2016

Pitch quantification Part 1: between pitcher comparisons of QOP with conventional statistics

Jason Wilson
Biola University

Pitch Quantification Part 1: Between-Pitcher Comparisons of QOP with Conventional Statistics

Jason Wilson[1,2]

1. Introduction

The Quality of Pitch (QOP) statistic uses PITCHf/x data to extract the trajectory, location, and speed from a single pitch and is mapped onto a -10 to 10 scale. A value of 5 or higher represents a quality MLB pitch. In March 2015 we presented an LA Dodgers case study at the SABR Analytics conference using QOP that included the following results[1]:

1. Clayton Kershaw's no hitter on June 18, 2014 vs. Colorado had an objectively better pitching performance than Josh Beckett's no hitter on May 25th vs. Philadelphia.
2. Josh Beckett's 2014 injury followed a statistically significant decline in his QOP that was not accompanied by a significant decline in MPH.

These, and the others made in the presentation, are big claims. Can they be substantiated? Is QOP a valid statistic? Could it be a reliable new stream of information about pitch quality? The purpose of this paper is to provide evidence to MLB analysts, analyst-types, and fans of the affirmative.

We originally developed QOP with the desire to provide a more comprehensive measurement for individual pitches that could be used by pitching coaches for player development and by scouts to identify player potential. The analytical and medical applications in our presentation came later. Before we can go about the business of developing QOP for such high-level applications, however, the MLB analysts with whom we have spoken are asking for evidence of the validity of QOP. *How does it compare with conventional statistics like ERA and WHIP and FIP, or BABIP and OPS? Other results? Is it consistent? Is it better?* This paper presents that comparison. I will make a case that while QOP values can be shown to associate with traditional pitch classification at a macro level, direct comparisons of QOP with conventional statistics have low correlation due to the difference between the statistics and the loss of variation inherent in averaging hundreds of pitching and hitting events. However, when QOP is compared with raw results data -- number of bases earned per hit -- and conditioning by pitch designation, it is shown to strongly associate with results. Additionally, I demonstrate a positive way to relate QOP to ERA is as a predictor. QOP has a 76 to 97% success rate for predicting a pitchers' drop in ERA or FIP in subsequent seasons for pitchers with QOP average over 5.0.

In order to be as clear as possible at the outset, let me remark on what QOP is, and is not. My colleagues and I have designed the QOP function to calculate the three key properties of a pitch which

[1] Biola University, 13800 La Mirada, CA 90639; jason.wilson@biola.edu

are: trajectory, location, and speed. Specifically, these properties are determined from the following components, shown with their relationship to QOP:

1. Increased rise → lower QOP (for curveballs)
2. Increased total break → higher QOP
3. Increased late vertical break → higher QOP
4. Increased horizontal break → higher QOP
5. Closeness to corners of strike zone → higher QOP
6. Increased velocity → higher QOP

Thus, the calculation of QOP is completely objective because it incorporates the above objectively measurable pitch components. There are no subjective evaluations in these calculations[2]. As the old English measurement system had inconsistent lengths between the "foot" measurement of different merchants, so contemporary major league baseball has inconsistent assessments of pitch quality between observers. As the "foot" inconsistency was resolved by standardized ruler, so we would like QOP to become a "ruler" of pitch quality for baseball. Therefore, the QOP model does not include pitch tempo, pitch sequence, or pitch intention. Neither do we consider pitcher deception or batter anticipation to be measurable properties for inclusion in QOP. We include only measurable components in the QOP "ruler".

In our theory of pitch quantification, we distinguish between pitch quality (QOP), pitch context (count, tempo, sequence, base runners, etc.), and pitch effectiveness (results). Pitch quality is the foundational issue. To the degree that the measurement of pitch quality is flawed, the estimation of batter effects and prediction of results will be flawed. We hypothesize that when accurate pitch quality is correctly combined with pitch context, then pitch effectiveness/results can be accurately modeled. QOP is our attempt to measure pitch quality accurately. In this paper, we are not addressing the relationship between pitch quality and pitch context. Such development of our theory remains for future work. In order to get there, the prior issue is the establishment within the baseball analytics community of the validity of QOP, which is the focus of this paper. Once QOP, an accurate measurement of pitch quality, is established as a valid metric, future development can occur.

