The Intentional Base-on-ball Phenomenon in Baseball:

A Statistical Analysis and Strategic Recommendations

by

Andrew Kappy

A thesis presented to the University of Waterloo

in fulfillment of

the thesis requirement for the degree of

Master of Applied Science

in

Management Sciences

Waterloo, Ontario, Canada, 2005

Author's Declaration for Electronic Submission of a Thesis

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The game of baseball is amenable to a variety of strategies that affect short-term outcomes. This paper employs regression analysis, simulation, and cognitive analysis of mental biases to analyze the strategic scenario known as the "Intentional Base-on-Balls" and proposes a model to explain that strategy and predict its effectiveness.

The results of this study suggest that managers are prone to Type II errors, that is, issuing an Intentional Base-on-Ball in a situation where objective analysis suggests otherwise. Results further suggest that the ratio of Type I errors to Type II errors is disproportional to the ratio of their respective costs. This imbalance points to a subjective component to the decision-making process, one that can be explained by biases and cognitive errors.

The results and model described in this paper may allow managers to avoid future mistakes and improve their decision-making ability.

Acknowledgements

I would like to acknowledge the help of my supervisor Scott Jeffrey for all his hard work. I would also like to thank Major League Baseball for years of entertainment.

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

In major league baseball, there is an action called the intentional base-on-balls (IBB), whereby a manager will make a strategic decision to instruct his pitcher to intentionally impart four consecutive unhittable pitches rather than allow the opposing batter the opportunity to swing freely. Also referred to as an intentional walk, the IBB is in theory a means of minimizing risk. The walk is usually used in the three following situations. 1) In order to set up a force out when first base is open with a runner on 2nd or 3rd base; 2) In a similar situation, to set up a double play; and 3) when the manager is afraid of the offensive potential of the current batter and would rather pitch to the subsequent batter.

While the use of the Intentional Base-on-Ball has its merit, and is beneficial in a number of situations, it can be argued that managers have recently resorted to this tactic as a projection of a particular bias as opposed to a rational strategic decision. For example, on June 22nd 2003, San Francisco Giant outfielder Barry Bonds was intentionally walked twice by the Florida Marlins with the bases empty, once with one out and once with two outs. Although this is a rare occurrence, it leads one to question the motives of the manager. Was he fearful of the offensive player, or was his faith in his pitcher so little that he would rather walk a batter than face him in a situation where only a home run would directly lead to a run scored? Adding to the complexity of determining the answer is the fact that Barry Bonds, during that same 2003 season, hit a homerun only 8.7% of his at-bats — roughly once every 11.5 chances.

This paper will attempt to analyze the strategic decision-making process with respect to the IBB, and the biases inherent in that process. In addition to these primary concerns, the paper will attempt to confirm two hypotheses. The first states that managers are tempted to offer an IBB in a situation where statistically it is not their most effective tactic. The second hypothesis states that managers are influenced by a number of biases and predispositions that will pressure them to vary from the normative model. These biases and predispositions include loss/risk aversion, incentive bias/agency problem, path dependence, confirmation bias/availability bias, regret theory/omission-commission, and representativeness.

The problem can be summarized in the following chart, which shows the four possible scenarios involving an IBB problem.

		Suggestion	
Action		*Pitch*	*Walk*
	Pitch	A	B
	Walk	C	D

Boxes ‘A’ and ‘D’ represent scenarios where the normative model and the manager’s decision agree. In these situations, no biases are presumed to be inherent and little investigation needs to be conducted.

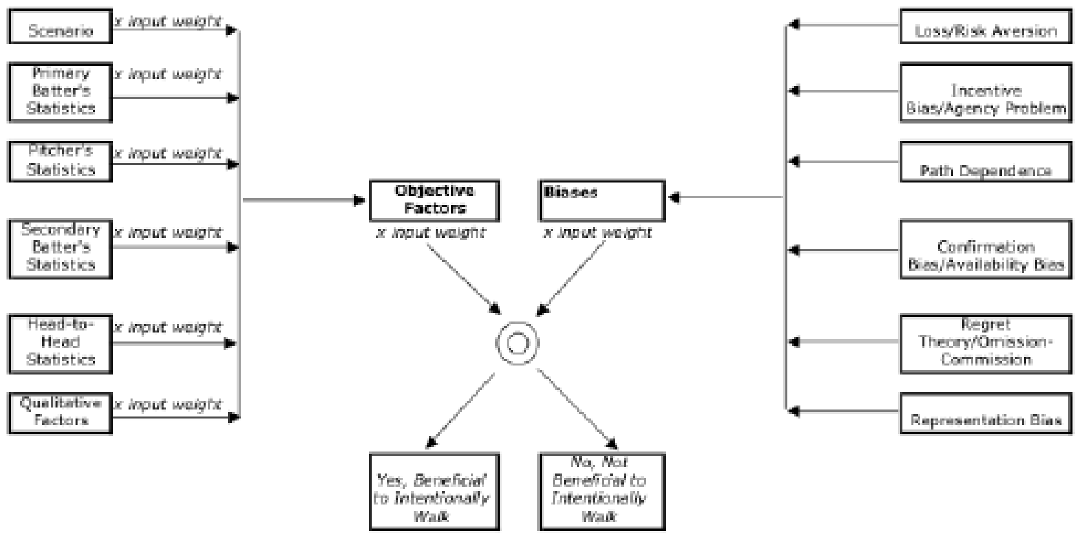
The focus of this paper will be on Boxes ‘B’ and ‘C’. Box ‘B’ represents a Type I error whereby the manager should have instructed the pitcher to intentionally walk the batter, but chose to order a pitch. Box ‘C’ represents a Type II error whereby the manager should have instructed the pitcher to pitch, but chose to issue an IBB. In these two situations, the normative model and the actions taken do not agree and are possibly influenced by a combination of biases.

I hypothesize that, in the case of IBB, Type II errors are more prevalent than Type I errors after taking into account the associated costs, or:

*Freq(Type II Errors)*Cost(Type II Errors) > Freq(Type I Errors)*Cost(Type I Errors)*

Quantifying the specific costs associated with each error may be difficult. However, should there be a dramatic variance between the frequency of each error, I maintain that the ratio of costs could not possibly be that high. For example, should the data suggest a 5:1 ratio between the frequencies of Type I errors versus Type II errors, a relative cost of 1:5 must exist to maintain balance. Since this ratio is probably unrealistic, I will be able to assume that some tendency towards this type of error exists. This paper will analyze the strategic decision involved in an Intentional Base-on-Balls, review historical analysis of the Intentional Base-on-Balls, develop a normative model and discuss the biases inherent in the decision to stray from the normative model.

1.1 Conceptual Model



1.2 Scope of Analysis

The scope of this project assumes that a number of variables are non-factors. Although this assumption may limit the conclusions, it discounts information that is not necessarily in the control of the players through whom the managers manage. These factors include defensive fielding as well as stadium characteristics.

1.2.1 Fielding

Behind every pitcher stands eight defensive players whose role is to assist the pitcher. It is reasonable to assume that certain players are better defensively than others. It may be reasonable to further assume that, at times, a defensive player will commit an error, or commit a spectacular play that cannot be predicted for the average player.

Errors, miscues, and exemplary play will lead to deviations from the normative model that may affect future results. However, for the purpose of the present model, I will not consider errors and such. This is a reasonable assumption, since managers cannot predict these actions, and hence managers should not be expected to take them into account when devising a strategy involving an IBB.

1.2.2 Stadium Characteristics

In 2004, major-league baseball had 30 teams, each playing in their own home stadium. Stadiums are not cookie-cutter constructions; they come in different sizes, dimensions and even altitudes. While the left-field fence at Boston's Fenway Park measures at 334 feet, the left-field fence at Detroit's Comerica Park measures at 402 feet. While New York's Yankees Stadium is situated at sea level, Colorado's Coors Field is one mile above sea level. Stadium variance will affect a batter's ability to get a hit or reach an additional base, thereby affecting the statistical relationship between pitchers and batters. It has been calculated that a home run hit at Yankee Stadium traveling 400 feet would travel 408 feet at Atlanta's Turner Stadium and 440 feet at Coors Field. Similarly, while a ball hit 340 feet to left field of Comerica Park will be an out, at Fenway Park it would be a homerun. Given the variation, stadium size and dimension will affect every relationship between a batter and a pitcher. Due to the number of stadiums and the complex disparities between these stadiums, this paper will not consider their effect on the intentional-walk phenomenon in baseball.

2.0 The Strategic Decision involved in an Intentional Base-on-Balls

In the article 'An Analysis of the Intentional Base of Balls', John F. Jarvis (1999) compiled two tables that he used to explain the use of the IBB. Table 2.0a outlines the historical performance of all batters between 1980 and 1996, while Table 2.0b outlines the historical performance of all batters, whose at-bat directly followed that of an IBB during that same time period.

Table 2.0a[1]: Hitting Summary for All At-Bats			
	BA	SLG	AB
NL	0.257	0.383	1,048,080
AL	0.264	0.405	1,184,055
Total	0.261	0.395	2,232,135

Note: Inclusive of at-bats in Table 2.0b

Table 2.0b: Hitting Summary for All At-Bats following an Intentional Base-on-Balls			
	BA	SLG	AB
NL	0.241	0.356	18,177
AL	0.261	0.394	15,012
Total	0.250	0.373	33.189

The above tables show that there appears to be an incentive to intentionally walk a batter since, on average, the subsequent batter has a reduced expected output level. While Table 2.0a shows a batting average of .261 or 26.1%, the hitters following an IBB have a batting average of .250 or 25%.

However, the article also points to the fact that, in a typical at-bat, hitters gain approximately 0.09644 bases per at-bat, while hitters gain 0.11171 bases per at-bat following an IBB. In light of this, it may be considered detrimental to offer an IBB. This paper will attempt to understand the discrepant predictions.

---

[1] BA is the ratio of hits per at-bat. SLG is the ratio of total bases per at-bat.

2.1 Historical Analysis of the Intentional Base-on-Balls

Since there are 3 bases and runners can be on any or all of them, there are 8 possible base combinations. Given that a team can be in an inning with 0, 1 or 2 outs, there are thus 24 total configurations. Much of the analyses conducted on intentional base-on-balls relate to the probabilities associated with these 24 scenario configurations.

Jarvis (1999) attempts to determine whether the expected future runs in a given scenario was increased due to an IBB by employing two distinct mathematical methods. The first is a historical comparison between actual and expected outcomes, and the second is a regression analysis between IBBs and actual runs.

2.1.1 Historical Comparison between Actual and Expected Outcomes

Jarvis is able to prove that by utilizing the IBB, managers were able to save 39 runs over the 16,160 that were scored in the situations he analyzed. Basing his study on observed historical records, Jarvis measured the expected future runs in each of the 24 possible scenarios.

Table 2.1.1a: Expected Future Runs

Outs/Runners	---	1--	-2-	12-	--3	1-3	-23	123
0	0.492	0.872	1.21	1.487	1.363	1.744	1.982	2.310
1	0.261	0.519	0.684	0.907	0.955	1.167	1.386	1.558
2	0.098	0.224	0.327	0.439	0.375	0.501	0.595	0.763

The table above indicates an overall trend of increasing expected runs, as (i) the number of runners increase, and/or (ii) the number of outs decrease. This trend matches expectations since these conditions would generate increased opportunities. However, the trend would imply that it is never beneficial to intentionally walk a batter, as the expected future runs will always increase. It is this point that indicates a limitation in the analysis of Jarvis, and begins to explain the theory extension this paper will propose.

