Post Hoc Ergo Propter Hoc? Does Getting Hit by a Pitch Affect a Batter's Performance in the Subsequent At-Bat?

David M. Lang
California State University, Sacramento

In 2019, there were 2,429 games played in Major League Baseball in which 1,984 batters who were Hit by a Pitch (HBP). This paper explores the effect of being hit by a pitch on the outcome of the subsequent at bat for that hitter in the same game. Results suggest that while, on average, there is no difference in subsequent performance on being hit by a pitch, that for the most prolific homerun hitters, being hit by a pitch increases the likelihood of making an out at the expense of decreases in reaching base - homeruns, in particular.

Keywords: major league baseball, hit by pitch, path dependence, incentives

INTRODUCTION

Academics seem to enjoy writing about baseball. In the past 20 years, there have been over 2,400 peer-reviewed journal articles published across the varied disciplines of the academy. Economists are no strangers to this literature having published almost 500 of these manuscripts. This is in no way surprising. Sports, with baseball leading the way, provide examples of labor economics, industrial organization, market structure analysis, asymmetric information, and game theory both theoretically and, with readily available performance data, rich empirical possibilities. Despite the prolific nature of this filed of research, there is currently, limited research that has attempted to analyze data concerning batters being hit by pitches. And, the research that has been conducted is limited to discussing the role and relationship of being hit by a pitch and the usage of the designated hitter in baseball.

This paper is really a paper about incentives and behavior. Do individuals -- even those that are highly skilled and highly paid professionals -- allow previous interactions to affect their motivation and outcomes going forward? In a sense, baseball is simply being used as the tool to address this question. In fact, this topic more generally is important to help understand behavioral economics and game theory related models. More specifically to baseball, however, the results of this paper are important for strategic purposes. If managers and pitchers know that a batter being hit by a pitch may behave and achieve differently in their subsequent at bat, especially with that same pitcher, then they might be much more careful in their approach with that batter.

Some of the more interesting related literature on baseball hitter performance considers both incentives and recent trends. O'Neil (2013) finds that hitter performance is greatly affected when playing during the final year in their contracts by considering 256 Major League Baseball free agent hitters playing under the 2006-2011 suggesting that opportunistic behavior is present when controlling for the intention to retire. Green and Zwiebel (2018) look at incentives and behavior more immediately by examining the presence of

"hot hand" performance in Major League Baseball and finds strong evidence for its existence. Further, they find that teams fail to respond appropriately to this hot hand effect. De Vany (2011) surprisingly finds that there has been no change in Major League Baseball home run hitting in recent years and that broken records are the result of chance variations in at bats, home runs per hit, and other factors. Koop (2002) cautions researchers and discusses the limitations of looking at singular statistics in baseball by expanding the literature concerning multiple output production to develop methods for comparing baseball players across several dimensions of performance.

More specifically, in the literature concerning players hit by a pitch, the focus has been on either *who* is hit by a pitch or whether there are retaliatory effects of being hit by a pitch. Kanago and Surdam (2020) demonstrate that while superior players are significantly more likely to be hit by a pitch, there is also a significantly higher rate of being hit by a pitch amongst Black batters. Trandel (2004) finds that despite the common belief that teams retaliate when a player is hit by a pitch, that season-long disaggregated data show little evidence of this. Goff, Shughart, and Tollison (1998) explore the moral hazard problem associated with the American League's 1973 adoption of the designated hitter rule and considering recent interleague play to estimate the magnitude of this problem.

This paper furthers this literature by considering the effect of being hit by a pitch on the next plate appearance in the same game. Estimates are disaggregated by the type of batter and the inning, weather, pitcher, and location are controlled for.