2. Overview of QOP

I begin with a brief description of how QOP is calculated and how it differs from conventional statistics. Those already familiar with QOP may skip this section. The formula for QOP is a multiple regression model with the following form (see Figure 1 for a graphical representation of components):

$$QOP = \beta_1 * rise + \beta_2 * break.point + \beta_3 * vert.break + \beta_4 * horiz.break + \beta_5 * location + \beta_6 * MPH$$

where

1. **rise**: the number of feet the ball rises vertically to the maximum height
2. **break.point**: the horizontal distance (in feet) from the release point to the maximum height
3. **vert.break**: the number of vertical feet from the maximum ball height to the point it crosses the plate

4. **horiz.break:** the number of horizontal feet, at home plate, from the "straight line" crossing point of the ball to the actual crossing point
5. **location**: a measurement of how easy the pitch is to hit, on average, based upon the location in the strike zone (the higher the value, the easier to hit, see Figure 2)
6. **MPH:** pitch speed

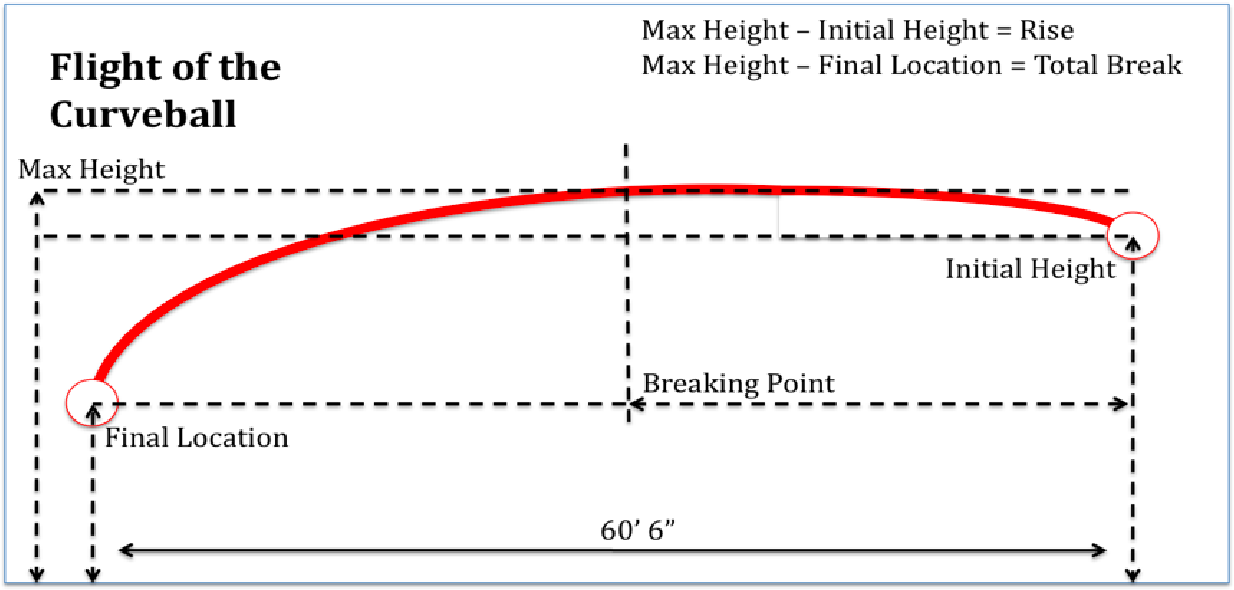

Figure 1. Curveball trajectory diagram. The trajectory components of a pitch used to develop the initial QOP model. In the model shown, the parameters were derived from the pitch measurements using a low-tech method. The current model uses PITCHf/x data.

```
# Residuals:
#    Min      1Q  Median      3Q     Max
# -2.1066 -0.4282 -0.0656  0.3274  3.3782
#
# Coefficients:
#            Std. Error t value  Pr(>|t|)
#   rise        0.69040  -2.809  0.00519 **
#   break.point 0.01837  12.956  < 2e-16 ***
#   vert.break  0.01964  50.342  < 2e-16 ***
#   loc         0.01537 -65.053  < 2e-16 ***
#   horiz.break 0.08130   4.050  6.1e-05 ***
#   ---
# Residual standard error: 0.6724 on 419 degrees of freedom
# Multiple R-squared:  0.9578,  Adjusted R-squared:  0.9573
# F-statistic:  1903 on 5 and 419 DF,  p-value: < 2.2e-16
```

