The 7th Inning Is The Key
By David W. Smith
Presented June 30, 2017
SABR47, New York, New York

It is now nearly universal for teams with a 9th inning lead of three runs or fewer (the definition of a save situation since 1975), to bring in a specialist (a "closer") to pitch. This strategy reflects the belief that the chances of winning the game are enhanced by the pitching change. I demonstrated last year that there is in fact no significant improvement in the chance of winning by this reliance on new pitchers for the 9th inning. For those who did not attend my presentation on closers at last year's SABR convention, you can find the complete paper on the Retrosheet site at: (http://retrosheet.org/Research/SmithD/MythOfTheCloser.pdf).

Since so few games are decided in the 9th inning, the logical question is to ask at what point in the game does the winning team take the decisive lead. We know that more runs are scored in the first inning than in any other, especially by the home team. I examined this in detail at the Houston SABR meeting in 2014. That presentation is also on the Retrosheet site at: (http://retrosheet.org/Research/SmithD/WhyDoHomeTeamsScoreSoMuchInTheFirstInning.pdf).
.
Although the lead certainly can change hands several times in a given game, the aggregate pattern is that once a team gets a lead it most often maintains it.

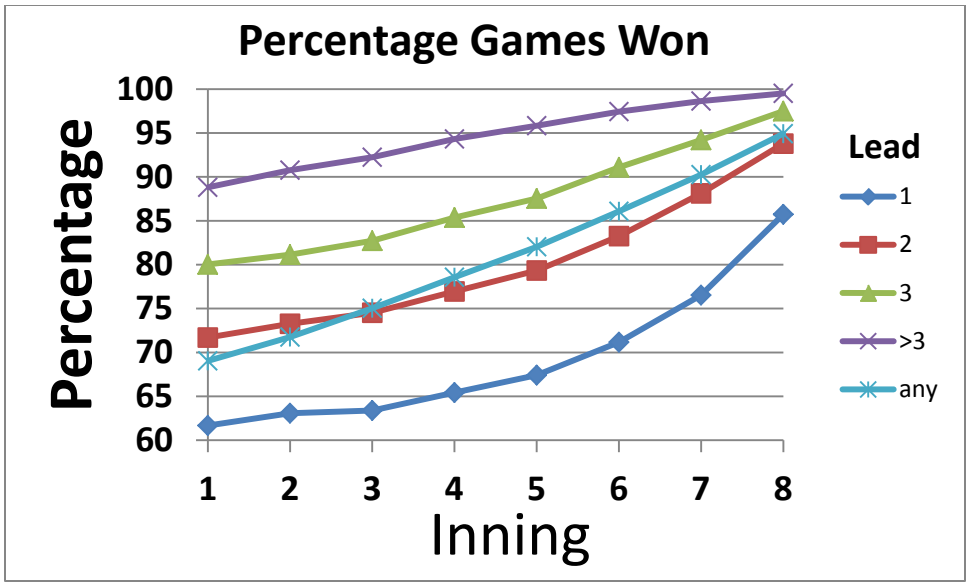
Before we get to the details, I present my usual summary of the size of the database I used. It is:

Table 1. Data used in present study

| | |
|---|---|
| Seasons: | 107 (1910 – 2016) |
| Games | 181, 705 |

As for maintaining the lead, Figure 1 shows a team which is ahead by even one run after the first inning wins over 63% of the time.