DATA

The data were collected from detailed play-by-play descriptions from www.retrosheet.org. In the 2019 season, there were 189,246 plate appearances. In order to isolate the effect of being hit by a pitch (HBP) on subsequent performance within the same game, the refined dataset looks exclusively at 10,420 of these plate appearances. To be included it this analysis, these plate appearances are immediately following either a walk (BB) or an HBP within the same game. Comparing the effect of an HBP to a BB is particularly informative and sensible. In terms of the immediate effect, these two events have exactly the same result in gameplay. In each case, the batter proceeds to first base. In each case, forced baserunners (those occupying first base, those occupying second base with first base already occupied, and those occupying third base with both first and second base already occupied) proceed to the next base. In terms of the credit the batter receives in his statistics, both situations result in a run batted in (RBI) if a runner scores, both fail to count as an official at-bat, both increase the batter's on-base percentage (OBA) by the same amount. As such, within the context of that inning and in terms of how a batter is perceived, these two events are simply identical. Therefore, comparing these two events isolates the effect of the HBP specifically.

Table 1 illustrates means and standard deviations for selected variables in the dataset for all observations and then also disaggregated depending on whether the previous plate appearance was an HBP or a BB. The first four variables listed are factors related to the weather and inning. The mean inning of 5.86 feels like a large number. Recall, however, that the plate appearances being considered must be at least the second plate appearance in the game for the batter (and potentially the third or beyond) as it must come immediately after he was hit by a pitch or walked. Note also that the inning, temperature, wind speed and wind direction are essentially identical regardless of previous plate appearance.

The next six variables are dummy variables for outcomes of the plate appearance being considered. For the entire dataset, roughly 23% of plate appearances resulted in strikeouts (K), another 42% resulted in other types of outs (Out), 13.5% resulted in Singles, 4.6% resulted in Doubles, 0.5% resulted in Triples, and 4% resulted in Homeruns (HR). The remainder of plate appearances consisted of BB, HBP, reaching on an error, and sacrifices (bunts or flys). Many of these percentages look relatively consistent regardless of previous plate appearance. However, there is some preliminary evidence that being hit by a pitch results in a higher likelihood of a Single (14.2% vs. 13.4%) or a Triple (0.7% vs. 0.4%) and a lower likelihood of a Double (4.3% vs. 4.6%) or a Homerun (3.5% vs. 4.1%).

The remaining variables presented consider attributes about the batter (HR Rate) or the pitcher (ERA and Same Pitcher). Overall, the mean HR Rate of 0.046 indicates that the average batter in the dataset hits

a homerun roughly every 22 at bats. Additionally, the Earned Run Average (ERA) of the pitcher faced averages 4.49 across the dataset and 47% of plate appearances are against the same pitcher who walked or hit that batter previously.

TABLE 1
SELECTED DESCRIPTIVE STATISTICS – MEANS (STANDARD DEVIATIONS)

	ALL	Last AB HBP	Last AB BB
Inning	5.86	5.89	5.85
	(2.24)	(2.21)	(2.24)
Temperature	73.15	73.71	73.08
(degrees F)	(10.82)	(10.43)	(10.86)
Wind speed	7.63	7.52	7.64
(mph)	(5.09)	(5.00)	(5.11)
Wind out?	0.41	0.42	0.41
(=1 if wind is blowing out)	(0.49)	(0.49)	(0.49)
K	0.23	0.24	0.23
(=1 if strikeout)	(0.42)	(0.42)	(0.42)
Out (non-K)	0.42	0.42	0.42
(=1 if groundout or flyout)	(0.49)	(0.49)	(0.49)
Single	0.135	0.142	0.134
(=1 if single)	(0.341)	(0.349)	(0.340)
Double	0.046	0.043	0.046
(=1 if double)	(0.210)	(0.203)	(.211)
Triple	0.005	0.007	0.004
(=1 if triple)	(0.068)	(0.084)	(0.065)
HR	0.040	0.035	0.041
(=1 if homerun)	(0.197)	(0.184)	(0.199)
HR Rate	0.046	0.044	0.046
(HR per AB for the Hitter)	(0.020)	(0.021)	(0.020)
ERA	4.49	4.48	4.50
(earned run average of pitcher)	(1.36)	(1.36)	(1.36)
Same Pitch?	0.47	0.49	0.47
(=1 if facing the same pitcher)	(0.50)	(0.50)	(0.49)
Last AB HBP?	0.109		
(=1 if HBP on previous AB)	(0.312)		
Number of Observations	**10,420**	**1,135**	**9,285**