Based on the probabilities of singles, doubles, triples, homeruns and outs from overall batter statistics during that same time period, as seen in Table 2.1.1b, Jarvis recalculated the expected future runs in each scenario.

Table 2.1.1b: Aggregate Probabilities – Prof. Jarvis										
	BA	SLG	AB	S	D	T	HR	TBB	IBB	HBYP
Overall Statistics	0.261	0.395	2,232,135	409,542	102,963	14,172	55,676	215,272	20,638	13,795

Table 2.1.1c: Expected Future Runs								
Outs/Runners	---	1--	-2-	12-	--3	1-3	-23	123
0	0.522	0.943	1.121	1.557	1.121	1.557	1.736	2.256
1	0.283	0.568	0.738	1.035	0.738	1.035	1.205	1.585
2	0.106	0.246	0.368	0.513	0.368	0.513	0.635	0.843

In order to properly calculate the number of runs that would have been scored should an IBB not have been imparted, Jarvis recalculated the expected future runs and weighted each hitter's statistics based on the number of IBBs they received. A hitter who received more IBBs would have his statistics weighted more heavily.

Table 2.1.1d: Weighted Offensive Earned Run Average								
Outs/Runners	---	1--	-2-	12-	--3	1-3	-23	123
0	0.644	1.106	1.270	1.750	1.270	1.750	1.914	2.481
1	0.353	0.670	0.832	1.164	0.832	1.164	1.326	1.748
2	0.135	0.293	0.412	0.578	0.412	0.578	0.698	0.933

Finally, using the available data concerning the number of times each of the 24 scenarios has occurred, it is possible to measure the total number of runs that would have been scored if an IBBs was not imparted.

Table 2.1.1e: IBB Counts								
Outs/Runners	---	1--	-2-	12-	--3	1-3	-23	123
0	1	1	220	3	63	78	406	0
1	1	11	3306	11	911	176	4140	0
2	13	30	6428	24	1387	102	2483	0

Using simple multiplication between the above two tables, a total of 16,160 runs would have been scored versus the 16,121 that were scored after issuing an IBB. This provides a value of –0.003 runs per IBB, indicated that the strategy did save a few runs, indicating value in an IBB at the aggregate level.

It should be noted that the above analysis speaks of the aggregate results of 16,000+ scenarios disregarding the individual abilities and achievement of the key playmakers. This analysis does not take the pitcher into account and is thus inadequate as a tool for proper decision-making. The present paper will build on Jarvis's results and propose a model that integrates individual player differences to create a tool that is adequate for better decision-making.

2.1.2 Regression Analysis between Intentional Base-on-Balls and Actual Runs

A second method of analysis is based on linear regression between offensive events and runs scored per single event. Jarvis based his findings on 424 team-season records (424 *162 games) and 16,121 total runs scored between the 1980 and 1996 seasons. When we use 11 independent variables and runs as the dependent variable, the correlations look as follows:

Table 2.1.2a: Parameter Weights (runs/event)	
Event	**Runs**
Outs	-0.101
Strikeouts	-0.099
Single	0.439
Double	0.679
Triple	0.815
Homerun	1.484
Walk + Hit by Pitch	0.308
Intentional Base-on-Balls	0.033
Stolen Base	0.087
Caught Stealing	-0.238
Ground into Double Play	-0.429

The value of 0.033 attributed to IBB means that every Intentional Base-on-Balls imparted led to an average of 0.033 runs. This number is significantly lower than the coefficient attributed to any other positive offensive scenario including the categories hit by pitch (0.308) and single (0.439). This indicates that an IBB leads to fewer runs than other offensive categories. Using only this piece of information, it could be argued that an IBB would always be preferred. This paper is an attempt to quantify argument by looking at hitter/pitcher head-to-head match-ups.

One quantifiable value for an IBB can be illustrated by utilizing the aggregate results of Table 2.1.1b and Table 2.1.2a. Table 2.1.1b documents the likelihood of each offensive outcome, while Table 2.1.2a indicates the projected runs associated with each offensive outcome. By multiplying the results for the variables Outs, Singles, Doubles, Triples and Homeruns, it is possible to determine the aggregate projected runs expected per at-bat, that is, the projected runs associated with a single at-bat that does not result in an IBB. This number can be compared to the projected runs associated with a single IBB (0.033).

The results of the multiplication is a project run value of 0.0421, meaning that a single at-bat that does not result in an IBB will lead on average to 0.0421 runs. This value is greater than the 0.033 expected runs for a single IBB. This result seems to validate the aggregate use of the IBB.

2.2 The Normative Model Extension

Although the above methods provide a good analysis of the effect of an IBB on future runs, such an analysis neglects two significant variables: the statistics of both the individual hitters and the statistics of the individual pitchers participating in this scenario. Knowing that an IBB was beneficial in an aggregate analysis is of little benefit to managers attempting to judge individual situations whose variables differ from that of the aggregate. A manager may make one decision when there is one out and a runner on second and the batter is superstar Alex Rodriguez followed by weak-hitting Tony Clark. He may make a completely different decision in a scenario where weak-hitting Chris Woodward is followed by the superstar Carlos Delgado.

The proposed normative model allows the user to input pre-specified data and receive an outputted decision stating whether it is beneficial to intentionally walk the batter in the given situation. The model combines simulation and regression techniques. The development of the model required three key steps. These were: gathering raw data, testing data to determine historical relationships that predict future at-bat behavior and creating a computerized tool that will simulate the remainder of an inning.

2.2.1 Raw Data

The first step was gathering raw data to be used in the creation of mathematical relationships between historical statistics and projected outcomes. The difficulty in this step was to eliminate unnecessary data from the complex web of baseball statistics. Since there were 540 batters in the 2004 season, with their participation ranging from 1 to 690 at-bats, a reduced sample was necessary to limit those batters whose limited at-bats would skew the overall sample. Such a skew would occur should a batter hit a home run in their only career at-bat against a given pitcher, especially if that pitcher is not known for giving up homeruns. Using data from a 1994 data set, only those batters with greater than one thousand career at-bats were included. Similarly, I used only the statistics of a reasonable sample of pitchers, those whose number of career at-bats faced was greater than one thousand,.

Necessary information required for both hitters and for pitchers can be found in the following table:

Table 2.2.1a – Information Required (A)	
For Batters	**For Pitchers**
At-Bats	At-Bats Against
Singles	Singles Allowed
Doubles	Doubles Allowed
Triples	Triples Allowed
Homeruns	Homeruns Allowed
Strikeouts	Strikeouts Allowed
Outs (less strikeouts)	Outs (less strikeouts) Allowed
Hit By Pitch	Hit By Pitch Allowed
Walks	Walks Allowed

The above data was gathered at CBS Sportsline.com's baseball statistics section, where one is able to view career statistics on all active and retired players. When necessary, I employed more detailed data from the Retrosheet's website, an official publication for Sabremetrics (the statistics of baseball) analysis.

The above information was necessary to compute historical individual data. Further data was required to represent the head-to-head data between the pitcher and the hitter. This information includes:

Table 2.2.1b – Information Required (B)
For Head-to-Head Batter vs. Pitcher
Head-to-Head At-Bats
Head-to-Head Singles
Head-to-Head Doubles
Head-to-Head Triples
Head-to-Head Homeruns
Head-to-Head Strikeouts
Head-to-Head Outs (less strikeouts)
Head-to-Head Hit By Pitch
Head-to-Head Walks

I gathered head-to-head data from two primary sources. The first source was a publication entitled Bill James Presents STATS 1994 Batter versus Pitcher Matchups, (1994). The publication breaks down every hitter versus pitcher match-up prior to the 1994 season and surveys both active offensive and defensive players. Secondary information was compiled from CBS Sportsline.com's pitcher versus batter section. This source provided updated information for the years 1995-2005.

2.2.2 Simulation Data

In the development of the model, I used raw data compiled from historical baseball statistics to develop a general probability equation, for each individual batter/pitcher pair.

2.2.3 Simulation

The third step involves the creation of a baseball simulator. This will enable a user to input historical information and, based on the allotted probabilities, will perform a number of simulations to determine what would most likely occur throughout the remainder of that given inning.

3.0 Normative Model

An important component of the simulation tool was the relationship between the individual batters and the individual pitchers when determining the at-bat outcome. Although much data is available on individual performance, it is the relationship between the performances of these individuals that matters most in the proposed simulation. It would be extremely easy to record only the individual performance of a pitcher and/or batter and extrapolate as though outcomes were independent of second parties. Unfortunately, it is improbable that players of varying abilities will be as good in all situations. Tom Hanrahan discusses this exact point in an article he published in 2001 entitled Does Good Hitting Beat Good Pitching? Using a sample of all hitters who, between 1984 and 1996, had at least 446 at-bats, and all pitchers who, during the same period, faced at least 100 batters, Hanrahan was able to come up with the following table.

Table 3.0a: American League Pitcher/Batter Performance				
Player Category	**Poor Batter (avg.<.253)**	**Average Batter (.252<avg<.283)**	**Good Batter (avg>.282)**	**All Batters**
Good Pitchers (opposing average less than .253)				
At Bats	93,137	106,936	81,393	281,466
Hits	19,215	25,320	22,272	66,807
Batting Average	.2063	.2368	.2736	.2374
Average Pitchers (opposing average between .253 and .283)				
At Bats	93,321	111,247	86,180	290,748
Hits	21,995	29,988	26,020	78,003
Batting Average	.2357	.2696	.3019	.2683
Poor Pitchers (opposing average greater than .283)				
At Bats	63,793	77,928	61,820	20,3541
Hits	16,822	23,437	20,664	60,923
Batting Average	.2637	.3080	.3343	.2993
All Pitchers				
At Bats	250,251	296,111	229,393	775,755
Hits	58,032	78,745	68,956	205,733
Batting Average	.2319	.2659	.3006	.2652

Unsurprisingly, this table indicates that a pitcher will fare better against a poor batter than a good batter and a batter will fare better against a poor pitcher than a good pitcher. This necessitates the creation of a model that will quantify the relationship between an individual pitcher and an individual batter.

One solution was developed by Dan Levitt (1999). In this article, Levitt reintroduces an equation developed in 1983 by Bill James that attacks the same pitcher/batter relationship. The equation not only determines the relationship between the ability of the hitter and the ability of the pitcher, it also normalizes that relationship against the league average. The purpose of the equation is to determine the most likely future average between a batter and a pitcher. The relationship James introduced, and tested against historical data, is the following:

$$= \frac{(BatAvg*PitAvg)/(LgAvg)}{(BatAvg*PitAvg)/(LgAvg) + ((1-BatAvg)*(1-PitAvg))/(1-LgAvg))}$$

Where

BatAvg = Batter's average, shown as percentage of hits per official at-bats

PitAvg = Aggregate batting average against a particular pitcher

LgAvg = The aggregate league batting average

Levitt (1999) tests the above equation against real-life data to see if the equation was still valid approximately 15 years after it was initially introduced. Levitt's results indicate that Bill James's equation is not only still effective, but predicts behaviour within only a few percentage points over 18 situations. Table 3.0b shows a breakdown of Levitt's analysis based on 1995 statistics.