Table 1. Statistical software output for QOP model. It may be noticed that MPH does not appear in the output. This is because we combine this trajectory and location parameter model with MPH separately.

The *rise* and *location* coefficients are negative, while the *break.point, vert.break, horiz.break*, and *MPH* are all positive. This makes sense, because a rising pitch and bad location make for a low quality pitch whereas a big break and fast speed make for a good quality pitch. Under extensive model testing, there were no quadratic relationships or interactions found. It is noteworthy that this is a zero intercept model. The model fit is excellent with a p-value < 10^{-5} for all coefficients (except *rise*, which is 0.005), and adjusted R^2 = 0.957. All QOP work has been done in R[3] and the model output is in Table 1. Unfortunately, the patent pending model is proprietary and the coefficients cannot be released at this time.

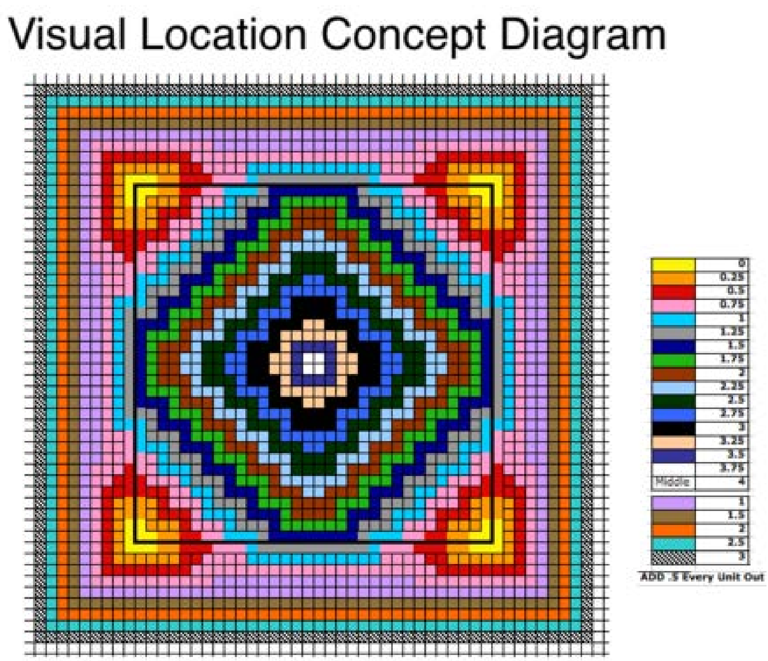

Figure 2. Diagram used for developing the location variable in the current QOP model. Different colors indicate different points assigned to the location parameter. The best location is near the strike zone corners, and near the edge of the strike zone. The worst location is the exact center of the strike zone, and far out of the strike zone. The edge of the strike zone, inside and out, rates well because it can lure the batter into swinging, while still difficult to hit. This is a conceptual diagram only; this is NOT the exact way location is calculated for the QOP model.