Table 2 further disaggregates these data by looking at the means of these variables broken into the following six subcategories: previously walked with a HR Rate at least one standard deviation greater than the mean, previously hit by pitch with a HR Rate at least one standard deviation greater than the mean, previously walked with a HR Rate within one standard deviation of the mean, previously hit by pitch with a HR Rate within one standard deviation of the mean, previously walked with a HR Rate at least one standard deviation less than the mean, and previously hit by pitch with a HR Rate at least one standard deviation less than the mean. The idea here is to see if in addition to potential differences in outcome determination by previous events, to identify if this effect varies for different types of hitters (prolific HR

hitters, average HR hitters, and anemic HR hitters). While there are many potentially interesting differences in this table, the most noteworthy are to be found amongst the most prolific HR hitters. Note that compared to those who walked previously, those who were hit by a pitch are more likely to have an out (either a strikeout or another type of out), and less likely to have a Single or Homerun and, therefore have much lower On Base Percentages (25.2% vs. 37.5%) and much lower Slugging Percentages (.410 vs. .575). Slugging Percentage, it should be noted, is a batting average that weights an at bat by the number of bases achieved with the hit (where Singles, Doubles, Triples, and Homeruns count for one, two, three, and four total bases, respectively).

TABLE 2
MORE SELECTED DESCRIPTIVE STATISTICS - MEANS

	Last: BB, HR: >1SD	Last: HBP, HR: >1SD	Last: BB, HR: ±1SD	Last: HBP, HR: ±1SD	Last: BB, HR: <1SD	Last: HBP, HR: <1SD
HR Rate	0.079	0.081	0.046	0.045	0.016	0.017
K	0.23	0.26	0.24	0.24	0.21	0.21
Out (non-K)	0.38	0.48	0.42	0.41	0.47	0.45
Single	0.128	0.077	0.130	0.146	0.153	0.173
Double	0.037	0.026	0.050	0.048	0.042	0.039
Triple	0.002	0.006	0.004	0.005	0.007	0.013
HR	0.072	0.058	0.041	0.041	0.013	0.004
OB%	0.378	0.252	0.339	0.352	0.309	0.320
SLG	0.575	0.410	0.459	0.471	0.345	0.345
# obs.	**1,423**	**155**	**6,312**	**749**	**1,550**	**231**

RESULTS

Table 3 displays overall Logit regression results for the entire dataset for five separate dummy dependent variables (K, Out, Hit, HR, and OB). For each of these specifications, weather (temperature and windspeed), batter statistics (homerun rate and batting average), pitcher statistics (earned run average and whether the batter is facing the same pitcher again), stadium fixed effects, and of course, whether or not the batter was hit by a pitch on the previous plate appearance were included.

For the K regression, homerun rate significantly increases the likelihood of a strikeout while batting average, earned run average, and facing the same pitcher significantly decreases the likelihood. For other types of outs (Out), batting average and facing the same pitcher significantly increase the likelihood of an out while homerun rate significantly decreases that likelihood. For any type of hit (Hit), batting average, earned run average, facing the same pitcher, and temperature all significantly increase the likelihood of a hit. For homeruns (HR) more specifically and not surprisingly, homerun rate significantly increases the likelihood of hitting a homerun. Finally, for getting on base (OB), homerun rate, batting average, earned run average, and temperature all significantly increase the likelihood of reaching base. However, of particular note, in none of these specifications does being hit by a pitch significantly affect these outcomes. The coefficient on that variable remains insignificant. This suggests that batter performance may be unaffected comparing the events of being hit by a pitch to being walked in the previous plate appearance.

