OPTIMAL APPROACH FOR INFIELDERS ON STOLEN BASE ATTEMPTS

JONAH LUBIN

SMT DATA CHALLENGE 2023

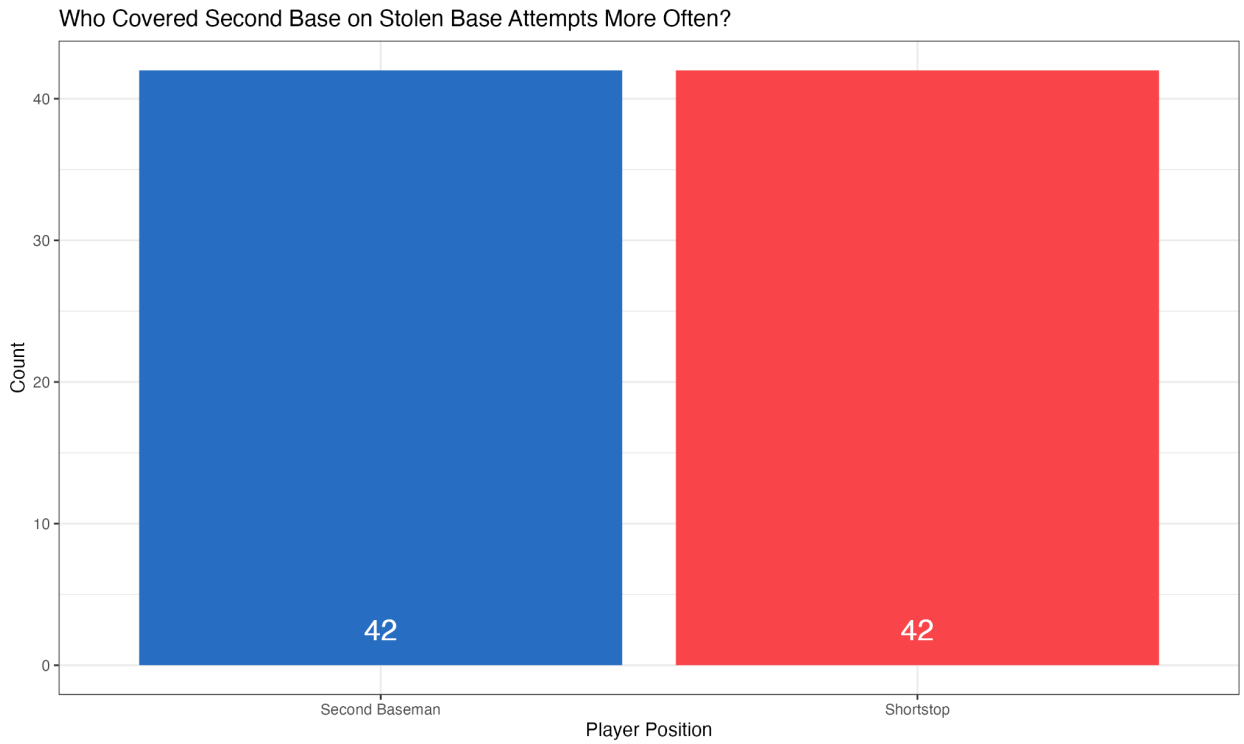
BACKGROUND AND IMPORTANCE

Imagine that you're a shortstop currently playing in the infield. The runner on first base runs to steal second. What should you do? If you have a general knowledge of baseball, you'd know that either the second baseman or the shortstop runs to cover second base in order to attempt to tag the runner out. However, how do you know if you or the second baseman should run, and if it is not you, where should you go? The answers to these questions are known by every middle infielder in professional baseball. Although these players know what to do in this certain situation, they are not necessarily correct, in fact, the wrong approach is taken quite often.

Given baseball is a relatively low scoring game, every run saved matters. When looking at the statistics from the 2022 MLB season, the difference in runs scored per game between the teams with the most and least runs was 1.73, clearly illustrating the importance of saving any amount of runs. When baseball teams take an approach that does not optimize the amount of runs being saved, it is a clear indication that they need to change the approach.