Table 3.0b: Summary – Batter versus Pitcher Matchups									
		All Batters		**Good Batters**		**Average Batters**		**Poor Batters**	
		Actual	*Formula*	*Actual*	*Formula*	*Actual*	*Formula*	*Actual*	*Formula*
AL	*Pitchers:*								
	Good	.250	.249	.276	.275	.251	.251	.223	.220
	Average	.290	.290	.317	.319	.292	.293	.262	.259
	Poor	.317	.325	.340	.356	.323	.328	.287	.291
NL	*Pitcher*								
	Good	.245	.247	.266	.274	.251	.247	.218	.219
	Average	.284	.283	.314	.313	.280	.283	.259	.253
	Poor	.317	.321	.351	.353	.322	.321	.279	.289

By using the above equation, we can calculate projected batting averages for all present batter-versus-pitcher match-ups. For example, in a head-to-head match-up between All-Star pitcher Roger Clemens and All-Star hitter Barry Bonds, the projected batting average for the 2004 season would be the following:

$$= \frac{(BatAvg*PitAvg)/(LgAvg)}{(BatAvg*PitAvg)/(LgAvg) + ((1-BatAvg)*(1-PitAvg))/(1-LgAvg))}$$

$$= \frac{(.362*.208)/(.272)}{(.362*.208)/(.272) +((.638)*(.792))/(.728))}$$

= .285 projected batting average.

Similarly, if we remove the strong Barry Bonds and insert the weaker Adam

Dunn, the projected batting average would fall to .257

3.1 Application of Levitt's Theory

The equation introduced by Levitt, by way of Bill James, produces a useful and effective tool to predict future batting averages of head-to-head matchups. The question becomes: Can one extrapolate the frequency of singles, doubles, triples, etc. from a unitary statistic such as a batting average?

To initially compare trends in performance versus batting average, I used Batter Versus Pitcher Match-ups (Bill James, 1994) to collect data on 194 batters and 635,921 at-bats prior to the 1995 season. I grouped players by batting average, and calculated their associated ratios of single, doubles, triples, homeruns, strikeouts, non-strikeout outs, walks and hit batsmen. The purpose of this analysis was meant to assess initial high-level trends in the data.

Table 3.1a: Trends, By Average

Range	No.	S/H	D/H	T/H	HR/H	K/Out	Non-K/Out	BB/AB	HBP/AB
.330-.340	1	74.3%	19.8%	2.1%	3.8%	11.5%	88.5%	15.9%	0.32%
.320-.330	1	60.8%	20.7%	1.2%	17.3%	24.2%	75.8%	22.2%	0.53%
.310-.320	1	71.9%	17.2%	2.7%	8.2%	19.4%	80.6%	5.8%	0.73%
.300-.310	7	67.8%	19.6%	2.8%	9.7%	14.2%	85.8%	10.0%	0.53%
.290-.300	12	73.0%	16.8%	3.0%	7.2%	16.4%	83.6%	11.8%	0.66%
.280-.290	25	70.2%	17.0%	3.2%	9.6%	18.3%	81.7%	9.9%	0.56%
.270-.280	20	68.7%	18.3%	2.4%	10.6%	20.5%	79.5%	9.8%	0.52%
.260-.270	40	69.8%	17.8%	2.4%	10.0%	21.0%	79.0%	9.2%	0.71%
.250-.260	42	68.9%	17.7%	3.0%	10.4%	23.3%	76.7%	8.7%	0.71%
.240-.250	29	68.2%	17.5%	2.8%	11.5%	25.5%	74.5%	9.6%	0.65%
.230-.240	9	70.9%	18.2%	1.6%	9.4%	24.0%	76.0%	9.0%	0.91%
.220-.230	5	65.6%	17.5%	2.1%	14.9%	30.3%	69.7%	10.4%	0.96%
.210-.220	2	57.9%	19.0%	0.6%	22.5%	39.1%	60.9%	13.5%	1.04%
	194	69.2%	17.7%	2.7%	10.4%	21.8%	78.2%	9.7%	0.67%

Given the low number of batters with averages greater than .300 and lower than .230, these 'outside' ranges were eliminated since their small sample size might distort true relationships.

Table 3.1b: Condensed Trends, By Average									
Range	**No.**	**S/H**	**D/H**	**T/H**	**HR/H**	**K/Out**	**Non-K/Out**	**BB/AB**	**HBP/AB**
.290-.300	12	73.0%	16.8%	3.0%	7.2%	16.4%	83.6%	11.8%	0.66%
.280-.290	25	70.2%	17.0%	3.2%	9.6%	18.3%	81.7%	9.9%	0.56%
.270-.280	20	68.7%	18.3%	2.4%	10.6%	20.5%	79.5%	9.8%	0.52%
.260-.270	40	69.8%	17.8%	2.4%	10.0%	21.0%	79.0%	9.2%	0.71%
.250-.260	42	68.9%	17.7%	3.0%	10.4%	23.3%	76.7%	8.7%	0.71%
.240-.250	29	68.2%	17.5%	2.8%	11.5%	25.5%	74.5%	9.6%	0.65%
	194	69.2%	17.7%	2.7%	10.4%	21.8%	78.2%	9.7%	0.67%

The condensed chart seems to indicate a positive relationship between batting average and singles/hit, non-strikeout outs/out, and walks/at-bat while indicating a negative relationship with doubles/hit, home runs/hit, strikeouts/out and hit batsmen/at-bat. Interesting the one category that does not show a positive or negative relationship is triples/hit. However, this category shows a curvilinear relationship, with peaks at the extreme.

In order to take the analysis from high-level trends to a more detailed driven analysis, I more closely scrutinized each of the above eight variables and divided them into three major headings: hits (containing singles, doubles, triples and homeruns), outs (containing strikeout, and non-strikeout) and misc. (containing walks and hit batsmen).

3.1.1 Section I – Outs

A batter can be retired in one of two ways: by a strikeout or by a hit-out. The following subsection is devoted to analyzing the relationship between the two options and determining whether one can predict their future likelihood based on historical statistics.

Using Batter Versus Pitcher Match-Ups (1994) as a source for data, I compiled the historical statistics from 194 batters and 14 pitchers, each within the American League. The data set includes the same 635,921 at bats used above as well as 118,785 batters faced (754,706 total at-bats). The data set includes only hitter-versus-pitcher relationships that contain greater than 20 at-bats. Using this dataset, I analyzed the strikeout-per-out (k/out) ratio.

A regression analysis was conducted using the head-to-head k/out ratio as the dependent variable and the historical ratios of both the pitchers and hitters as the independent ratios. As the input data is in the form of percentages, the typical method of *ordinary leased squared* does not apply. Therefore, *negative binomial regression* was used to rid the data of its sigmoid-shaped curve. I performed this alteration by altering both the left and right side of the input data into the form of:

= LN(proportion)/(1-proportion)

When originally conducted with all 1,146 runs, the R-Squared was approximately 20%. However, in a secondary test, runs that contained extreme values were removed, leaving an R-Squared of 48%. Extreme values were represented by data sets where both the career percentage of the batter and the career percentage of the pitcher varied from the head-to-head percentage by greater than 10%. This discrepancy was interpreted to mean that the head-to-head percentage was inconsistent and not representative of the true relationship. Using this new data, containing 505 total runs, the output looked as follows:

Table 3.1.1a: Regression Results, K/Out				
Variable Name	**Estimated Coefficient**	**Standard Error**	**T-Ratio 77 DF**	**P-Value**
Pitcher's K/Out	0.6489009	0.057797	11.22723	0
Batter's K/Out	0.5357182	0.033805	15.84722	0
Constant	0.1906349	0.07794	2.4459	0.007

R-Squared	0.484
R-Squared (adjusted)	0.482
F-Stat	237.135

Therefore the equation to calculate the head-to-head Strikeout/Out ratio can be represented as follows:

Head to Head K/O = 0.1906349+ P-K/O(0.6489009) + B-K/O*(0.5357182)*

**Note that K/O is the ratio of strikeouts per out, P-K/O is the pitcher's historical K/O ratio and B-K/O is the batter's historical K/O ratio.

The simulation model employs this equation to determine the relative ratio of strikeouts versus outs. However, please note that each of the three percentages must be converted (and reconverted) to take into account the transformation to the negative binomial regression form. This information can also be used to determine the percentage of non-K outs. As Levitt's theory determines the total percentage of outs, and the above regression analysis determines the percentage of K/out, the remainder would fall in to the non-K category.

3.1.2 Section II – Hits

This section determines a reasonable approach for estimating the percentage of hits that will be singles, doubles, triples or homeruns. Each of these constitutes a variable.

3.1.2.1 Homeruns per Hit

Similar to previous sections, data was regressed, removing all extreme and abnormal data points[2]. The final data set included 556 runs. The negative binomial regression data looks as follows:

Table 3.1.2.1a: Regression Results, HR/H				
Variable Name	**Estimated Coefficient**	**Standard Error**	**T-Ratio 77 DF**	**P-Value**
Pitcher's HR/H	5.478921	1.441312	8.981361	0
Batter's HR/H	1.9979163	0.130760	15.27923	0
Constant	12.944944	0.617769	8.868912	0

R-Squared	0.4004
R-Squared (adjusted)	0.4026
F-Stat	187.351

Given the above data, the equation to determine the future head-to-head HR/H ratio is as follows:

$$HR/H = 12.944944 + (B\text{- }HR/H)*1.9979163 + (P\text{- }HR/H)*5.4789421$$

3.1.2.2 Ratio of Singles, Doubles and Triples per Hit

This section created a little more difficulty than that of homeruns. When negative binomial regression analysis was run, the total variances explained by the regression were 4.8% and 5.8% respectively for doubles and triples. These numbers were far too small to create useful simulation data. I was able to overcome this on the basis of my experience with baseball and my knowledge of professional players: there are certain types of hitters who are more inclined to hit doubles and triples. Especially in respect to triples, speedy base-runners are far more likely to hit triples than slower, more powerful, batters. Furthermore, data trends show that the ability to hit safely for triples is not significantly effected by the particular pitcher. Therefore it was deemed reasonable to use historical triples/hit ratios for given batters. This could be logically extended to include doubles, as again, speedy batters are more inclined to hit them, and the data indicates that this doesn't vary greatly depending on the pitcher. Given that it was is possible to determine the total number of hits, percentage of homeruns, percentage of doubles and percentage of triples, I obtained the category of singles by subtration. (singles = hits – home runs – doubles - triples)

[2] Representing all data sets where both the batter's historical HR/H ratio and the pitcher's historical HR/H ratio varied from the head-to-head ratio by greater than 10%.