TABLE 3
LOGIT ESTIMATES OF SPECIFIC OUTCOMES

Independent Variable	K	Out	Hit	HR	OB
Temperature	-0.003	-0.003	0.009*	0.008	0.003
	(-0.13)	(-1.61)	(3.94)	(1.61)	(1.55)
Windspeed	-0.007	-0.004	0.005	-0.012	0.009*
	(-1.46)	(-1.12)	(1.12)	(-1.19)	(2.16)
Last AB HBP?	0.019	0.014	0.005	-0.141	-0.026
	(0.25)	(0.21)	(0.06)	(-0.83)	(-0.39)
HR Rate	5.19*	-5.82*	-1.03	22.56*	2.93*
	(4.35)	(-5.71)	(-0.86)	(8.97)	(2.77)
ERA	-0.10*	-0.01	0.08*	0.08*	0.09*
	(-5.53)	(-0.87)	(4.47)	(2.29)	(-5.58)
Same Pitcher?	-0.19*	0.15*	0.11*	0.16	-0.01
	(-4.10)	(3.84)	(2.36)	(-1.57)	(-0.18)
AVG	-6.14*	1.53*	4.87*	-1.15	3.20*
	(-8.55)	(2.47)	(6.51)	(-0.72)	(-4.90)
Constant	0.74*	-0.22	-3.52*	-4.91*	-2.29*
	(2.99)	(-1.03)	(-13.56)	(-8.87)	(-10.16)
Observations	10,420	10,420	10,420	10,420	10,420

z-scores are reported in parentheses.
* indicates significance at the 90% level.
Additional controls: Stadium fixed effects

Table 4 reconsiders the five specifications from Table 3 but looking exclusively at the anemic homerun hitters – those with homerun rates at least one standard deviation below the mean homerun rate in the dataset. While some significant dependent variables lost their significance, some remain and almost no new variables became significant. For the K regression, batting average continues to significantly decrease the likelihood of striking out. For other types of outs (Out), batting average continues to significantly increase the likelihood of an out. For any type of hit (Hit), batting average, facing the same pitcher, and temperature all continue to significantly increase the likelihood of a hit. For homeruns (HR) amongst these relatively anemic homerun hitters' homerun rate no longer matters but batting average significantly increases the likelihood of hitting a homerun. Finally, for getting on base (OB), earned run average continues to significantly increase the likelihood of reaching base. And as was the case across the whole dataset, in none of these specifications does being hit by a pitch significantly affect these outcomes.

TABLE 4
LOGIT ESTIMATES OF SPECIFIC OUTCOMES: ANEMIC HR HITTERS

Independent Variable	K	Out	Hit	HR	OB
Temperature	0.004	-0.005	0.011*	-0.004	0.001
	(0.76)	(-1.02)	(2.04)	(-0.23)	(0.11)
Windspeed	0.009	-0.013	-0.006	-0.035	0.007
	(0.76)	(-1.37)	(-0.55)	(-0.80)	(0.65)
Last AB HBP?	0.040	-0.100	0.070	-1.124	0.059
	(0.23)	(-0.70)	(0.42)	(-1.09)	(0.39)
HR Rate	8.47	-3.31	-3.46	55.02	1.28
	(0.97)	(-0.47)	(-0.41)	(1.51)	(0.17)
ERA	-0.06	-0.03	0.04	-0.11	0.08*
	(-1.45)	(-0.86)	(0.97)	(-0.65)	(2.12)