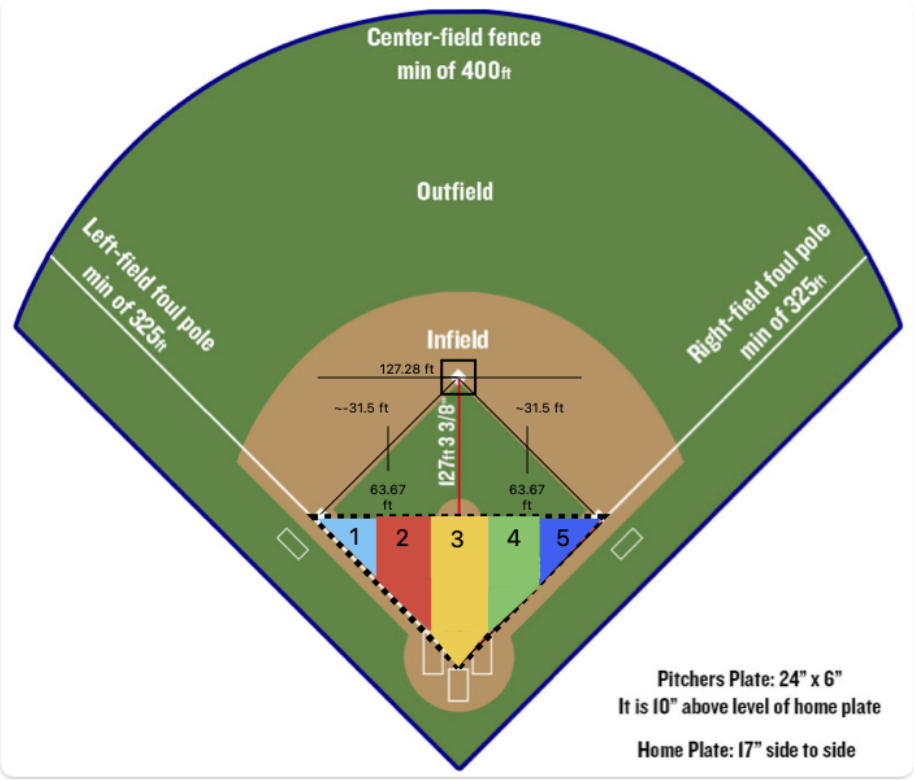
CURRENT APPROACH FOR COVERING SECOND BASE

I first wanted to see if the second baseman or shortstop cover second base more frequently, as well as whether it was relatively the same frequency. After looking at the entirety of the data and plays in which the runner on first attempted to steal second base, it showed that both infield positions covered second base 42 times, which is promising because it shows that there is an equal amount of data for each position, so trying to find an optimal approach is more accomplishable. After determining they each covered the same number of times, I wanted to know when they would decide who exactly would cover second base.



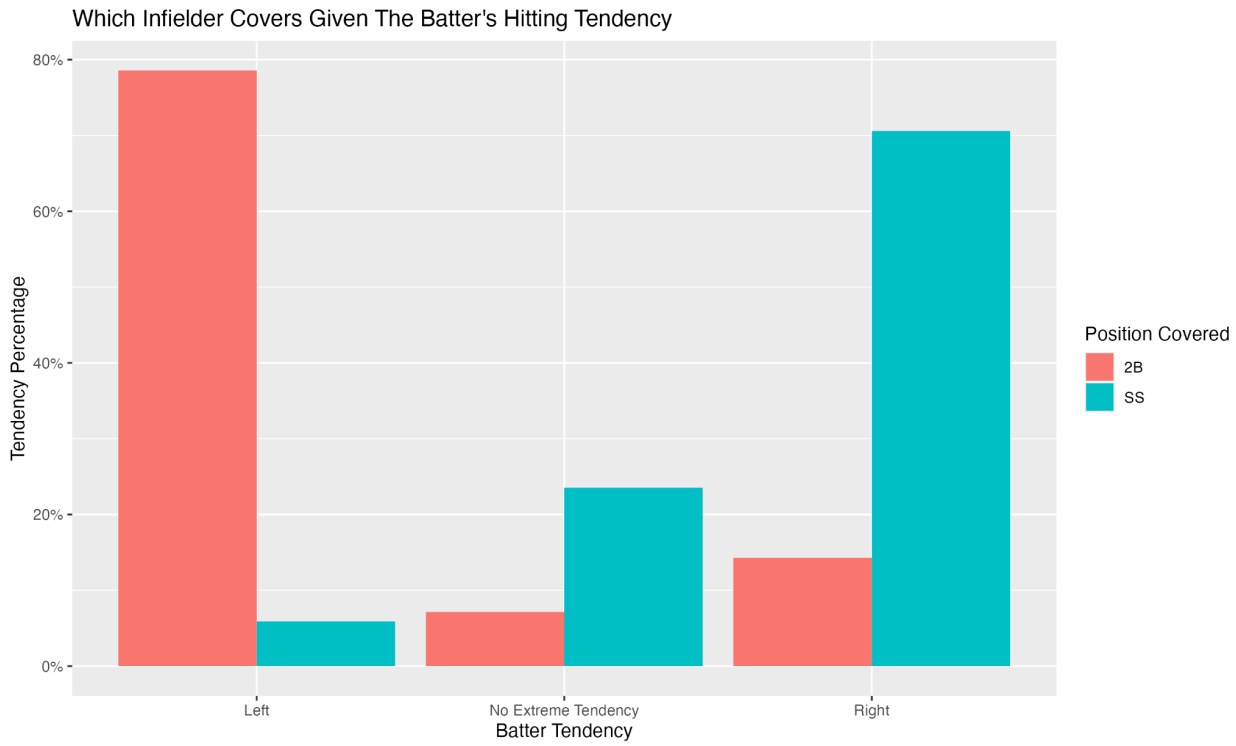
We fail to generally think about the space infielders leave when they run towards second. Both the shortstop and second base would leave a hole between second base and either first/third base that was not originally there. When deciding which infielder should cover second base, we have to also think about which infielder should leave a hole in the infield for the hitter. If the second baseman covers, then there is a hole on the right side of the infield, meaning it would be favorable for the hitter to hit it right. Logically thinking, the hole should be left on the side that the hitter is less likely to hit it towards. If a player who predominantly hits the ball to the right, the second baseman leaving a hole there due to covering second base would be suboptimal because they are likely to hit it into the hole. Therefore, we could predict that the shortstop should cover second base if the batter has a right-hitting tendency, and the second baseman should cover second base if the batter has a left-hitting tendency.

