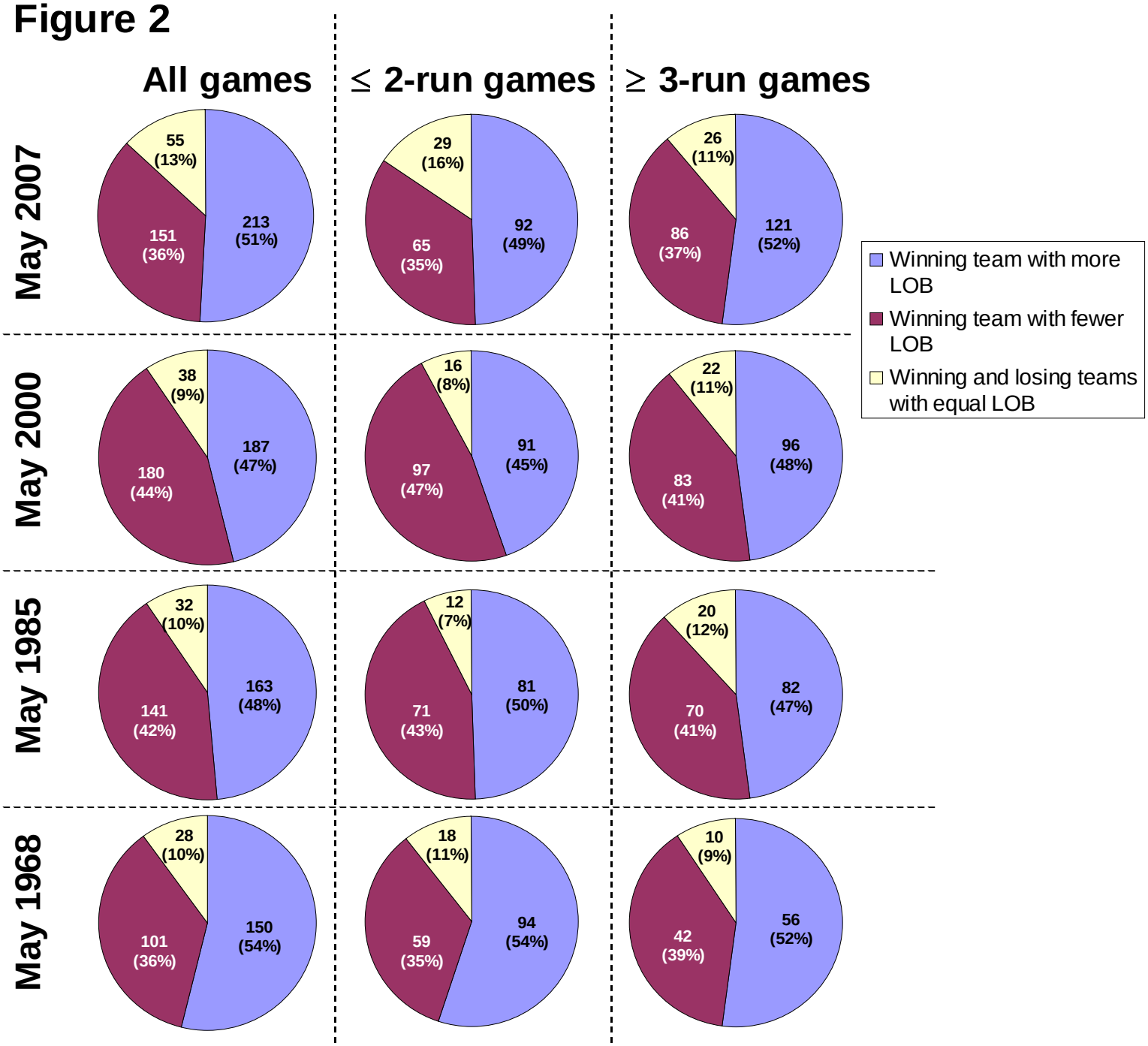

Figure 2. The correspondence between winning and leaving the greater number of runners on base for games played in May of 2007, 2000, 1985 and 1968. All games (first column; same as first column in Figure 1) are subdivided into games that were decided by two runs or less (second column) and games decided by three runs or more (third column).

Summary:

Traditional baseball lore, or at least the rhetoric put forth by many baseball announcers, holds that leaving runners on base foretells a team's ill fortune. Under such a point of view, each runner left on base is a lost opportunity to score. Do this more than a few times a game and the costs will add up to a loss for your team. In contrast to this expectation, an examination of the data for runners left on base (LOB) and runs scored in games played in the month of May of four MLB seasons (2007, 2000, 1985, and 1968) reveals a consistently positive correlation between LOB and runs scored. Although the magnitude of the correlation is modest (on the order of less than 0.2), when considering a sufficient number of games (such as the hundreds played during a month of MLB action), the correlation is highly statistically significant. Moreover, the pattern is qualitatively similar for home and visiting teams and is robust across eras. This positive correlation between LOB and runs translates into a positive correspondence between LOB and winning. Although the margins are modest, leaving more runners on base is more often associated with winning than losing. This pattern is also robust across eras (although the pattern is weakest in the May 2000 dataset). Moreover, this pattern is not overtly dominated by one-sided games. Even in games that are decided by a narrow margin (two runs or less), leaving more runners on base is more often associated with winning. In conclusion, leaving runners on is not evidence of failure or lost opportunity. Rather, if examined over a sufficiently large number of games, leaving runners on base is evidence of the ability to put runners on base. In the "get 'em on, get 'em over, get 'em in" world of baseball, this ability forebodes scoring runs and winning baseball games.