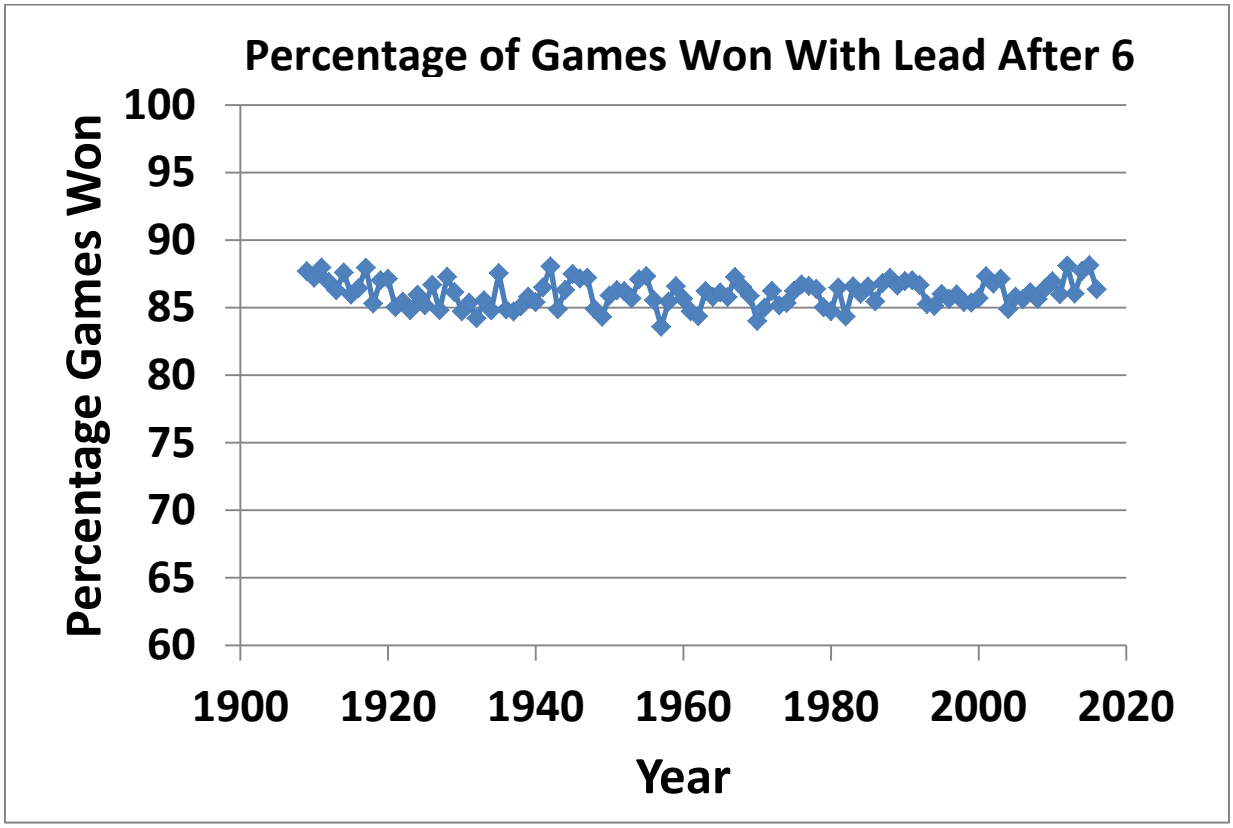
Figure 1. Winning percentage when leading by indicated margin after each inning, 1910-2016.




That percentage increases with each passing inning so that the team with a one-run lead at the end of the 8th wins over 85% of the time. This pattern was found to be remarkably consistent across all 107 years studied even though total scoring changed significantly in different eras over this last century plus, as we shall see in a few minutes. It is reasonable to ask how much the size of the lead matters and that is addressed on the same graph by adding three more lines. These are the probabilities of winning the game when leading by two, three, or more than three runs at the end of the indicated inning. I saw no point in generating separate lines for four, five, six run leads, etc, so they were all combined. As the game proceeds, the value of the lead increases although the rise is not steady, with a proportionally greater value in the last three innings. As the size of the lead increases, so does the translation to a win, with a first inning lead of over three runs resulting in a win 88% of the time. The final line for this figure combines all these different values to give the win probability for a lead of any size. This is a more linear response than the individual leads. Note that after the sixth inning, the team with a lead - no matter what the margin is - wins 86% of the time. It is arbitrary to be sure, but a situation in which the chance of winning exceeds 85% is my candidate for the "deciding point" of a game with a decisive lead, although that may be too ambitious a term.