To determine the tendency of the batters, I decided to divide the infield into five sections from left to right. Splitting it evenly between first and third base, each section was approximately 25.47 feet wide. Since there were 5 sections in total, sections one and two were the two leftmost sections and sections four and five were the two rightmost sections. I did this because splitting the field into 2 sections, left and right, would not be as precise because hitting it barely to the left of second base does not necessarily mean that you typically hit it left, and the same applies with hitting it slightly to the right.



Using those sections, I sorted all the batters in the database into three categories: left-hitting tendency, right-hitting tendency, and no extreme tendency. Batters were deemed to have no tendency if the percentages were within 15% of each other, signaling that they hit it to both sides enough to not warrant them having any certain tendency.

I only accounted for balls that were hit in play and bounced before going 60 feet, signaling a ground ball. Using this information, I looked at which infielder covered second base given the batter's hitting tendency to see if they recognize the optimal positioning based purely off of the batter's tendency.



What this data showed was that the infielders did have a pretty good idea on which infielder should cover based off of the batter's tendency. There is still some room for improvement, as although 80% is great, it is 100%, which is the perfectly optimal approach. Overall, this data showed that at least there is some recognition of the optimal approach to infielder's positioning.

FINDING OPTIMAL APPROACH FOR SECOND INFIELDER

Now we have to determine what the other infielder should do. There are basically two options for the other infielder: stay where they are or back up second base.

The positives of the infielder staying where they are is that it prevents there being holes on both sides of the infield. Given that the infielder covering second is doing so because it is less likely for the batter to hit it that way, we can infer that they are likely to hit it to the side of the other infielder. Therefore, if that infielder decided to back up second base and leave a hole, it would logically cost them more runs because it is more likely the batter hits it in that newly left hole.

The positives of the infielder backing up second base is that it helps prevent runners from stealing from 1st to 3rd if the catcher makes a bad throw to the infielder covering second. Here is an example of one infielder covering second base and one infielder backing up second base when it works optimally: ("Infielder Backing Up Throw - MLB Film Room" in the Repo)
Infielder backing up second - MLB Film Room.mov

In this example, the second baseman covered second base, but due to a bad throw by the catcher it went past the second baseman, causing the runner to run to third base, thinking it was open. However, the shortstop ran to back up second base once the player started stealing from first base to second base and was able to get the ball before it went into the outfield, allowing him to make a throw to third base, which got the runner out. This is the positive of having the second infielder back up second base.

To determine which approach is better, I decided to look at run expectancy values and compare which saves more runs.

RUN EXPECTANCY ALLOWED WHEN INFIELDER BACKS UP SECOND BASE

I will start by going over the run expectancy allowed when the infielder backs up second base. This would only be detrimental if the batter makes contact and hits the ball into play because if the batter takes the pitch, there is no possible way he could get it through the newly left hole. Due to this, I had to calculate how often a batter hits the ball into play, which turned out to be 35.14% of the time. If only the second baseman covers second base, there is a 15.77% chance the batter gets a single to the right, meaning there would be a runner on 1st and 2nd now. If the shortstop backs up second base as well, then there is now a 74.98% chance the batter gets a single, given the probability that the batter hits it either to the left hole or right hole, meaning that would be the frequency of runners being on 1st and 2nd base. The same process was then taken for if only the shortstop covers second base. Averaging out the shortstop and second baseman data equally, due to the identical frequency, if no one backs up the throw, there is a 16.03% chance the better gets a single to the hole left by the covering infielder. If an infielder does back up the throw, there is now a 78.50% chance the batter gets a single to either of the holes left by the two infielders, which is an increase of 62.47% of players being on first and second bases.