3. Results

How does QOP compare with pitching statistics like FIP and batting statistics like OPS? These questions form the subject of the first two sub-sections, below. Since QOP provides a number for every pitch, we can look at either an average or the individual pitch values. To distinguish these, we let **QOPV** refer to a set of individual pitch values and **QOPA** refer to their average. While QOPV is normally on a -10 to 10 scale, there are no theoretical limits to the endpoints, since it is a regression model. Some particularly bad pitches score negative numbers, and a few exceptional pitches score over 10 (the model has rated five pitches at or above 10 since 2008). Naturally, QOPA in MLB typically falls in the 3 to 6 range, depending on the subject.

3.1 QOPV by Pitch Classification

Our first dataset consists of n=9990 pitches thrown in 2014. A random subset was used instead of the entire dataset for simplicity and robustness. Each of the three analyses shown below was replicated, with the same results, on two additional samples of around 10,000 pitches -- one from 2014 and another from 2013.

Let us begin at the highest level of the data. Figure 3 shows boxplots of QOPV for each pitch designation. At a glance, different designations have higher QOPV than others. The rankings all make sense. The blue boxes are the 'bad' pitches: *Pitchout, Intentional Ball, Hit By Pitch, Ball in Dirt*, and *Balls*. It is clear that the bulk of these pitches are well below the quality of the others. The green boxes are the different strikes. The pink boxes are those in which the batter made contact with the ball, but did not get any bases. The red boxes are those where bases were made. The Called Strikes have the highest median QOP. It will be shown later that this highest average is real, and cannot be explained by randomness.

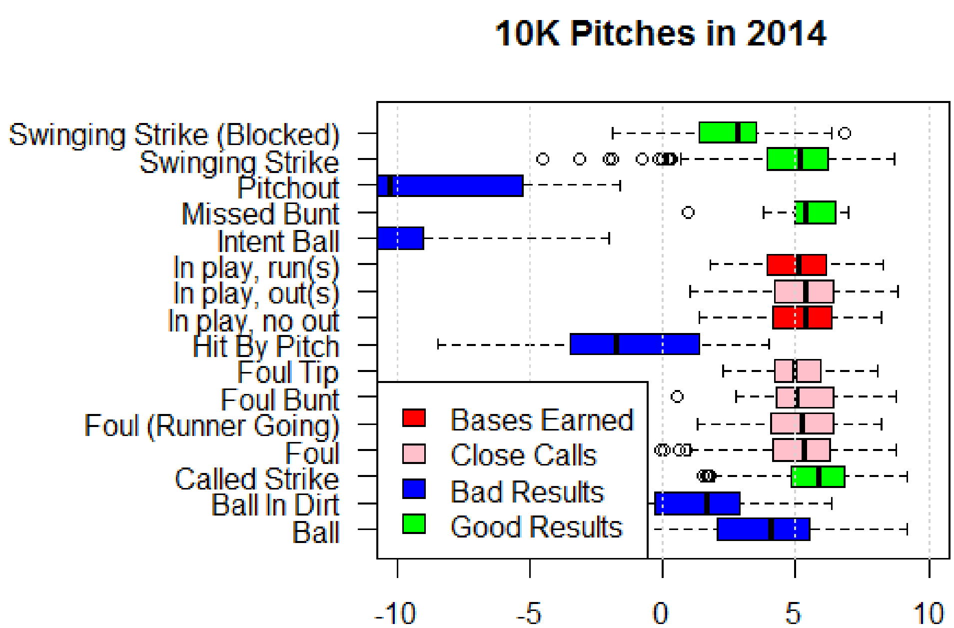

Figure 3. Boxplots of QOPV by pitch classification. Each box shows the range of QOPV for its pitch type. A box contains the middle 50% of the data, for that pitch type, with the dark line in the middle at the median. The remaining 50% of the data is split evenly on each side. The colored labels are from the perspective of the pitcher. The bulk of the Bad Results (blue) are below the quality of the others. The green Called Strikes have the statistically significant highest QOPV.