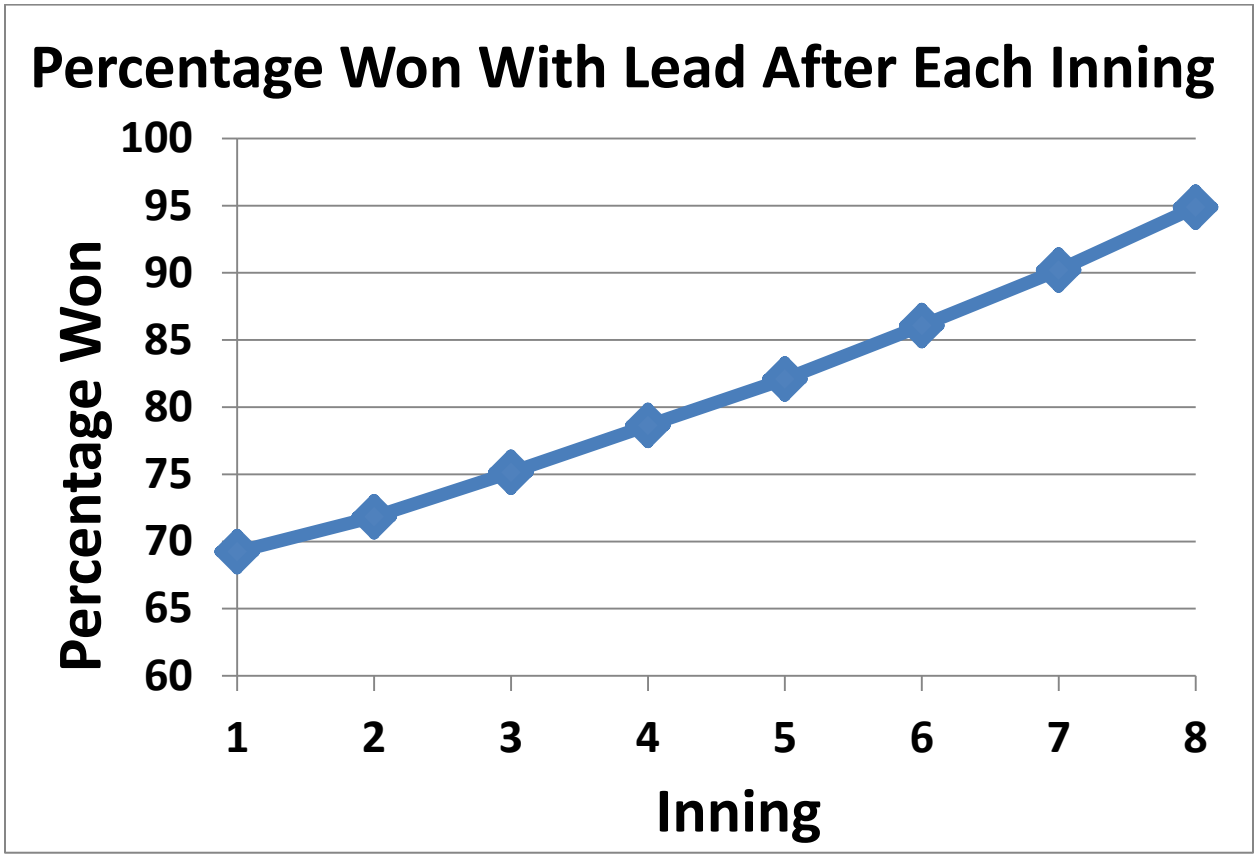
I mentioned that these patterns of lead and probability of winning were constant across the 107 seasons and I address that in the next two figures. Figure 2 shows the probability of winning from 1910 to 2016 when leading by any amount after the sixth inning.

Figure 2.




The results for other innings show the same narrow range of variation. Rather than present eight separate slides, one for each inning, I prepared Figure 3 which shows the average value for each inning across the 107 years I studied. This is the same line as the summary line in Figure 1.

Figure 3.

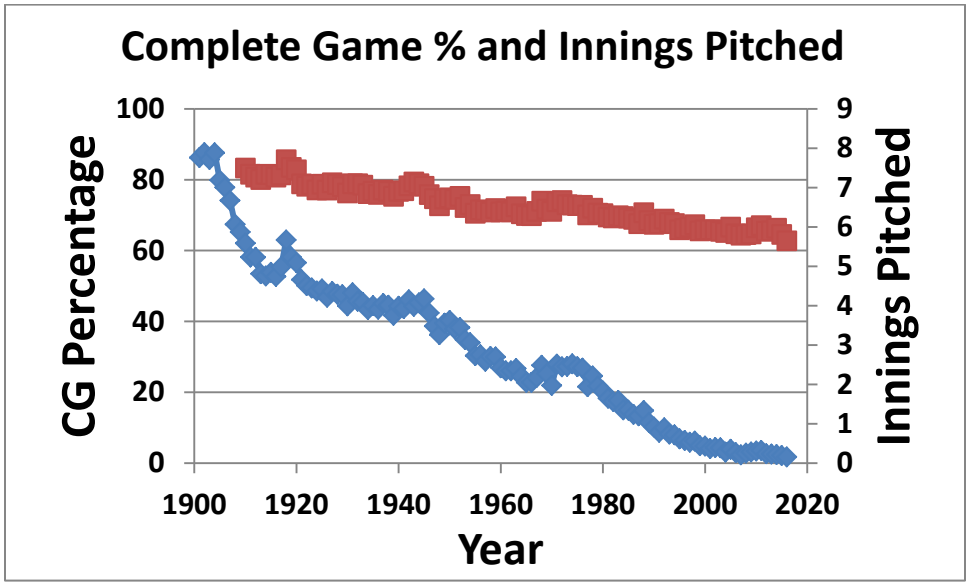



The conclusion is inescapable. A lead is always important, no matter how early it comes and the larger the lead, the more likely it will lead to a win. It is good to remember at this point that these are aggregate values from the more than 181,000 games I examined. I do not claim that any individual game will follow this pattern, although there is still value in this general result.