3.1.3 Section III - Other

3.1.3.1 Walks per At-Bat

Similar to the above regression data that I ran on the strikeout/out ratio and the homerun/hit ratio, I collected and regressed data with respect to the number of walks a player will receive in the course of their total number of at-bats. (Total at-bats include hits and outs). Again, I eliminated extreme and abnormal data, leaving 940 total runs[3]. Given this analysis, the output looked as follows:

Table 3.1.3.1a: Regression Results, BB/AB				
Variable Name	**Estimated Coefficient**	**Standard Error**	**T-Ratio 77 DF**	**P-Value**
Pitcher's BB/AB	2.2616779	0.281364	8.03825	0
Batter's BB/AB	2.2539445	0.172654	13.0546	0
Constant	7.2333445	0.765607	9.44785	0

R-Squared	0.201
R-Squared (adjusted)	0.200
F-Stat	118.540

Therefore, the equation to properly determine the number of walks a player will generate per at-bat is equal to:

*BB/AB = 7.2333445 + (P-BB/AB)*2.2616779 + (B-BB/AB)*2.2539455*

3.1.3.2 Hit-By-Pitcher per At-Bat

Similar to the situation with doubles and triples, regression with respect to hit batsmen rendered statistically insignificant results. However, I feel that it is reasonable to assume that the HBP/AB ratio will closely mirror that of the individual pitcher as the pitcher is primarily in control of the location of the pitch.

3.1.4 Example

To clearly illustrate the use of the above data analysis and the creation of the normative percentages, the following example show a match-up between Barry Bonds, a batter on the San Francisco Giants, and Roger Clemens, a pitcher on the Houston Astros. All data used is from the 2004 season. The initial step was to determine, using Levine's theory, the projected battering average in a head-to-head match-up. Using the pitcher/batter matchup equation introduced in Section 3.0, the projected average between Barry Bonds (.362 average) and Roger Clemens (.208 average against) is .285.

Table 3.1.4a: Summary, Projected Averages		
Category	**Name**	**Average**
Batter	Bonds, Barry	0.362
Pitcher	Clemens, Roger	0.208
League	MLB	0.272
Projected Average		0.285

This number means that of every 100 at-bats between Barry Bonds and Roger Clemens, Barry Bonds will garner approximately 29 hits and 71 outs. Using the regression equations for HR/H, BB/AB and K/out, as well as historical information for HBP, doubles and triples, the following totals and percentages can be calculated.

[3] Representing all data sets where both the batter's historical BB/AB ratio and the pitcher's historical BB/AB ratio varied from the head-to-head ratio by greater than 10%.

Table 3.1.4b: Summary, Projected Probabilities								
	S	**D**	**T**	**HR**	**K**	**Non-K**	**BB**	**HBP**
Totals	14	6	1	8	19	53	49	-
Probability	9.4%	4.2%	0.5%	5.1%	12.5%	35.6%	32.7%	0.0%

4.0 Simulation

I employed a simulation tool to predict the aggregate outcome of all relationship that would occur during the remainder of that inning. The purpose of this simulation was to measure its strategic effect by predicting the expected future runs in a scenario in which an IBB was imparted, and the expected future runs in a scenario in which an IBB was not imparted.

The simulation tool itself was created in Microsoft Excel and takes advantage of numerous lookup tables and random number generation as it predicts future behaviour based on historical statistics. Using the Boston Red Sox as an example, a final scenario window would look as follows:

Table 4.0a: Simulation

Lineup	B/G Scenario	Outcome	End Scenario	Outs	Runs Scored
Johnny Damon					
Orlando Cabrera					
Manny Ramirez					
David Ortiz					
Kevin Millar					
Bill Mueller					
Dave Roberts					
Doug Mientkiewicz					
Pokey Reese					
Total					

In the above table, **Lineup** represents the hitters in the Boston Red Sox lineup that are due up during the inning under investigation. The model allows for as many as 9 hitters as less than 0.1% of all innings last beyond 9 at-bats.

B/G Scenario represents the game scenario, as of the first pitch of the at-bat. As there are 24 potential combinations of scenarios as determined by the number of base runners and outs, there are 24 scenarios that could occur prior to an at-bat. Each of the 24 possible scenarios was given a number from 1 to 24. The details of all 24 scenarios can be seen in the Appendix A binary chart, where one represents 'Yes' and zero represents 'No'.

The column titled **Outcome** represents the calculated outcome of a head-to-head at-bat between the given pitcher and the given batter based on historical data fed through a pre-specified regression program. That outcome is based on a random seed, which in turn is based on possible events and the likelihood of those events.

End Scenario represents the scenario that is a result of the initial scenario and the outcome of the present at-bat. Similar to the B/G scenario, the end scenario is allotted a number between 1 and 24.

The column labeled **Outs** represents the number of outs in the inning at the end of the at-bat, based on the relationship between the initial scenario and the end scenario.

The final column, labeled **Runs,** describes the number of runners who crossed home plate due the outcome of that lone at-bat. The value of the runs generated is based on the initial scenario and the at-bat's outcome.

The subsequent section will discuss the individual columns as well as the mathematics and modeling techniques behind them.

4.1 Calculating Outcome

The column in Table A entitled **Outcome** represents the key output to the relationship between the pitcher and primary batter. This relationship is expressed as a mathematical relationship between historical statistics, and is presented in the form of probable percentages that each of the eight possible scenarios will occur (single, double, triple, homerun, strikeout, walk, out, hit by pitch). Given these probabilities, a randomly generated number, in combination with the weights of these percentages, will give the calculated result.

4.1.1 Step 1: Outcome Probabilities

In the example of an at-bat between a Boston Red Sox hitter and New York Yankee pitcher Mike Mussina, I determined the following probabilities based on historical statistics calculated by the methods described in the previous section:

Table 4.1.1a: Hitter Specific Probabilities								
	Probability							
Hitter	**Single**	**Double**	**Triple**	**Homerun**	**Walk**	**Non-K**	**Strikeout**	**HBP**
Johnny Damon	18.2%	5%	0.9%	3.5%	7.3%	51.3%	13.9%	0%
Orlando Cabrera	16.5%	5.6%	0.4%	2.3%	3.2%	58.4%	13.5%	0%
Manny Ramirez	14%	7%	0%	6.5%	8.7%	44.9%	18.9%	0%
David Ortiz	13.1%	7.3%	0.5%	6.2%	7.7%	45.5%	19.7%	0%
Kevin Millar	17%	6.3%	0%	3.8%	6.6%	48.9%	17.4%	0%
Bill Mueller	16.1%	6%	0.2%	3.3%	7.6%	51.5%	15.4%	0%
Dave Roberts	15.3%	3.8%	1.9%	1.9%	7.1%	53.6%	16.3%	0%
Doug Mientkiewicz	13.9%	5.4%	0.2%	2%	7.3%	55.1%	16.1%	0%
Pokey Reese	15.6%	2.6%	0.7%	1.8%	3.7%	53.2%	22.4%	0%

4.1.2 Step 2: Random Number Generation

As a means of simulating an actual at-bat, a random number is uniformly generated between 0 and 1, using the function =RAND() in Microsoft Excel, which will determine the outcome of the at-bat, given the above probabilities. For each batter a random number is generated. In the sample case, the following numbers were randomly generated:

Table 4.1.2a: Random Numbers	
Hitter	**Random No.**
Johnny Damon	0.47
Orlando Cabrera	0.93
Manny Ramirez	0.89
David Ortiz	0.91
Kevin Millar	0.95
Bill Mueller	0.76
Dave Roberts	0.22
Doug Mientkiewicz	0.0
Pokey Reese	0.72

4.1.3 Step 3: Reported Outcome

The final step is to use the randomly generated number as a tool to help simulate the outcome of an at-bat. Based on the random number, using the Boston Red Sox's Johnny Damon as an example, the outcome will look as follows:

Table 4.1.3a: Outcome Possibilities	
Range	**Outcome**
0 – 0.18	Single
0.18 – 0.23	Double
0.23 – 0.24	Triple
0.24 – 0.27	Homerun
0.27 – 0.38	Walk
0.38 – 0.90	Out (non-strikeout)
0.90 – 1.0	Strikeout

Therefore, in the given scenario, where the randomly generated number is 0.47, the outcome will be an *Out* as shown in Table 4.1.3a. This method is repeated for every subsequent batter for the rest of the inning.

4.2 End Scenario

The *end scenario* represents the scenario that occurs at the end of the at-bat, or at the beginning of the subsequent at-bat. Based on simplistic analysis of running speed, and taking into account only the final destination of the batter, the end scenario could be generated through the table found in Appendix B.

Intuitive reasoning and experience with baseball suggests that not all hits lead to the same outcome, nor to the same base destination for the lead runner. For example, where a single is hit with a runner on first base, the lead runner won't necessarily remain at second base and may make their way to third. Similarly, a runner may be able to advance an additional base given that a single was hit to right field as opposed to left field.

This idea was emphasized in an article written by Dan Levitt entitled Hitters and Baserunner Advancement (1999). Levitt discusses the aggregate analysis of historical statistics between the years 1980 and 1983 and shows that there is in fact a statistical difference given the placement of a batter's hit, but that it varies depending on the situation.

Table 4.2a: Single with Runner on First

		Runner Destination – Percent				
Fielder	**# of hits**	**Out**	**Second**	**Third**	**Home**	**Bases/hit**
N/A	1,727	1.5%	69.3%	28.5%	0.8%	1.3
LF	8,952	2.1%	77.8%	19.1%	1.0%	1.21
CF	8,465	2.2%	61.5%	34.6%	1.8%	1.38
RF	8,757	2.2%	46.9%	49.4%	1.6%	1.52
INF	3,231	1.9%	87.4%	9.2%	1.5%	1.12
Total	**31,132**	**2.1%**	**65.2%**	**31.3%**	**1.4%**	**1.34**

As shown in the above table, analyzing the 31,132 singles that occurred with a runner on first, on average 65.2% of the lead runners ended up on second, 31.3% on third, 1.4% at home and 2.1% of the lead runners were out.

By segmenting those hits between left field (LF), center field (CF), right field (RF) and infield (INF), it is possible to see variations with respect to those destination percentages. With most of the variance occurring between second base and third base, one can see that, based on historical statistics, runners are more likely to advance to third when the single was hit to RF (49.4%) than if the ball was hit to CF (34.6%) or LF (19.1%). These statistics seem reasonable as the throw from RF to third base is the longest, while the throw from LF is the shortest.

Similar analysis was conducted using a single with a runner on second, a single with a runner on third, a double with a runner on first, and a double with a runner on second. Their respective tables and probabilities can be found in Appendices C through F.

With respect to the simulation created for the purpose of this project, I decided not to model the destination of the hit, but rather the overall statistics that describe the advancement of the runner.

Based on the information gathered by Levitt, I deduced the probability of a runner advancing to a particular base, given the type of hit. For example, given a single where a runner is on first, the lead runner will end up out 2.1% of the time, end up at second base 65.2% of the time, end up at third base 31.3% of the time and will end up at home 1.4% of the time. It is these probabilities that were inserted into the simulation model to more realistically simulate actual events.