Next, let us focus on the most important pitches: *Called Strike*; *In play, no out;* and *In play, run(s)*. Figure 4 shows a multiple histogram of the QOPV of each, which offers a clearer view of the higher QOPA. On average, the best pitches resulted in strikes. The second best were in play, but no runs. The third best were in play, with runs. Again, this ranking agrees with intuition.

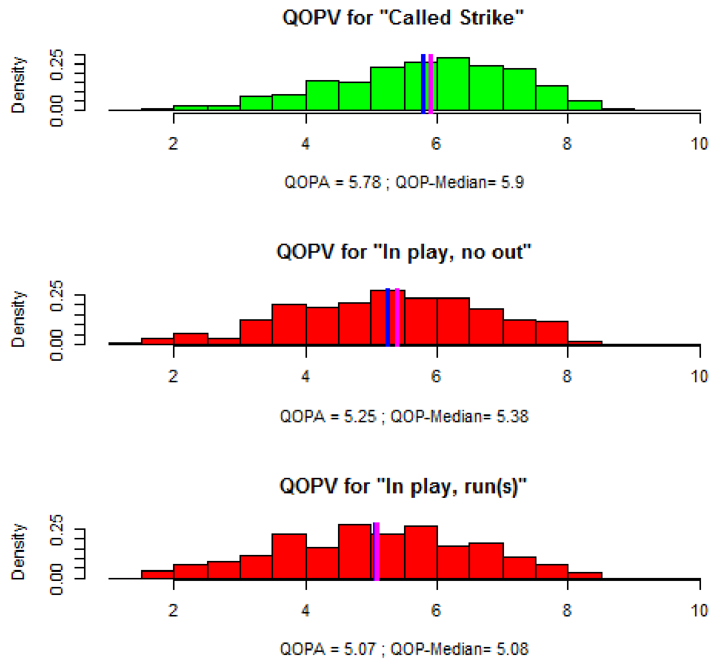

Figure 4. Multiple histogram of QOPV for the most important three pitch designations. The "Called strikes" are the best, followed by "In play, no outs", and lastly by "In play, run(s)".

Finally, I ran an ANOVA test on *Called Strike, Swinging Strike, Foul, In play-no out, In play-out(s),* and *In play, run(s)* (F = 39.07, p-value < 2×10^{-16}) to distinguish whether the differences observed above were due to real differences in pitcher performance, or mere randomness. The p-value means there is less than a 2×10^{-16} probability that the differences observed would occur if there actually were no differences. To see where the specific differences occurred, I submitted the designations to Tukey's Highest Significant Difference (HSD) test. The results in Figure 5 show that QOPA for *Called Strikes* are significantly higher than the *In play* pitches as well as *Fouls* and *Swinging strikes*. The differences in QOPA for the other categories are not distinguishable.

The above analyses, the macro-level pitch classification and the specific QOPA comparison between the best designations, show that QOP performs rationally.

3.2 QOPV Compared with Conventional Statistics

One of the benefits of QOP is that it offers a numeric value for every pitch. The richness of this information, however, makes it difficult to compare with conventional statistics like ERA, FIP, BABIP, and OPS. When such comparisons are made they do not reveal much, presumably due to the loss of information caused by conventional statistics averaging over so many pitches. Comparisons, however, with directly observable data, like number of bases per batted ball in play, reveal significant association with QOPV, as we will show. For technical reasons, when working with conventional statistics, I am using 2013 data with about 2/3 of the pitchers and batters.[4]