I will continue the analysis with the end of the sixth inning as my “deciding point”, hence the title of this presentation. My approach is very much concerned with starting pitcher usage and effectiveness, which are two distinct measures. My paper on closers showed a major change in reliever usage starting in 1980 which I will summarize below. Starters are no longer routinely expected to pitch 9 or even 8 innings as bullpens have a variety of "roles" with a general design of using them in sequence after starter has hopefully completed six. This is connected to pitch counts as well, but I want to look at very long term trends and pitch data are only routinely available beginning in 1988.

I will start with the baseline of changes in starting pitcher usage. Figure 4 shows two measures of usage.

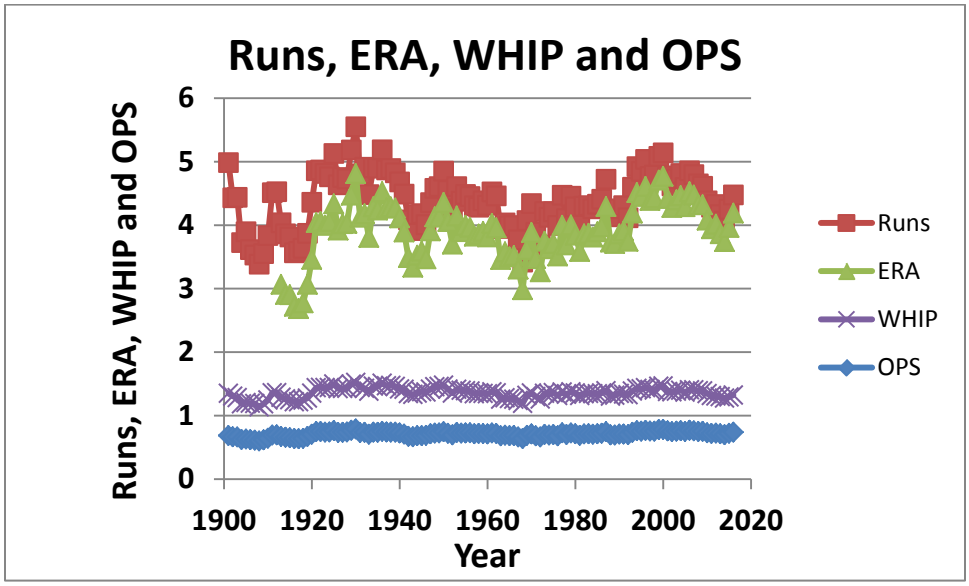
Figure 4.




The first measure is the percentage of complete games and for this line the data go back to 1901, which uses the left-hand vertical axis. The high point was 87.6% in 1901 and the low was 1.7 % in 2016. In fact, last year was the first in history with fewer than 100 complete games, as the two leagues combined for a total of 83 (39 NL, 44 AL). The second measure, on the right-hand vertical axis, is the average innings pitched per start. This also shows a fairly steady decline since 1910, but much more gradual. There are interesting blips in both lines in WWI, WWII, and the years just before the advent of the Designated Hitter. The value of 5.65 innings pitched per start in 2016 is the lowest in history and it has been below 6.0 for 17 of the last 18 years, with only 2011 crossing that line with an average of 6.03. Of course, there is a corresponding increase in relief pitcher innings, but as I have shown before (Figure 9 in http://retrosheet.org/Research/SmithD/MythOfTheCloser.pdf ) there is also a clear pattern of increased number of relief pitchers used in each game, so that the average relief appearance now consists of facing 4.3 batters, while the average starter sees 24.5 batters per start.

In addition to patterns of pitcher usage, we must consider performance. The first way I chose to look at this is ERA. As many people have shown, there is a great deal of annual variation as Figure 5 shows.

Figure 5.