Same Pitcher?	-0.17	0.03	0.31*	0.63	0.09
	(-1.23)	(0.33)	(2.65)	(1.38)	(0.88)
AVG	-6.10*	2.29*	3.41*	-9.85*	2.31
	(-3.84)	(1.72)	(2.04)	(-1.92)	(1.59)
Constant	-0.00	-0.09	-3.19*	-2.17	-1.89*
	(-0.01)	(-0.19)	(-5.43)	(-1.13)	(-3.72)
Observations	1,781	1,781	1,781	1,781	1,781

z-scores are reported in parentheses.
* indicates significance at the 90% level.
Additional controls: Stadium fixed effects

Table 5 once again reconsiders the five specifications, now looking exclusively at the prolific homerun hitters – those with homerun rates at least one standard deviation above the mean homerun rate in the dataset. Again, there are some commonalities compared to the overall data regressions. For the K regression, homerun rate continues to significantly increase the likelihood of a strikeout while batting average, earned run average, and facing the same pitcher continues to significantly decrease the likelihood. For other types of outs (Out), homerun rate continues to significantly decrease the likelihood. For any type of hit (Hit), batting average, earned run average, and facing the same pitcher all continue to significantly increase the likelihood of a hit. For homeruns (HR) amongst these relatively prolific homerun hitters' homerun rate no longer matters but earned run average significantly increases the likelihood of hitting a homerun. Finally, for getting on base (OB), batting average and earned run average all continue to significantly increase the likelihood of reaching base.

Interestingly, however, amongst these prolific homerun hitters, we find significant effects of previously being hit by a pitch compared to previously being walked. In particular, compared to being walked, being hit by a pitch significantly increases the likelihood of making an out and significantly decreases the likelihood either getting a hit, reaching base, or hitting a homerun, in particular.

TABLE 5
LOGIT ESTIMATES OF SPECIFIC OUTCOMES: PROLIFIC HR HITTERS

Independent Variable	**K**	**Out**	**Hit**	**HR**	**OB**
Temperature	0.003	-0.004	0.007	0.005	0.003
	(0.48)	(-0.79)	(1.23)	(0.57)	(0.64)
Windspeed	0.014	-0.014	0.006	-0.005	0.003
	(1.17)	(-1.33)	(0.47)	(-0.25)	(0.30)
Last AB HBP?	0.099	0.458*	-0.410*	-0.295*	-0.589*
	(0.50)	(2.66)	(-1.81)	(-1.81)	(-3.01)
HR Rate	18.14*	-15.69*	-7.16	-4.28	3.08
	(2.77)	(-2.72)	(-1.08)	(-0.39)	(0.53)
ERA	-0.15*	-0.03	0.11*	0.20*	0.14*
	(-3.15)	(-0.78)	(2.68)	(3.12)	(3.64)
Same Pitcher?	-0.26*	0.06	0.21*	0.12	0.12
	(-2.12)	(0.60)	(1.77)	(0.62)	(1.12)
AVG	-5.24*	-0.44	4.27*	-4.55	4.04*
	(-2.39)	(-0.23)	(1.88)	(-1.24)	(2.04)
Constant	-0.74	1.36*	-2.97*	-2.37*	-2.80*
	(-0.86)	(1.81)	(-3.36)	(-1.65)	(-3.63)
Observations	1,578	1,578	1,578	1,578	1,578

z-scores are reported in parentheses.
* indicates significance at the 90% level.
Additional controls: Stadium fixed effects

As Logit coefficients are notoriously difficult to interpret on face value, Table 6 aids in the interpretation of the coefficients on the Last AB HBP variable in particular for each of the five specifications for the prolific homerun hitters displayed in Table 5. In each row, the Logit coefficient and its accompanying z-statistic is presented. Columns II and III display predicted probabilities for the dependent variable, controlling for all other variables previously indicated, comparing previously walked to previously hit by a pitch. Finally, column IV calculates the percentage change associated with being hit by a pitch compared to being walked in the previous plate appearance.