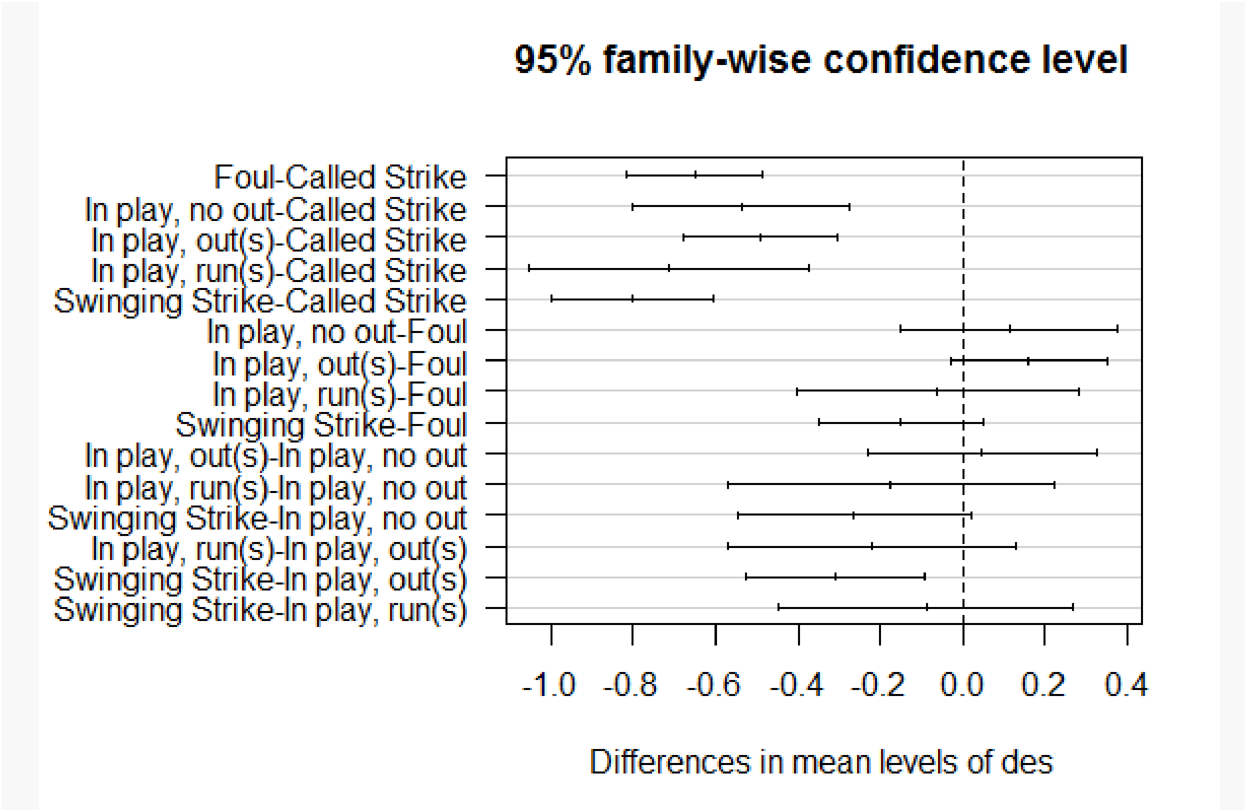

Figure 5. Confidence intervals of the difference between pitch designations for the Tukey HSD hypothesis tests. Each bar represents the 95% confidence interval for the difference in QOPA between the two pitch designations. If the interval overlaps zero, there is no statistically significant difference. All of the differences with Called Strike are below zero, meaning Called Strike has higher QOPA than the categories compared.

Starting with pitching statistics, FIP is a respected conventional statistic for pitcher performance. Since pitchers normally have one FIP statistic per season, FIP cannot be directly compared with QOPV for a single player. A pitcher's QOPA may be computed during the same season, however, and multiple players may be compared. Using 593 pitchers from the 2013 season, I find low negative correlation between the pitchers' QOPA and their FIP (Spearman's rho = -0.13). See the scatterplot in Figure 6. The correlation is significant (p-value = 0.002). This means that higher QOPA goes with a lower FIP, which makes sense, but the relationship is weak. An examination of other pitcher statistics, such as ERA, strikeout rate (SR), and walk rate (WR) similarly reveal little to no correlation, see Table 2.