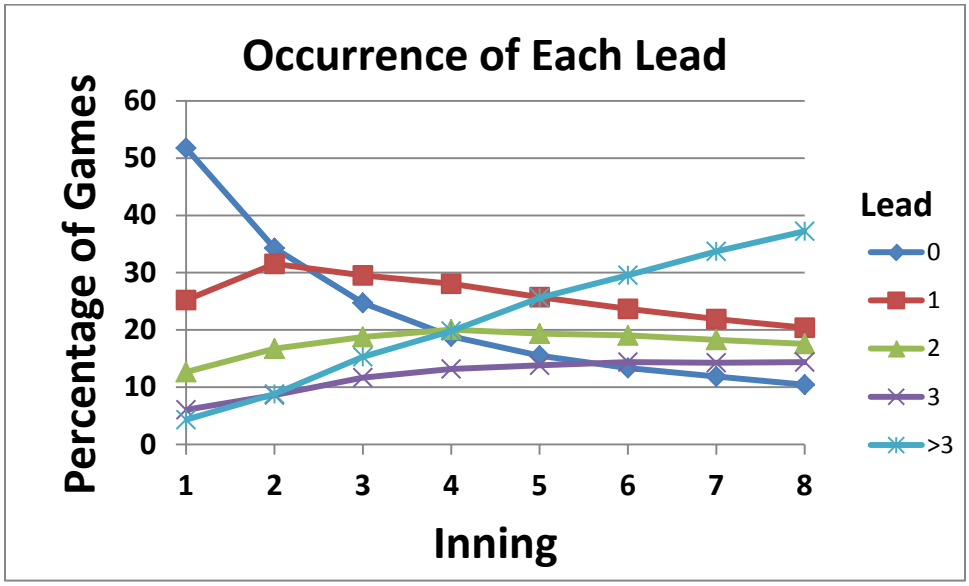



From 1913 (first year that earned runs were officially recorded) through 2016 the range was a high of 5.5 per team per game in 1930 and a low of 2.68, which came in 1917, not in 1968! This pattern reflects total scoring per team as the second line on the graph indicates, but although the lines mirror each other, the difference between the two has changed as unearned runs have become much less common. In 1913 there was just under one unearned run per game per team (0.97) and in 2016, this value was down to 0.29 as errors have become progressively less common. The pattern of ERA changes is interesting in light of the other performance measures I calculated, which are WHIP (walks plus hits per inning) and OPS (on-base average plus slugging average). As the last lines in Figure 5 show, WHIP and OPS have been much more stable than ERA, although there are certainly differences. For WHIP, the high was 1.52 in 1930 and the low was 1.19 in 1968. For OPS, the range is a low of .602 in 1908 and a high of .790 in 1930, to no one's surprise! I find it surprising that scoring has varied so much more than what I like to call the components of scoring over the last century, especially when these component measures have been so stable.

This relation of having a lead to the chance of winning is discussed commonly in broadcasts, usually in terms such as "Our team is 33 and 5 when they lead after six innings". That is a winning percentage of .869 which is certainly impressive, but what is usually not mentioned is what I think is the proper comparison or context, namely that all teams combined have a winning percentage of around .860 when they lead after six. So it is wise to take pronouncements by announcers with a grain of salt.

What I am mostly interested in is what effect pitcher usage has on this relation between having a lead and winning the game. Before I do that, I must digress to another aspect that is not often mentioned. That is, how often do these situations occur? Remember that Figure 1 showed how increasingly important having the lead is, but we need the context of how often each occurs. Figure 6 shows us how often a lead of each size occurs after each inning.

Figure 6.




The first line shows that the most common situation after the first inning is a tie score, which occurs in 52% of all games. Of these games, 93% are 0-0, which is hardly a surprise since most would predict that a scoreless first inning is the most frequent occurrence. The average game quickly leaves the realm of being tied with a steady but nonlinear decline in tied games with a low of 10% being tied after 8 innings, very few of which are still 0-0, of course.

As the ties disappear, what patterns arise? The next three lines present the data for leads of one, two, three, or more than three runs after each inning. Games with a one-run lead have an interesting bump to just over 30% in the second inning, followed by a very steady decline through the rest of the game. Leads of two runs show a gentle increase through the fourth inning and then a very slight drop thereafter. Similarly games with three-run leads show a similar gradual rise, but not much decline. However, the biggest surprise to me occurs with leads of more than 3 runs. A lead of this size is rare after the first inning (4%), but rises very steadily throughout the rest of the game until they reach 37% of all games after the 8th. I suggest that a game with a score differential of more than three runs is not a close game in any meaningful sense and furthermore that after the sixth inning, these large leads collectively predominate and that dominance continues to rise.