Table 4.2b: Aggregate Information to be used in Simulation Tool				
	Lead Runner Destination - Percent			
Scenario	**Out**	**Second**	**Third**	**Home**
Single w/ Runner on 1st	2.1%	65.2%	31.3%	1.4%
Single w/ Runner on 2nd	3.6%	1.2%	29.9%	65.3%
Double w/ Runner on 1st	3.1%		53.6%	43.3%
Double w/ Runner on 2nd	0.1%		1.4%	98.4%
Single w/ Runner on 3rd	0.1%		0.9%	99.0%
Total	**0.1%**		**1.4%**	**98.4%**

This information was inserted into the simulation through the use of look-up tables and simple spreadsheet calculations. By way of illustration, the remainder of this section will discuss possible outcomes of scenario 4: a runner on first with zero out.

As described in Table 1, given a single, there are four possible outcomes, each representing a different destination for the lead runner:

Scenario 1 – Outcome includes runners on first, runner on second (65.2% probability)

Scenario 2 – Outcome includes runner on first, runner on third (31.3% probability)

Scenario 3 – Outcome includes runner on first, runner out (2.1% probability)

Scenario 4 – Outcome includes runner on first, runner at home (1.4% probability)

Similarly, given a double, there are three possible outcomes:

Scenario 1 – Outcome includes runner on second, runner on third (53.6% probability)

Scenario 2 – Outcome includes runner on second, runner at home (43.3% probability)

Scenario 3 – Outcome includes runner on second, runner out (3.1% probability)

Given the above scenarios, I used a randomly generated number in conjunction with a selection process — similar to that utilized in the Outcome section of the model — to select a fair representation of the simulation process.

The overall table, which represents all available options with respect to lead runner destinations, can be found in Appendix G.

4.3 Out

The column entitled Outs represents the number of outs in the inning at the end of the at-bat. This number is generated using a look-up table based on the beginning scenario, the outcome and the final scenario. The complete look-up table can be found in Appendix H.

4.4 Runs

The final column, entitled Runs, describes the number of runners that crossed home plate due to the outcome of that lone at-bat. The value of the runs generated, based on the initial scenario and the at-bat's outcome, is represented by a look-up table that can be found in Appendix I.

4.5 Aggregate Tool

Therefore, based on the look-up tables, and information generated from both internal and third-party analysis, the final simulation table depicts the most probable run production for the remainder of any given inning. The following situation takes into account not only the starting lineup of the Boston Red Sox and the starting pitcher of the

New York Yankees, but also the in-game scenario. Based on this information, the table predicts a quick two-batter inning and no runs scored.

Table 4.5a – Simulation					
Lineup	**B/G Scenario**	**Outcome**	**End Scenario**	**Outs**	**Runs Scored**
Johnny Damon	8	Out	9	2	0
Orlando Cabrera	9	Strikeout	End	3	0
Manny Ramirez	End	Strikeout	End	End	End
David Ortiz	End	Strikeout	End	End	End
Kevin Millar	End	Strikeout	End	End	End
Bill Mueller	End	Out	End	End	End
Dave Roberts	End	Homerun	End	End	End
Doug Mientkiewicz	End	Single	End	End	End
Pokey Reese	End	Out	End	End	End
Total					**0**

I must stress that the above example is only one of numerous possible outcomes that are highly dependent on the random number generated for each player. To give a more accurate prediction, I simulated the total inning a number of times and averaged and analyzed the final run tally. As can be seen in Table 4.5b below, the average runs scored in this situation would be 0.6.

The question then becomes, what would the outcome of the scenario be if Johnny Damon were given an Intentional Base-on-Balls? This can be easily computed by substituting the word ‘walk’ in the Outcome column. Simulating the scenario 20 times gives the following results:

Table 4.5b: 20 Scenario Result		
Scenario No.	**No IBB**	**Yes IBB**
1	1	0
2	1	1
3	0	4
4	0	3
5	1	0
6	0	3
7	0	0
8	0	0
9	0	2
10	2	1
11	1	2
12	0	2
13	1	0
14	0	0
15	3	0
16	1	3
17	0	1
18	0	0
19	0	4
20	1	0
Average Runs	**0.6**	**1.3**

Based on 20 scenarios, and based on the validity of the imported data, it seems apparent that in the situation where Johnny Damon is up against Mike Mussina with a runner on second and one out, the best strategy would be to pitch to him. Not only are the average runs dramatically higher with an IBB, but the variance also poses a heavy risk since, in 8/20 scenarios, more than one run would be scored. This type of simulation and analysis can be used for any combination of hitters and batters and applied to any in-game scenario.

5.0 Disparity between Normative Model and Visible Action

With the normative model, we can analyze the historic use of the IBB and verify the success of those decisions.

5.1 Aggregate Analysis of the Intentional Base-on-Balls

While it is easy to criticize a manager's decision based on a single situation, the more accurate method would be to analyze the aggregate results of many decisions. It is for this reason that I analyzed all 47 IBBs issued during the major-league week of June 1, 2004. This selection represents a typical week, covers the spectrum of intentional base-on-ball usage, and represents the range of major league managers' decision-making abilities and strategies. It does not restrict my analysis to any one of the three major IBB usage techniques, but tests all three. The 47 situations were run through the simulation model to verify the effectiveness of the decision. My model indicated that in 22 of the 47 situations an IBBS was the correct decision, in 25 of them, the manager would have been better off pitching to the batter, saving their team an estimated 0.28 runs in the current inning.

		Suggestion	
Action		*Pitch*	*Walk*
	Pitch	A	B
	Walk	25	23

A failure rate greater than 50% is poor by any business or sport standard. Managers can ill-afford to make 25 strategic errors per week on a minor decision like the IBB. Furthermore, in 12 of the 47 examples, independent of the strategy used, the offensive team was able to score at least one run. In these examples, the average runs scored per inning were slightly higher for the situations in which the normative model suggested a pitch as opposed to the situations where it suggested an IBB.

In a sport wrought with statistics and variables, managers must learn to utilize the objective tools available to them to minimize errors, and minimize the future runs their errors would create. As managers are not typically statisticians or mathematicians, the usefulness of these tools may not be apparent. It is for this reason that I believe that the proposed simulation model can act as an effective tool for Major League managers.

5.2 Examples

Example 1

Date: April 14, 2004
Game: San Francisco Giants against the Milwaukee Brewers
Situation: 3-0 Milwaukee in the seventh inning

- Two out with runners on first and second base
- Milwaukee's Jason Bennett is pitching to Barry Bonds

In the above example, the manager of the Milwaukee Brewers seems to make a cardinal mistake walking Barry Bonds to load the bases when his team is leading by three runs late in the game. By walking Barry Bonds, the manager is bringing the winning run to the plate, meaning that a homerun by Edgardo Alfonzo would propel the San Francisco Giants into the lead. In this instance, logic might suggest that the manager used the IBB when it was unnecessary.

Upon running the simulation in turns out that the manager's decision was, in fact, correct in that the projected future runs given an IBB is 0.65 and the projected future runs without an IBB is 0.9. Furthermore, not only did San Francisco score in more simulated situations with the IBB, the average number of runs scored per instance was higher as well. This matches with the actual occurrence of a fly-out by the subsequent batter Edgardo Alfonzo.

So in this situation the suggestion of the normative model reflects the decision of the manager, where traditional baseball thinkers might have acted otherwise.

Example 2

Date: April 11, 2004
Game: San Francisco Giants against the San Diego Padres
Situation: 3-0 San Diego in the eighth inning

- One out and runners on second and third base
- San Diego's Jay Witasick is pitching to Barry Bonds

In the above situation Barry Bonds was intentionally walked to load the bases with only one out with the San Diego Padres leading 3-0 in the eighth inning. What actually occurred was a 5-run outburst propelling the Giants to a 5-3 lead at the end of the inning. Simulating the situation in my model, the results suggest that no IBB should have been issued. While the projected future runs with an IBB are 1.9, they are only 1.5 without an IBB.

Example 3

Date: April 20, 2004
Game: Toronto Blue Jays against the Boston Red Sox
Situation (Part A): 3-0 Boston in the seventh inning

- One out with a runner on second base
- Toronto's Acquilino Lopez is pitching to Manny Ramirez

Situation (Part B): 4-0 Boston in the seventh inning

- Two out with runners on second and third base
- Toronto's Valerio De Los Santo is pitching to Mark Bellhorn

The above situation contains two separate IBBs, each of which can be analyzed for their strategic accuracy. Simulating the first IBB indicates that it would have been better to pitch to the batter, in that the projected runs is 0.85 with an IBB, and 0.35 without. This is further reinforced in that the manager of the Toronto Blue Jays chose to walk Manny Ramirez and the pitcher subsequently let up a run-scoring double. Simulating the second IBB also indicates a mistake, with the projected runs being 0.45 for an IBB and 0.25 for not issuing an IBB.

5.3 Aggregate Analysis of a Non-Intentional Base-on-Balls

Similar to Section 5.1, the normative model was used to analyze 50 random decisions that did not result in an IBB. The purpose of this analysis is to measure the degree of Type II errors. Using 50 randomly selected at-bats during the same week beginning June 1, 2004, in only three cases did the model call for walking the batter when the manager decided to pitch to them. This suggests that managers rarely make the mistake of not intentionally walking someone in a scenario where it is deemed acceptable.

		Suggestion	
Action		*Pitch*	*Walk*
	Pitch	47	3
	Walk	C	D

The three scenarios in which the model suggested an IBB were extremely similar in that they involved the batter directly preceding a poor-hitting pitcher in the National League. In these scenarios, the pitcher had a batting average below .200 with less than 1 homerun per hitter. However, it may be argued that the manager did not impart an IBB, in fear of a pinch-hitter with stronger statistics. Aggregating the tables in section 5.1 and 5.3 shows the following results:

Action		**_Suggestion_**	
		Pitch	*Walk*
	Pitch	47	3
	Walk	25	22

This analysis suggests that the original hypothesis — that Type II errors will occur more prevalently than Type I errors — was correct by a factor of 8.

$$25 * \mathit{Cost}(\mathit{Type\ II\ Error}) > 3 * \mathit{Cost}(\mathit{Type\ I\ Error}), \text{ or}$$

$$\frac{25}{3} > \frac{\mathit{Cost}\ (\mathit{Type\ I\ Error})}{\mathit{Cost}\ (\mathit{Type\ II\ Error})}$$

Although it is difficult to determine the cost of each error, it is reasonable to assume that the cost of a Type I error is not eight times that of a Type II error, leaving room for necessary qualitative interpretations of the variance.

6.0 The Inherent Biases in the Decision

In the sport of baseball, the game-to-game play calling is managed by an imperfect set of metrics and forecasted situations. Presently, no manager possesses a tool that objectively analyzes the information at hand and computes most-likely outcomes of given strategic decisions. Instead, the sport is governed by 'experience', 'tradition', and the fear of scrutiny by the media and upper management. The simulation model proposed in this paper is meant to provide the objective tool presently unavailable.

Due to the current regime of subjectivity, there is no governing formula with respect to the IBB. As can be seen in the following chart, there is a large discrepancy with respect to the number of IBBs ordered by each team's manager per 100 innings pitched.