	QOPA	ERA	FIP	SR	WR
QOPA	1.000	-0.114	-0.126	-0.060	-0.123
ERA	-0.114*	1.000	0.640	-0.333	0.215
FIP	-0.126*	0.640	1.000	-0.559	0.419
SR	-0.060	-0.333	-0.559	1.000	-0.066
WR	-0.123*	0.215	0.419	-0.066	1.000

Table 2. Correlation coefficients of QOPA vs. pitching statistics. There is a weak or correlation between QOPA and the pitching statistics. Spearman's correlation was used. An * by correlation in the QOP column indicates the correlation is statistically significantly different than zero. The expected sign depends on the nature of the statistic, e.g. it is negatively correlated with FIP. It should be positively correlated with the strikeout rate (SR), but it is not, although it should be noted that that correlation is statistically equivalent to zero.

The reader may be wondering, "If QOPA doesn't correlate with FIP, or other pitching statistics, then it's no good, right?" Wrong. FIP, ERA, WR, SR and frankly all of the conventional statistics are *highly batter dependent*. Therefore, they represent a combination of both pitcher and batter factors. QOPA, on the other hand, is *batter independent*. Therefore, it is devoid of the variation due to the batter. For example, we can have a well executed pitch (high QOPV), with a hit by a good batter. This counts neutral for FIP, SR & WR. Conversely, we can have a poorly executed pitch (low QOPV), with a hit. This counts neutral for FIP, SR & WR. These two features will "average out" in QOPA, to some extent. On a technical note, attempting to dig further, I did not find any noteworthy interactions when regressing SR, WR, and other such statistics onto QOPA, or fastball-only QOPA, or off-speed-only QOPA.

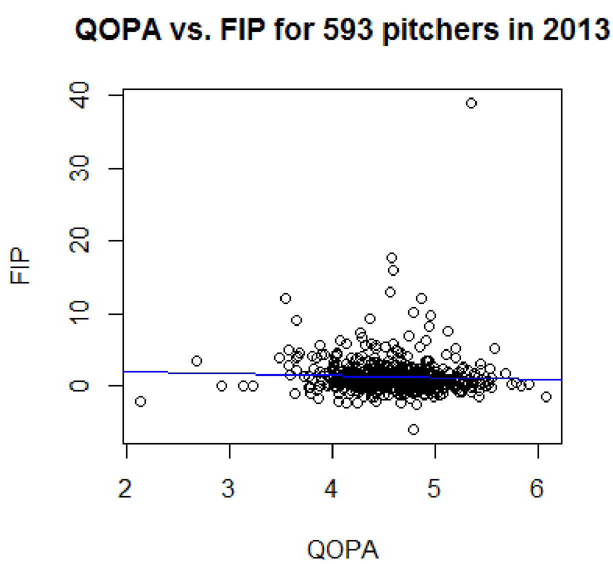

Figure 6. Scatterplot of QOPA vs. FIP with trendline. There is a weak negative correlation between QOPA and FIP (Spearman's rho = -0.126, p-value = 002).

How about batting statistics? As with pitching statistics, since batters typically have one statistic per season, it is not possible to directly compare the QOPV of a single pitcher-batter match. For the next best thing, I took a random sample of 10,000 pitches, from which I took only those which the batter swung at, plus called strikes[5], and compared their QOPV with the batting statistics of the batter who faced each pitch. For example, suppose Clayton Kershaw pitches twice to Andrew McCutchen (who swung or got a called strike) with a 4.7 and 6.8 QOPV, and Josh Beckett also pitches twice to McCutchen with a 3.7 and 7.8 QOPV. Since McCutchen's 2013 batting average was 0.317, we would have four datapoints: (4.7, 0.317), (6.8, 0.317), (3.7, 0.317), and (7.8, 0.317).