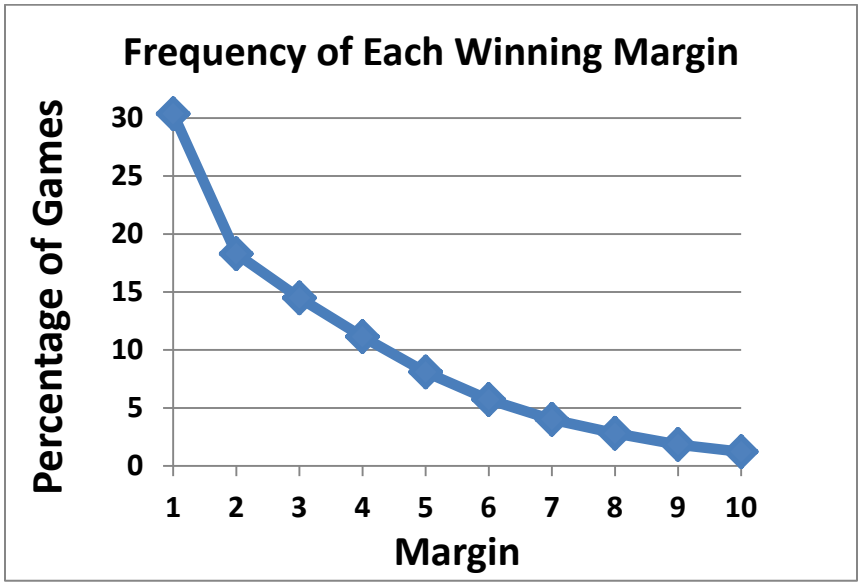
The frequent occurrence of a large lead raises another question, namely what is the average margin of victory? It turns out to be larger than most people suspect. First, Table 2 shows the five most common margins of victory from 1910 to 2016 with the percentage of each score.

Table 2. Most common final scores, 1910-2016.

| Final Score | Occurrences | Percentage |
|---|---|---|
| 3-2 | 10425 | 5.8 |
| 4-3 | 10073 | 5.6 |
| 2-1 | 8371 | 4.7 |
| 5-4 | 7895 | 4.4 |
| 4-2 | 6614 | 3.7 |

There were differences over the eras, but the general pattern holds pretty well. Note that the two most common are very close in frequency and these are the two that vie for the lead in most seasons. These are all close games (one or two runs), but together this group is only 24% of all games. Figure 7 shows the frequency of each winning margin up to 10.

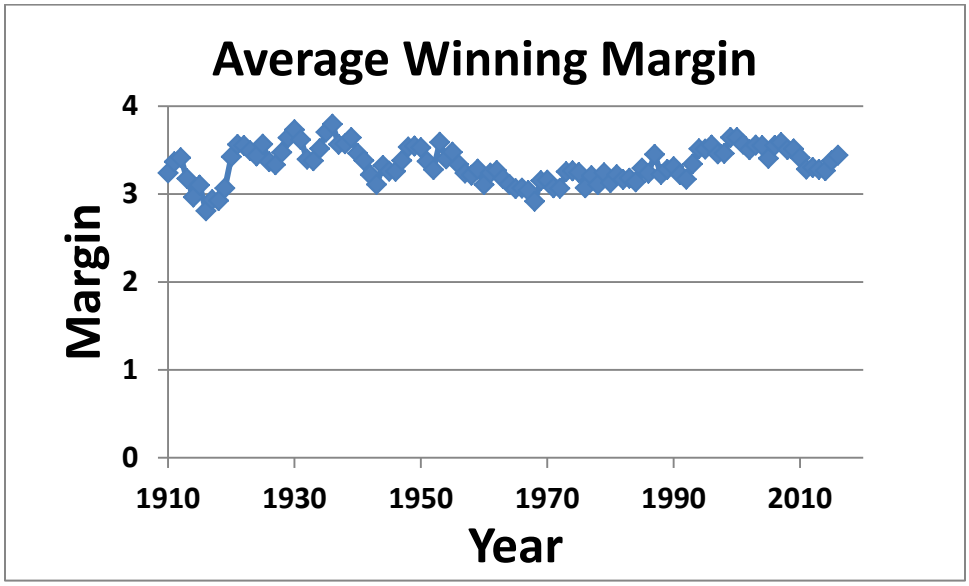
Figure 7.