Table 6.0a: 2004 Intentional Base-on-Balls, By Team

Team	**IBB/100 IP**	**Team**	**IBB/100 IP**	**Team**	**IBB/100 IP**
Colorado	7.5	Cincinnati	4.0	Los Angeles	2.5
Arizona	7.2	Oakland	3.9	Minnesota	2.3
Pittsburgh	5.8	Chicago (N)	3.4	Milwaukee	2.2
Philadelphia	5.1	Baltimore	3.2	Anaheim	2.2
Montreal	5.0	Detroit	3.0	Boston	2.2
Atlanta	4.4	Tampa Bay	3.0	Seattle	2.0
Florida	4.3	New York (A)	2.9	Chicago (A)	1.6
Cleveland	4.3	Kansas City	2.8	Texas	0.6
Houston	4.2	San Diego	2.6	**Total**	**3.5**
Toronto	4.2	St. Louis	2.5		
New York (N)	4.0	San Francisco	2.5		

I hypothesize that, in the case of IBB decisions, there are a number of biases, as well as psychological and sociological phenomena that will shift the basis for a manager's decision from the objective and rational towards the subjective and utility-maximizing. In this section, I suggest why disparity might exist between the normative model and the reported actions.

It is hypothesized that six psychological phenomena may play a role in the overuse of the IBB: loss aversion/risk aversion, representativeness bias, confirmation bias/availability bias, regret theory/omission-commission, path dependence, and an agency problem.

Furthermore, it may be argued that even if management possessed an objective, forecasting tool such as the normative model presented in this paper, management would still be influenced by the following subjective biases and phenomena.

6.1 Loss Aversion/Risk Aversion

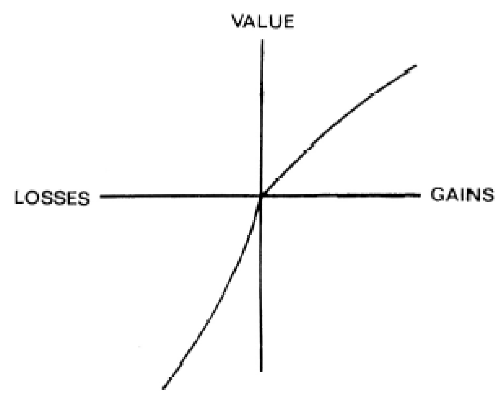
In the realm of business, relationships or sport, winning is always preferred to losing. Winning provides self-confidence, enhances a participant's self-image and improves social status while losing incurs the opposite. All decisions entail the element of win-versus-lose element and the element of risk, that is, the possibility that the decider will make the choice that leads to the 'loss' as opposed to the 'win'.

Kahneman et al (1991) discuss the element of risk and state that individuals dislike losing so much that they will overcompensate towards a risk-averse position. Studies show that an individual faced with two scenarios, the first being an 85% chance of winning $1,000 and a 15% of winning nothing and the second being a 100% chance of winning $800 will almost always select the second alternative even though the present value of the first alternative is much greater. This example represents the element of loss aversion/risk aversion. Even in the case of equal expected values, individuals tend to lean towards the option with less variance in the outcomes. Bernoulli (1954) suggests that "people do not evaluate prospects by the expectation of their monetary outcome, but rather by the expectation of the subjective value of these outcomes."

It is my hypothesis that a manager, facing a situation that might merit an IBB, may overcompensate (Type II error: call for an IBB when it might not be appropriate) to avoid a heavier loss associated with a homerun or extra-base hit. This overcompensation would not be based on a fear of the batter, but rather a fear of allowing additional bases (or runs) that could have been avoided if an IBB had been initially bestowed. Based on historic patterns, a batter is more likely to garner an IBB if he has a history of high homerun totals which are correlated with high strikeout totals. Therefore, a manager will choose to walk a riskier batter for a batter with less extreme statistics. This notion goes against traditional expected utility theory that focuses on objective utility based on statistical analysis, and is more in line with prospect theory, suggested by Kahneman and Tversky (1979), which emphasizes the value function.

One major portion of prospect theory is loss aversion. Loss aversion refers to the tendency for people to strongly prefer avoiding losses to making gains. In the course of studying monetary risk, Kahneman and Tversky's developed prospect theory (1979), which holds that, while individuals were risk averse in a positive domain, they were risk seekers in a negative domain — a relationship that Kahenman and Tversky illustrated in the following figure.

Figure 6.1a: Kahneman and Tversky's Value Function



It is further interesting to note that the scorekeeping practice of professional baseball also works against the manager, since the game focuses on the scoring of runs and the advancement of runners as opposed to the defensive aspects of that same play, which is in line with with Kahneman and Tversky's (1979) statement that "losses loom larger than gains". This implies that managers may be more risk averse than necessary. They fear the negative repercussions of their actions, whereas, if they focused on the positive/defensive outcomes of their decisions, they would be less risk-averse.

6.2 Representativeness Bias

Ideally, one should assess probability based on exact historical information. In the case of professional baseball, if a manager is interested in understanding the ability of a given batter, the manager must ensure that he is taking all variables into account that will affect the on-field outcome. This type of analysis must include not only an analysis of the batter's ability versus that of the given pitcher but also the exact scenario, where the scenario includes runner, outs, and secondary information such as inning, stadium, fatigue and an athlete's career curve as a representation of skill level.

However, it has been argued that the representative bias will dissuade managers from seeking exact information in exchange for representative information, that is, information that is representative of the present scenario, but is not necessarily similar. Just as important, the representative bias addresses scenarios where no exact historical situation can be found and related scenarios must be substituted.

The notion of representativeness implies basing a judgment on facts and signs, similar to those in the present situation, that one compares in order to make an objective, as opposed to subjective, decision. In order to forecast the outcome of alternatives, individuals often compare the situation at hand versus base case scenarios that are "representative" of the problem presently faced. But the term "representative bias" refers to a distinction between what is representative and what is probable. For example, J. Tenenbaum and T. Griffith state that 'being divorced four times' is more representative of a Hollywood actress than is 'voting democratic,' but in reality the former is certainly less likely. The issue, then, is to determine what the base-case scenarios are, and which ones are most representative.

With respect to baseball and the IBB phenomenon, managers will often compare the present situation to some historical scenario as a means of capturing the most likely outcome of each alternative. The manager may use his own experience, the historical stats of the batter, the historical statistics of the pitcher, or any of a multitude of statistical compilations.

Tversky and Kahneman (1982b) state that people often evaluate the probability of an uncertain event or a sample 'by the degree to which it is (i) similar in essential properties to its parent population and (ii) reflects the salient features of the process by which it is generated'.

Using the first criterion – that of parent population – Tversky and Kahneman state that the situation at hand will be compared to similar historical situations. However, as baseball is rich in statistics, it is difficult to know which statistics are the most important and relevant to the situation at hand. Does one focus on the positive categories of hitters such as hits and homeruns, or does one focus on the positive categories of pitchers such as wins and strikeouts? Does one focus on the negative categories of hitters such as strikeout and double-plays, or does one focus on the negative categories of pitchers such as hits allowed and walks. With each individual statistic, there are a multitude of computations and permutations that create even more numerical data. It is easy for managers to lose track of all this information.

Similarly, a manager may inadvertently group his hitter in a larger category of hitters with similar historical statistics and base a decision on the abilities of the category given the same situation. For example, a manager facing a situation that involves a young hitter who holds a high batting average may utilize historical information comparing the present hitter to a prototypical hitter with a high career batting average, someone such as Wade Boggs, Tony Gwynn and Paul Molitor.

Tversky and Kahneman's second criterion for determining the probability of an uncertain event – that of the "salient features of the process" –cannot be directly used to either support or weaken the given hypothesis. It entails an analysis of what items will initially jump into the cognitive awareness of the manager and which instances are easiest to recall. Salience may play itself out as a recent situation where a manager chose to impart a base-on-balls and its subsequent outcome. This aspect of representativeness may have a cross-over effect with biases such as availability and confirmation bias, which I will discuss in subsequent sections.

The overall concern with this bias is that managers, overcome with the pressure of the situation and the plethora of statistical data, may be blinded to the true base-case scenarios to which they should be comparing the situation at hand. Kahnman and Tversky (1972), discuss the bias' similarity to the gambler's fallacy. The fallacy suggests that a fair coin that has turned up HHH, is more likely to turn up T on its next flip than it is likely to turn up H. Similarly, a batter who has recorded 5 successive hits is more likely to record an out during their next at-bat.

The gambler's fallacy differs from the *hot-hand* hypothesis put forward by Gilovich et al (1985), which states that a batter who has recorded 5 successive hits is more likely to record a hit during their next at-bat. This hypothesis suggests that an individual's skill, coupled with confidence, can lead them into a streak of either good or bad outcomes.

A hypothetical example would be a batter with a lifetime batting average of .200, with 100 career at-bats, who had a .600 batting average against a pitcher in 5 career at-bats. The hot-hand hypothesis would suggest that the batter is likely to get a hit in his at-bat against the pitcher, whereas an objective observer might suggest otherwise.

I would suggest that these biases — the gambler's fallacy and the hot hand syndrome — can co-exist in people's beliefs. However, that individual will choose the bias that best suites his or her experience and personality. A risk-taking baseball manager may select the gambler's fallacy because he is 'taking a chance' that something different will occur in the next at-bat, while a risk-averse individual might look at trends and therefore choose to be influenced by the hot-hand syndrome.

6.3 Confirmation Bias/Availability Bias

When faced with a scenario that may or may not call for an IBB, the manager must use information available to him to make the ultimate decision. In this situation, the manager must decide which information to use and how to interpret it. These two questions can be answered by the “availability bias” and by the “confirmation bias”. These two biases presuppose that the manager already has a hypothesis as to which solution would be most effective in this situation. This hypothesis is a byproduct of other biases and psychological/sociological phenomena. Risk aversion, path dependence, regret theory and representativeness typically contribute to the foundation for a hypothesis geared to the impartment of an IBB.

The availability bias answers the first question: ‘What information does the manager chose from?’ More specifically, the availability bias states that one uses information that is readily available, and easier to recall, as a means of solving more complex problems. For example, Tversky and Kahneman (1982a) use an example in which one may assess the divorce rate in a given community by recalling divorces among one’s acquaintances. Similarly, a manager may allow information that is readily available to him to bias his decision-making process. Given that Barry Bonds is constantly in the headlines for his offensive prowess and that managers may be constantly questioned as to how they will pitch to him, managers may be overly fearful of his ability and consequently adopt strategies that overcompensate.

Similarly, when facing a potential IBB situation, a manager may recall instances that have a more powerful recall effect and that therefore bias his decision-making. For example, if Barry Bonds beat a manager last night with a homerun, the manager may be more willing to intentionally walk Bond in a similar situation, even though the probability of him hitting a homerun has not changed. This is similar to the fallacy of misleading vividness, which suggests that vivid events tend to have more probability assigned to them than is really there. Whereas objective analysis would suggest an emphasis on statistical evidence, subjective analysis will be blurred by a particularly dramatic event

The confirmation bias states that, when faced with a decision, an individual will seek out information that confirms their hypothesis and will not seek out information that refutes it (Klayman, 1995). Therefore, when originally seeking information, which in the case of baseball is often in the form of historical statistics, the manager may only seek out that information that confirms his original hypothesis.

An example might be head-to-head statistics between the hitter and the pitcher. Suppose that the overall, career-historical head-to-head statistics favour the pitcher and suppose that this season's historical head-to-head statistics favour the batter. Should his original inclination be to intentionally walk the batter, the manager may overweigh the significance of the latter piece of information and underweigh the significance of the former, independent of their relative likelihoods.