Using run expectancies from the Run Expectancy Matrix from 2022, we know that a runner on first and second base with 0 outs expects 1.435 runs, 1 out expects .902 runs, and 2 outs expects .44 runs. We can now multiply these values by 62.47% to see the run expectancy difference, which is .896397, .5634497, and .274854 runs for 0, 1, and 2 outs respectively.

Now we have to multiply each of those run expectancy values by the chance the batter makes contact with a ball on a pitch, which was previously determined as 35.14% of the time. The run expectancy allowed by having both infielders cover on a single pitch is .32028, .20131, and .0981 runs on 0, 1, and 2 outs respectively.

RUN EXPECTANCY ADDED WHEN INFIELDER BACKS UP SECOND BASE

Now we have to calculate the run expectancy added for when an infielder backs up second base. When an infielder backed up the throw to second base, the catcher threw the ball poorly causing it to get past the infielder covering second base a total of 9 times. In 5 of those 9 times, the runner attempting to steal second base was able to steal 3rd base. This shows that when someone backs up the throw, the runner still advances to 3rd base a good amount of time.

There is a 55.56% chance that the runner goes from second to third base if the base is backed up on a bad throw. This is compared to 100% on a bad throw if an infielder does not back up second base, which is a 44.444% increase. We now need to calculate the run expectancy difference using the run probability matrix. I previously stated the run expectancy for runners on first and second base. A runner on first and third base has a run expectancy of 1.753, 1.147, and .5 runs with 0, 1, and 2 outs respectively. Knowing this and the frequency in which the runner gets to third base, it can be concluded that the run expectancy for when the infielder backs up second base is 1.61167, 1.03811, and .473334 runs with 0, 1, and 2 outs respectively. It would be 1.753, 1.147, and .5 runs with 0, 1, and 2 outs respectively if neither infielder backed up the throw. This means you are saving .13133, .01889, and .02667 runs depending on the number of outs. All of this would

only matter if the catcher makes a bad throw, which happens only 24.79% of the time. Using this frequency, the adjusted run expectancy equals .03503, .02699, and .006611 runs saved for every stolen base attempt if someone backs up the throw.

After examining the run expectancy allowed and added, we can get the total run expectancy of an infielder backing up second base. By having an infielder back up second base, you are saving .14133, .10889, and .02667 runs depending on the number of outs, but are then allowing .32028, .20131, and .09821 runs in return. This leads to an expected .17895, .09242, and .07154 runs for the opponents, compared to if only one infielder covered the base.

COMPLETE OPTIMAL APPROACH FOR INFIELDERS

Therefore, we can ultimately conclude that the optimal approach for the shortstop and second baseman to take on a stolen base attempt from first to second base is twofold. The first step is to decide on who covers second base. If the batter is more likely to hit it to the right, the shortstop should cover second base, but if the batter is more likely to hit it to the left, the second baseman should cover second base. Whichever infielder is not covering second base should stay where they are to prevent a second hole in the infield being left. The disadvantages of the second infielder backing up second base outweighs the advantages, so they should not back up the throw.

ACKNOWLEDGMENT

I would like to thank both SMT and Carnegie Mellon University for hosting this challenge, providing me with the opportunity to participate in such a unique and stimulating experience. Specifically, the opportunity to use tracking data was something I felt is extremely special, as it is growing in importance in the Sports Analytics industry.

I would also like to thank the judges and all those that have taken their time and expertise to help me through this entire process and ultimately examine the submissions.

Finally, I would like to thank Rice University and their Sport Management program for sharing this challenge with me, allowing me to get more experience and knowledge.

Thank you for providing me with this incredible opportunity in the Sports Analytics field!

Sincerely, Jonah Lubin

REFERENCES

Drummey, D. (2022, November 17). *Adjusting for the current run expectancy matrix*. Pitcher List. https://www.pitcherlist.com/adjusting-for-the-current-run-expectancy-matrix/

MLB Film Room. (n.d.). *Kevin Pillar Steals (7) 2nd base. Kevin Pillar out at 3rd, second baseman Hernan Perez to third baseman Mike Moustakas.: 06/14/2019*. MLB.com. https://www.mlb.com/video/kevin-pillar-steals-7-2nd-base-kevin-pillar-out-at-3rd-second-baseman-hernan-perez-to-third-baseman-mike-moustakas?q=ResultRunnerOnFirst+%3D+%5B%22C

AUGHT_STEALING%22%2C%22STOLE_BASE%22%5D+AND+RunnersOutAt+%3D+%5B3%5D+Order+By+Timestamp+DESC&cp=MIXED&p=0

MLB team runs per game. MLB Stats - MLB Team Runs per Game | TeamRankings.com. (n.d.). https://www.teamrankings.com/mlb/stat/runs-per-game