The clear leader is the one-run win with each larger margin being less. I stopped at 10 runs because almost 98% of all games are decided by 10 runs or fewer, which is hardly a surprise. The games decided by one, two, or three runs are collectively 63% of the total

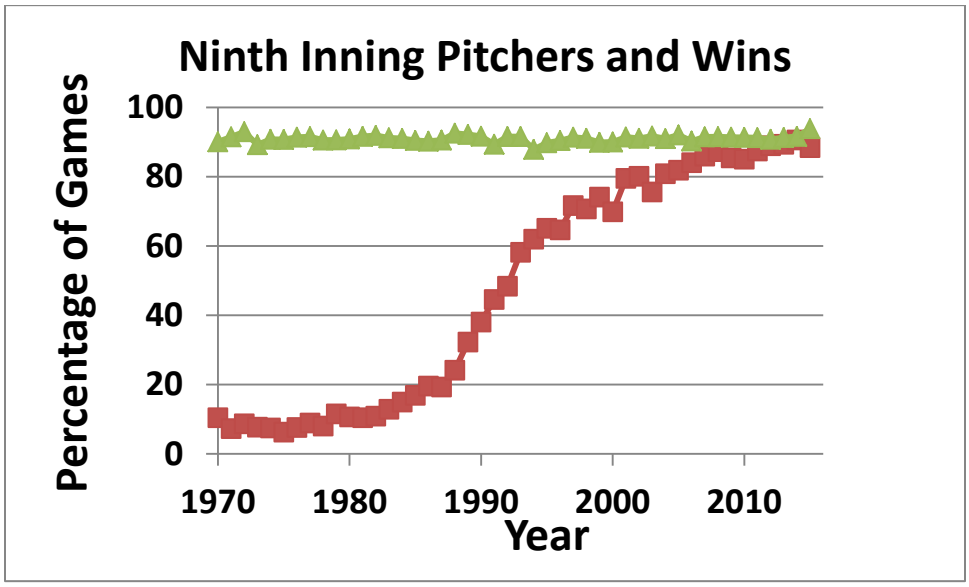
So what is the average margin of victory? In light of the dominance of one-run wins, it may be surprising to see in Figure 8 that although it has varied, it has almost always been over 3, with a grand average of 3.34. We can conclude that it is necessary to be careful not to confuse the most common with the average.

Figure 8.




How does all this fit into pitcher usage? One of the more dramatic findings I had last year was that starting in 1980, there was a surge in bringing in a fresh pitcher to start the ninth inning in a save situation. Figure 9 is an extract of the data from Figure 1 of last year's paper

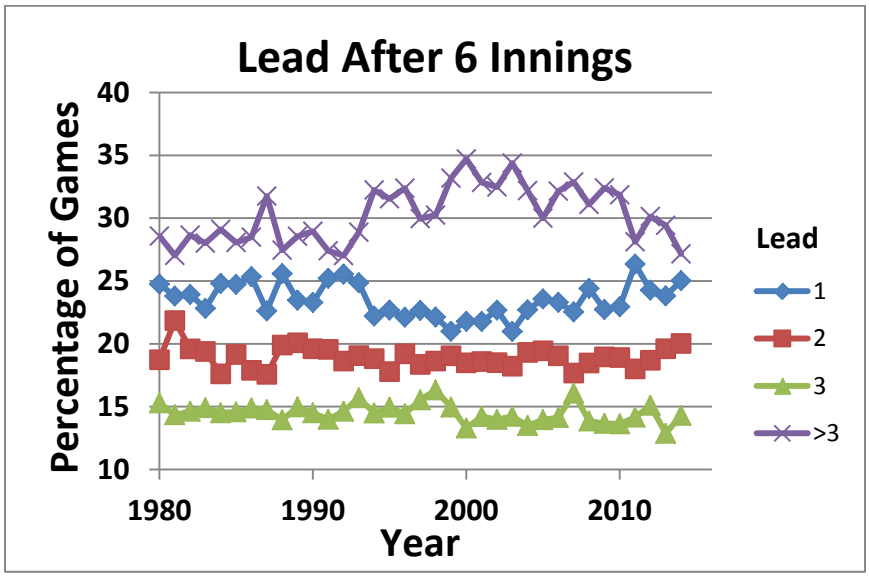
Figure 9.




Beginning in 1980, there was an increase in the entry of new pitchers to start the ninth inning in save situations. However, there is no change in the chance of winning the game, which led me to seriously question the now nearly universal use of closers in these situations (again that is an aggregate statement and does not refer to individual games). Given my emphasis in the current presentation on the significance of the first six innings, I decided to revisit this question of