The coefficient for the K regression was not statistically significant, but is displayed here nonetheless form completeness and comparison. For the regression on Out, controlling for the other variables present, a batter previously having walked has a 37.8% likelihood of making an Out (non-K) compared to a 49.0% likelihood for a batter having previously been hit by a pitch. That is a 29.6% increase in the probability of making an out – even greater than that if strikeouts were also incorporated given the direction of that effect. Similarly, a batter having previously walked has a 23.5% of getting a Hit compared to a 16.9% for a batter having previously been hit by a pitch. Similar sizeable effects can be found for homeruns in particular and reaching base more generally (decreases in 24.6% and 33.2%, respectively).

TABLE 6
CHANGES IN PREDICTED PROBABILITY: LAST AB HBP VS. BB, PROLIFIC HR HITTERS

	I	II	III	IV
Dependent Variable	**Logit**	**after BB**	**after HBP**	**Change**
K	0.099 (0.50)	0.231	0.249	Increase: 7.8%
Out	0.458* (2.66)	0.378	0.490	Increase: 29.6%
Hit	-0.410* (-1.81)	0.235	0.169	Decrease: 28.1%
HR	-0.295* (-1.81)	0.069	0.052	Decrease: 24.6%
OB	-0.589* (-3.01)	0.376	0.251	Decrease: 33.2%

CONCLUSIONS

Major League Baseball players are, by definition, professional and exception at their jobs. Yet, it has been argued that there is perhaps no more challenging thing to do in any sport than hitting a baseball. After all, the most successful hitter in 2019 successfully reached base in a mere 43.8% of plate appearances by a combination of hits, walks, and times hit by pitch (Mike Trout). For a team to be successful on offense, it must score as many runs as possible and this means reaching base and hitting homeruns as frequently as it can while minimizing strikeouts and other outs. On the other side, a team pitching is successful by maximizing outs and limiting baserunners and homeruns.

On its face and in the box score, a batter being walked and a batter being hit by a pitch are identically treated. However, what this analysis demonstrates is that while that is true at the time of that particular plate appearance, the differential effect of being hit by a pitch lingers and may affect the performance during the subsequent plate appearance. While this differential effect does not show up in the overall results for the more anemic of homerun hitters, the effect is substantial and significant for the more prolific homerun hitters. Specifically, for these hitters, being hit by a pitch compared to walking leads to a nontrivial reduction in the probability of reaching base, getting a hit, and hitting a homerun in particular.

The results suggest that first, there is a path dependence to batting within a game of baseball. Plate appearances are not independent events within a baseball game. Secondly, note that a particular game

theoretic outcome comes forth: if confronted with the choice of walking a prolific homerun hitter or hitting that same hitter with a pitch, there may be a significant benefit to hitting that batter in terms of performance in the next plate appearance.

REFERENCES

De Vany, A. (2011). Steroids and home runs. *Economic Inquiry*, *49*(2), 489-511.

Goff, B.L., Shughart, W.F., II, & Tollison, R.D. (1998). Moral hazard and the effects of the designated hitter rule revisited. *Economic Inquiry*, *36*(4), 688-692.

Green, B., & Zwiebel, J. (2018). The hot-hand fallacy: Cognitive mistakes or equilibrium adjustments? Evidence from Major League Baseball. *Management Science*, *64*(11), 5315-5348.

Kanago, B., & Surdam, D.G. (2020). Intimidation, discrimination, and retaliation: Hit-by-pitches during the integration of Major League Baseball. *Atlantic Economic Journal*, *48*(1), 67-85.

Koop, G. (2002). Comparing the performance of baseball players: A multiple-output approach. *Journal of the American Statistical Association*, *97*(459), 710-720.

O'Neill, H.M. (2013). Do Major League Baseball hitters engage in opportunistic behavior? *International Advances in Economic Research*, *19*(3), 215-232.

Trandel, G.A. (2004). Hit by pitches: Moral hazard, cost-benefit, retaliation, or lack of evidence? *Journal of Sports Economics*, *5*(1), 87-92.