The confirmation bias also answers the question: 'How does he choose to interpret it?' It deals with information that is ambiguous with respect to the hypothesis. The theory of confirmation bias states that, when faced with a decision, an individual will interpret ambiguous information in favour of his hypothesis (Klayman, 1995). In the above example of overall career head-to-head statistics, the manager will find a way to either (a) internalize the information in favor of the hypothesis, or (b) discredit the information. One way would be to say that either the pitcher's effectiveness has declined over his career, or the batter's ability to read the pitcher has improved and therefore, only recent information should be taken into account.

This issue is further complicated by the fact that the sport of baseball is permeated by statistics. There are infinite ways of analyzing a particular situation and just as manner ways of distorting those numbers to fit one's convenience. When determining whether to intentionally walk a batter, a manager could focus on any number of statistics. Example might be the age of parties involved and their relative decline in production, weather conditions as they relate to player's performance and momentum generated by one team during the game. In addition there are a number of available statistics to consider in the course of making a decision: Overall batting statistics, batting statistics when facing the same situation, overall pitcher's statistics, pitcher's statistics when facing the same situation, overall head-to-head statistics for the pitcher, head-to-head statistics for the pitcher when facing the same situation, etc. Even when the decision regarding which statistics is made, the manager must also consider whether to use all-time statistics, season-to-date statistics, that week's statistics or that day's statistics. Given the game's obsession with statistics and the plethora of available comparisons, it is probable that one can support any hypothesis by using historical records.

6.4 Regret Theory/Omission-Commission

Life involves making decisions between available options. Typically, an individual will weigh the options based on information available and, based on predetermined criteria, choose the option he or she hopes will maximize utility. In some cases, the path chosen turns out to be the optimal path, and in other cases, the chosen path is suboptimal. Much research has shown that individuals, choosing the suboptimal path, will experience a negative emotion in response to their choice: regret (Zeelenberg). The basis of regret is that the individual had the opportunity to choose a better alternative yet, for one of many reasons, did not and now find themselves in a worse position. Zeelenberg (1999) defines regret as a negative, cognitively-based emotion that we experience when realizing or imagining that our present situation would have been better, had we decided differently. Zeelenberg argues that the fear of anticipated regret will lead to risk-averse decision-making.

A manager, facing a scenario that may merit an intentional base-on-ball would be forced to decide between two options, (a) to impart an intentional base-on-ball or (b) not to impart an intentional base-on-ball. Should the outcome of the decision be poor, the manager may feel regret with respect to their decision. Zeelenberg (1999) would argue that a manager, fearing the eventual backlash of regret, would be inclined to lean towards the risk-minimizing decision to impart an intentional base-on-ball, since this decision decreases the likelihood of short-term defeat, that is, defeat by the player presently at bat).

In a second article Zeelenberg et al (2002) breaks regret into two separate categories. The first category is commission (or hot regret) and the second category of omission (or wistful regret). Hot regret is the direct emotional reaction to the outcome, while wistful regret is the less intense emotion associated with pleasantly sad fantasies of what might have been. The paper argues that regret is more intense in the case of commission than in the case of omission. In a related paper, Gilovich and Medvec (1995) reinforce Zeelenberg's argument by suggesting that that actions tend to generate more regret in the short term, but inactions tend to be more troubling in the long run.

In similar fashion, a manager faces the potential of two suboptimal results when deciding upon an IBB. If the manager chose to impart an IBB — an act of commission — and the outcome was negative, he might feel a negative sense of hot regret as his direct action led to a suboptimal result. The manager might wonder whether his directive was the cause of the result. However, if the manager chose not to impart an intentional base-on-ball and the outcome was negative — an act of omission — the manager might feel a negative sense of wistful regret, since his lack of action led to a suboptimal result. The manager would subsequently wonder whether the outcome would have differed had he imparted the IBB.

Zeelenberg (2002) would argue that a manager would opt towards not imparting the IBB (omission), as the action would minimize the potential regret. The author suggests that the regret associated with omission is less than the regret associated with commission as commission is a direct result of a conscious decision that can be directly associated to a single decision-maker. Gilovich and Medvec (1995) similarly argued that a manager would opt towards the omission option because it would minimize his regret in the short term. Omission minimizes regret in the short term as it takes the blame and responsibility out of the manager's hand and puts its squarely on the ability of the players.

While the above arguments seem to reflect the earlier hypothesis, it should be noted that an article by Ritov and Baron (1995) maintains that an individual will be biased towards commission in cases where the outcomes of the option not chosen will never be known. As it is impossible to know the outcome of unselected alternatives in baseball, it can be argued that managers would be biased towards an IBB. Therefore, walking the hitter when in doubt is the best alternative.

6.5 Path Dependence

Suppose you have ten dollars in your pocket. Would you feel any different if the money fell out of your pocket as opposed to your being robbed of it? Path dependence says that the manner in which the action (losing the $10) occurs will have different effects on the individual and will result in different levels of happiness/sadness. I call this "path dependence."

In order to relate path dependence to baseball, consider the following scenario: It is a tie game in the bottom of the ninth inning. There are two outs and a runner on second. A team's best hitter is up with the team's worst hitter on deck. What do you do? Logic would argue that you walk the present batter and pitch to the team's worst hitter. Now imagine two scenarios. In the first scenario the manager walks the initial, better batter, but the second batter (the worst hitter) hits a game-winning homerun. In the second scenario the manager chooses to pitch to the batter and he hits a homerun. Which loss would feel worse to the manager? Will the manager's initial decision be biased to decrease the pain associated with one type of losing over another?

The problem with this question is inherent in the circular logic. In the above example, using the case where the best hitter produced the homerun, the managers might be upset at themselves as they should have known better and gone with the rational decision. However, in the case where the worst hitter produced the homerun, the manager might be upset that his team allowed the worst hitter to beat them, realizing that it was a fluke scenario. In this scenario, would the manager be better off being beaten by talent or by luck?

With respect to baseball, I would argue that the manager would intentionally walk the stronger hitter, as the core issue is attribution of blame. In the initial scenario (luck), the attribution is external and lays blame on the pitcher who was unable to retire a weak batter or to the luck of the batter. In the latter scenario (talent), the expectation on the batter is high and therefore the attribution would be internal. The manager was aware of the batter's ability and chose to ignore it, thus calling his own decision-making ability into question.

6.6 Incentive Bias/Agency Problem

As of 2004, George Steinbrenner had owned the New York Yankees for 29 years. Over that period the Yankees had 21 different managers. In 1977, the Texas Rangers became the first team to have four different managers in the same season. For this reason, the office of a major-league baseball manager has often been compared to a 'revolving door' due to the tremendous number of hiring and firings that occur each season. Although it is ultimately the players that must perform in a game situation, it is the managers that receive the bulk of the criticism. Ownership often argue that, while it is their responsibility to build the initial network of players, it is the responsibility of the manager to get the most out of those players through preparation, motivation and strategic play-calling. Managers are ultimately responsible to the general manager and club ownership for the results of their team's performance. This situation is at the heart of the agency problem, with the manager acting as the agent to the club's ownership (the principle).

I hypothesize that the nature of this relationship leads to an incentive bias that may propel a manager to act against the objective norm and act, not in the best interest of the team, but in their own personal self-interest.

This follows from a paper written in by Eisenhardt (1989) that states that "the agency problem arises when (a) the desires or goals of the principal and agent conflict and (b) it is difficult or expensive for the principal to verify what the agent is actually doing" (page 58). In the case of professional baseball, both the ownership and manager are focused on the overall objective of winning. However, it is often argued that while the team owners is focused on short-term success, the manager should be more focused on building long-term success. This may play out in a situation where a manager is willing to continue with a struggling pitcher to build that player's confidence, whereas the owners would prefer the manager to act in the best interest of the game at hand.

Secondly, as there is no communication between a manager and the owners during a particular game, owners may not be able to understand the inherent logic built into the manager's strategy. Similarly, hindsight may affect the agency problem, since the owner will chose to assign blame as opposed to admit defeat. Related to the confirmation and availability bias, owners will voluntarily seek out information to confirm their hypothesis that the manager's decisions were poor in order to support their original hindsight bias.

Finally, as Major League Baseball is built around one-on-one, head-to-head, pitcher-to-batter confrontations utilizing the world's top professional baseball players, there is no guarantee of offensive success even if the manager's strategy is perfect. Over the past 65 years, only one hitter has been able to hit safely in over 40% of their at-bats over the course of a single season. Ted Williams' remarkable batting average of .407 in 1941 still renders a 59% chance of failure. Ownership often overlooks the opposing team's caliber of players and the objective likelihoods of success when passing judgment on the manager.

Given the presence of the agency problem, and given the overzealous owners who seem to keep their finger of the proverbial 'fire manager' button, there is an obvious incentive for a manager to act in their own best interest. Managers may be inclined to direct their players in a manner that spotlights the success and failures of the players and de-emphasizes the manager's own contributions. With respect to the intentional base-on-ball, managers may feel inclined to walk a particular batter in a situation where they believe pitching is the best strategy because they fear the criticism of the owners should their own strategy not meet with success.

This issue is further complicated by the emergence of sports media coverage. Whether in newspapers, magazines, television shows or websites, commentators are extensively analyzing a manager's tactics and frequently subjecting the manager to unnecessary blame and scrutiny.

7.0 Conclusion

The intentional base-on-ball represents an important strategic tool that provides an advantage to those managers who are better able to understand its importance and how best to use it. Whereas existing literature supports the use of the IBB at an aggregate level, this paper extended that analysis to include an individual, head-to-head perspective utilizing data analysis techniques such as regression and simulation.

The present study compiled data from a 25-year period and analyzed it to uncover key relationships between variables previously undocumented. While the outcomes were statistically significant, the model was limited by the sample size (it used only batters and pitchers with great than 1000 at-bats or 1000 batters faced), input variables (items such as fielding and stadium size were disregarded) and computer applications (accuracy of applications). Further analysis can be conducted extending each of these factors, thus creating a more complete tool.

While the initial purpose of the model was to create a decision-making tool, prescribing instances in which an IBB was the preferred strategic technique, the model was also used to evaluate the historical use of the same technique.

The model illustrated the overuse of the IBB by managers in accordance with the paper's initial hypothesis. The paper concluded with an introduction to the reasons for the overuse, described in terms of subjective biases and cognitive errors. In this respect, further research can be conducted to test those reasons and determine the significance of each variable through direct testing of actual baseball managers.

The importance of this research lies in its ability to improve the decision making process of baseball managers and limit future Type I errors. Through the use of this model, subjective biases and cognitive errors can be reduced, thus allowing managers to make smarter and more informed decisions.