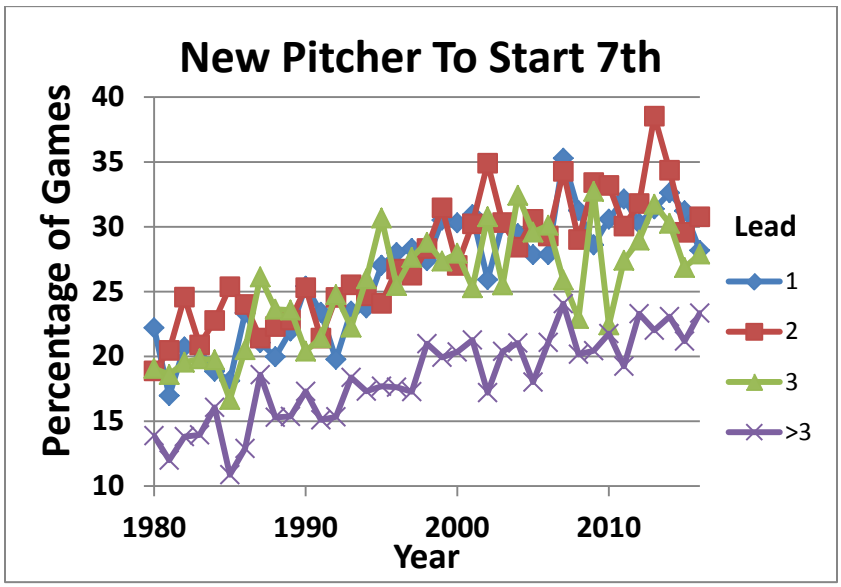
bringing in a new pitcher to start the seventh. Since the change in closer pattern began in 1980, I decided to analyze the seventh inning relievers the same way. These results are in Figures 10, 11, and 12. First I determined how often the different lead situations occurred,

Figure 10. Percentage of games with leads of 1, 2, 3, or more after sixth inning.



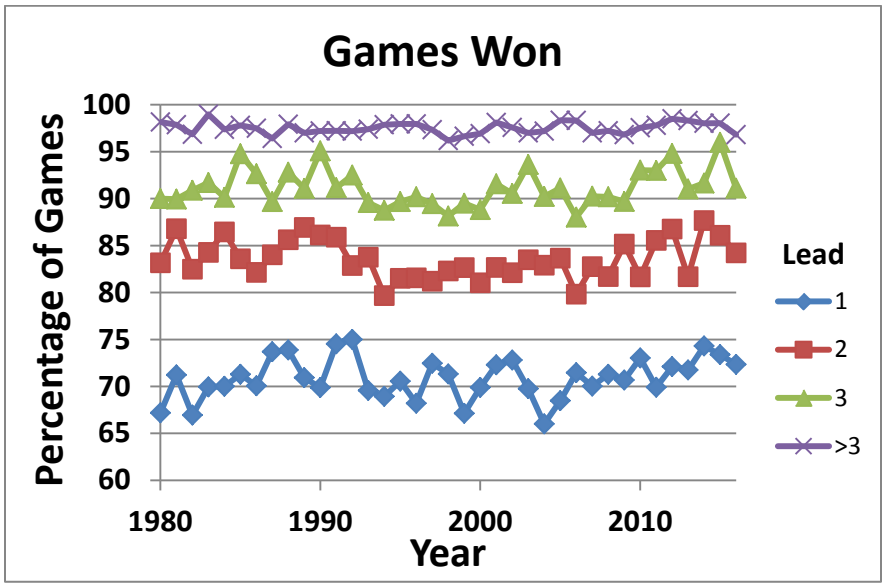

The lines appear in a possibly unexpected order, but remember from before that the large leads predominate at this point in the game. There is quite a bit of statistical noise in these lines but the general pattern shows no clear trends. Figure 11 shows how often a new pitcher was brought in, again as a function of the size of the lead.

Figure 11. Percentage of games with a lead that have new pitcher to start seventh inning.



I divided these entries based on the size of the lead since a direct comparison to the save situation of the ninth inning is not appropriate. There has certainly been an increasing trend of bringing in a reliever in these situations, but the three types of "close" game are similar in pattern, although there is a lot of noise. The fourth line is the result for large leads. The new reliever is called upon less often here, but the shape of the curve is similar, continuing the usage change I first demonstrated with the ninth inning. Finally, of course, we have to ask how effective these substitutions are in securing a win. That is addressed in Figure 12.

Figure 12. Percentage of games won after sixth inning lead of one, two, or three runs.




The results are remarkably similar to the ninth inning "closer" results from my 2016 presentation, namely that these new patterns of usage have had little effect on the final outcome.

Conclusions:

1. The patterns of scoring and winning margin have been remarkably stable across the last 107 seasons.
2. Getting the lead early, preferably a big one, is enormously important.
3. The changing use of starters and therefore relievers, has not mattered very much.
4. The "point of decision", as shown with a win probability of more than 85% is the end of the sixth inning.