Appendix

Appendix A - 24 Scenarios					
	Runners			Outs	
24 Scenarios	First	Second	Third	One	Two
1	0	0	0	0	0
2	0	0	0	1	0
3	0	0	0	1	1
4	1	0	0	0	0
5	1	0	0	1	0
6	1	0	0	1	1
7	0	1	0	0	0
8	0	1	0	1	0
9	0	1	0	1	1
10	0	0	1	0	0
11	0	0	1	1	0
12	0	0	1	1	1
13	1	1	0	0	0
14	1	1	0	1	0
15	1	1	0	1	1
16	1	0	1	0	0
17	1	0	1	1	0
18	1	0	1	1	1
19	0	1	1	0	0
20	0	1	1	1	0
21	0	1	1	1	1
22	1	1	1	0	0
23	1	1	1	1	0
24	1	1	1	1	1

Appendix B - Post At-Bat Scenario								
	At-Bat Outcomes							
24 Scenarios	**Single**	**Double**	**Triple**	**Homerun**	**Walk**	**Out**	**Strikeout**	**HBP**
1	4	7	10	1	4	2	2	4
2	5	8	11	2	5	3	3	5
3	6	9	12	3	6	End	End	6
4	13	16	10	1	13	5	5	13
5	14	17	11	2	14	6	6	14
6	15	18	12	3	15	End	End	15
7	13	7	10	1	13	8	8	13
8	14	8	11	2	14	9	9	14
9	15	9	12	3	15	End	End	15
10	4	7	10	1	4	11	11	4
11	5	8	11	2	5	12	12	5
12	6	9	12	3	6	End	End	6
13	22	19	10	1	22	14	14	22
14	23	20	11	2	23	15	15	23
15	24	21	12	3	24	End	End	24
16	22	19	10	1	22	17	17	22
17	23	20	11	2	23	18	18	23
18	24	21	12	3	24	End	End	24
19	22	7	10	1	22	20	20	22
20	23	8	11	2	23	21	21	23
21	24	9	12	3	24	End	End	24
22	22	19	10	1	22	23	23	22
23	23	20	11	2	23	24	24	23
24	24	21	12	3	24	End	End	24

(Please note that ‘End’ represents the end of an inning, or three outs).

Appendix C – Single with Runner on Second						
		Runner Destination - Percent				
Fielder	**# of hits**	**Out**	**Second**	**Third**	**Home**	**Bases/hit**
N/A	1052	1.5%	2.9%	41.5%	54.1%	1.51
LF	5117	4.5%	0.2%	26.9%	68.4%	1.68
CF	5476	2.4%	0.1%	14.8%	82.6%	1.82
RF	4374	4.3%	0.0%	23.9%	71.7%	1.72
INF	2380	4.2%	7.3%	76.6%	12.0%	1.05
Total	18399	3.6%	1.2%	29.9%	65.3%	1.64

Appendix D – Double with Runner on First						
		Runner Destination - Percent				
Fielder	**# of hits**	**Out**	**Second**	**Third**	**Home**	**Bases/hit**
N/A	673	1.5%		50.8%	47.7%	2.46
LF	3118	2.6%		56.9%	40.5%	2.38
CF	1160	4.6%		36.8%	58.6%	2.54
RF	2021	3.6%		58.8%	37.7%	2.34
INF	25	8.0%		80.0%	12.0%	2.04
Total	6997	3.1%		53.6%	43.3%	2.40

Appendix E – Single with Runner on Third						
		Runner Destination - Percent				
Fielder	**# of hits**	**Out**	**Second**	**Third**	**Home**	**Bases/hit**
N/A	541	0.2%		2.2%	97.6%	0.98
LF	2840	0.1%		0.0%	99.9%	1.00
CF	3080	0.1%		0.0%	99.9%	1.00
RF	2566	0.1%		0.0%	99.9%	1.00
INF	1106	0.4%		7.2%	92.4%	0.92
Total	10133	0.1%		0.9%	99.0%	0.99

Appendix F – Double with Runner on Second						
		Runner Destination - Percent				
Fielder	**# of hits**	**Out**	**Second**	**Third**	**Home**	**Bases/hit**
N/A	**399**	**0.0%**		**1.5%**	**98.5%**	**1.98**
LF	1975	0.2%		0.7%	99.2%	1.99
CF	772	0.1%		1.6%	98.3%	1.98
RF	1317	0.2%		2.4%	97.5%	1.97
INF	18	0.0%		11.1%	88.9%	1.89
Total	4481	0.1%		1.4%	98.4%	1.98

Appendix G - Lead Runner Destination/Outcome Chart				
	At-Bat Outcomes		Probability	
24 Scenarios	Single	Double	Single	Double
4	13	19	65.2%	53.6%
	16	7	31.3%	43.3%
	5	8	2.1%	3.1%
	4		1.4%	0.0%
5	14	20	65.2%	53.6%
	17	8	31.3%	43.3%
	5	9	2.1%	3.1%
	6		1.4%	0.0%
6	15	21	65.2%	53.6%
	18	9	31.3%	43.3%
	6	End	2.1%	3.1%
	End		1.4%	0.0%
7	4	7	65.4%	100.0%
	16		29.9%	0.0%
	5		3.6%	0.0%
	13		1.2%	0.0%
8	5	8	65.3%	100.0%
	17		30.9%	0.0%
	6		2.6%	0.0%
	14		1.2%	0.0%
9	6	9	65.3%	100.0%
	18		30.9%	0.0%
	End		2.6%	0.0%
	15		1.3%	0.0%
13	13	19	65.7%	100.0%
	22		30.3%	0.0%
	14		4.0%	0.0%
14	14	20	65.7%	100.0%
	23		30.3%	0.0%
	15		4.0%	0.0%
15	15	21	65.7%	100.0%
	24		30.3%	0.0%
	End		4.0%	0.0%

Appendix H: Number of Outs								
	At-Bat Outcomes, Outs							
24 Scenarios	**Single**	**Double**	**Triple**	**Homerun**	**Walk**	**Out**	**Strikeout**	**HBP**
1	0	0	0	0	0	1	1	0
2	1	1	1	1	1	2	2	1
3	2	2	2	2	2	3	3	2
4	0	0	0	0	0	1	1	0
5	1	1	1	1	1	2	2	1
6	2	2	2	2	2	3	3	2
7	0	0	0	0	0	1	1	0
8	1	1	1	1	1	2	2	1
9	2	2	2	2	2	3	3	2
10	0	0	0	0	0	1	1	0
11	1	1	1	1	1	2	2	1
12	2	2	2	2	2	3	3	2
13	0	0	0	0	0	1	1	0
14	1	1	1	1	1	2	2	1
15	2	2	2	2	2	3	3	2
16	0	0	0	0	0	1	1	0
17	1	1	1	1	1	2	2	1
18	2	2	2	2	2	3	3	2
19	0	0	0	0	0	1	1	0
20	1	1	1	1	1	2	2	1
21	2	2	2	2	2	3	3	2
22	0	0	0	0	0	1	1	0
23	1	1	1	1	1	2	2	1
24	2	2	2	2	2	3	3	2

Appendix I - Number of Runs Scored								
	At-Bat Outcomes: Runs Scored							
24 Scenarios	Single	Double	Triple	Homerun	Walk	Out	Strikeout	HBP
1	0	0	0	1	0	0	0	0
2	0	0	0	1	0	0	0	0
3	0	0	0	1	0	0	0	0
4	0	0	1	2	0	0	0	0
5	0	0	1	2	0	0	0	0
6	0	0	1	2	0	0	0	0
7	0	1	1	2	0	0	0	0
8	0	1	1	2	0	0	0	0
9	0	1	1	2	0	0	0	0
10	1	1	1	2	1	0	0	1
11	1	1	1	2	1	0	0	1
12	1	1	1	2	1	0	0	1
13	0	1	2	3	0	0	0	0
14	0	1	2	3	0	0	0	0
15	0	1	2	3	0	0	0	0
16	1	1	2	3	1	0	0	1
17	1	1	2	3	1	0	0	1
18	1	1	2	3	1	0	0	1
19	1	2	2	3	1	0	0	1
20	1	2	2	3	1	0	0	1
21	1	2	2	3	1	0	0	1
22	1	2	3	4	1	0	0	1
23	1	2	3	4	1	0	0	1
24	1	2	3	4	1	0	0	1

References

Bernoulli, D. 1954. Exposition of a new theory on the measurement of risk. *Econometrica*, 22:23-36.

CBC Sportsline, "Various locations depending on 10,000 player profiles", *http://www.CBSSporstline.com.*

Eisenhardt, K.M. 1989. Agency Theory: An assessment of review. *The Academy of Management Review*, 14:57-74.

Gilovich, T., & Medvec, V.H. 1995. The Experience of Regret – What, When, and Why. *Psychological Review*, 102(2):379-395.

Gilovich, T., Vallone, R., & Tversky, A. 1985. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17:295-314.

Hanrahan, Tom. 2001. Does Good Hitting Beat Good Pitching. *By The Numbers: The Newsletter of the SABR Statistical Analysis Committee*, Volume 8.

Jarvis, John F. 1999. An Analysis of the Intentional base-on-ball. Presented at SABR-29, Phoenix, AZ. (http://pacer1.usca.sc.edu/~jfj/IBBanalysis.html)

Kahneman, D., Knetsch, J., & Thaler, R.H. 1991. Anomalies: The Endowment Effect, Loss Aversion, and the Status Quo Bias. *Journal of Economic Perspectives*, 5(1): 193-206

Kahneman, D., & Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica*, 47: 263-291

Kahneman, D., & Tversky, A. 1972. Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology*, 3:430-454.

Klayman, Joshua. 1995 Varieties of confirmation bias, pp 385-418. In Busemeyer, Jerome, Hastie, Reid, and Medin, Douglas, Decision Making from a Cognitive Perspective, *The Psychology of Learning and Motivation*, Vol 32, San Diego, CA: Academic Press.

Levitt, Dan. 1999. The Batter/Pitcher Matchup. *By The Numbers: The Newsletter of the SABR Statistical Analysis Committee*, Volume 2. http://www.pathstoglory.com/danwritings.html

Levitt, Dan. 1999. Hitters and Baserunner Advancement. *By The Numbers: The Newsletter of the SABR Statistical Analysis Committee*, Volume 2. http://www.pathstoglory.com/danwritings.html

Retrosheet, "Various locations depending on 10,000 player profiles", *http://www.Retrosheet.org*.

Ritov, I., & Baron, J. 1995. Outcome Knowledge, Regret, and Omission Bias. *Organizational Behaviour and Human Decision Processes*, 64(2): 119-127.

Samuelson, W. F., & Zeckhauser, R. J. 1988. The Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty*, 1: 7-59.

Tenenbaum, J., & Griffith, T. The Rational Basis of Representativeness. *Department of Psychology, Stanford University*. (jbt,gruffydd @psych.stanford.edu)

Tversky, A., & Kahneman, D. 1982a. Availabiliy: A heuristic for judging frequency and probability. In D. Kahneman, P. Slovic, & A. Tversky (Eds.)*, Judgement under uncertainty: Heuristics and biases*, 84-98.

Tversky, A., & Kahneman, D. 1982b. Judgement of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*, 84-98. Cambridge: Cambridge University Press.

Zeelenberg, M. 1999. Anticipated regret, expected feedback and behavioural decision making. *Journal of Behavioural Decision Making*, 12(2): 93-106.

Zeelenberg, M., van den Bos, K, van Dijk, E., & Pieters, R. 2002. The Inaction Effect In the Psychology of Regret. *Journal of Personality and Social Psychology*, 82(3): 314-327.