The correlations are found in Table 3. It can be seen that the correlation between QOPV and all statistics are nearly zero, which is not perhaps what one might expect. For a sample visual, see the scatterplot of QOPV vs. OPS in Figure 7. The points are pretty spread out, no noticeable trend. There is no correlation.

The reader may be wondering, "If QOPV doesn't correlate with batter performance, then what good is it?" Like with the pitching statistics, some of the variation is being "averaged out". Here, each QOPV is preserved, but the same batting statistic is used for each batter for their numerous at bats, irrespective of whether they were having strong or weak performance for that pitch. I do not think that explains the zero correlation, though, since there still is a good amount of variation preserved by QOPV. Rather, the reason for the low correlation has to do with the nature of batting performance. Suppose that QOPV was perfectly correlated with the batter statistic BABIP. Then the high QOPV pitches would be hit most by high BABIP batters and the low QOPV pitches would be hit least by high BABIP batters. Conversely, high QOPV pitches would be hit least by low BABIP batters and low QOPV pitches would be hit most by low BABIP batters. But that is not what happens, is it? High BABIP batters probably hit *all QOPV pitches most* and low BABIP batters probably hit *all QOPV pitches least* (on average). For a given range of QOPVs faced by a particular batter, his BABIP is plotted at the same level on the y-axis. For a range of BABIPs, this would make a somewhat rectangular shape, which would have no correlation. For example, look again at the hypothetical two pitches from Kershaw and Beckett to McCutchen again: (4.7, 0.317), (6.8, 0.317), (3.7, 0.317), and (7.8, 0.317). The points plot on a horizontal line, which is the graph of zero correlation!

	QOP	BA	PA	TB	SlugPct	OBP	OPS	BABIP	wOBA
QOP	1.000	-0.006	-0.003	0.000	0.004	-0.009	-0.002	-0.007	-0.002
BA	-0.006	1.000	0.536	0.644	0.713	0.815	0.799	0.792	0.799
PA	-0.003	0.536	1.000	0.955	0.533	0.547	0.574	0.305	0.568
TB	0.000	0.644	0.955	1.000	0.723	0.633	0.741	0.409	0.726
SlugPct	0.004	0.713	0.533	0.723	1.000	0.719	0.964	0.561	0.942
OBP	-0.009	0.815	0.547	0.633	0.719	1.000	0.868	0.670	0.898
OPS	-0.002	0.799	0.574	0.741	0.964	0.868	1.000	0.636	0.995
BABIP	-0.007	0.792	0.305	0.409	0.561	0.670	0.636	1.000	0.645
wOBA	-0.002	0.799	0.568	0.726	0.942	0.898	0.995	0.645	1.000

Table 3. Table of correlations between QOPV and batting statistics. Spearman correlations are shown (Pearson correlations are about the same). QOPV, in the way used, has no correlation with batting statistics, which is explained by the nature of the comparison, rather than a lack of relationship.

3.3 QOPA Compared with Number of Bases Per Hit

So, what data does QOPV correlate with? Let us look at raw results – the number of bases run on a hit by a batter, or **Bases**, for short. Thus, Bases takes on only the values of 1, 2, 3, or 4 for a single, double, triple, or home run, respectively. This data has the advantage that it is not a statistic (summarized data); it can be directly compared with each QOPV; and it is a direct reflection of the quality of the batting performance. In theory, we would expect high QOPV pitches to have the lowest Base and low QOPV pitches to have the highest Base, on average. For this analysis, I used the 2015 regular season data from April to July (the time of writing). Since we are looking at the rarest and most valuable data – the pitches which resulted in hits that the batter got on base (including home runs), this limits us to only 26011 of the 438807 pitches recorded by PITCHf/x in this time frame. I am not using a subset, like the previous analyses, because the dataset is already dramatically reduced and I want it large enough to satisfy the assumptions of the chi-square test.

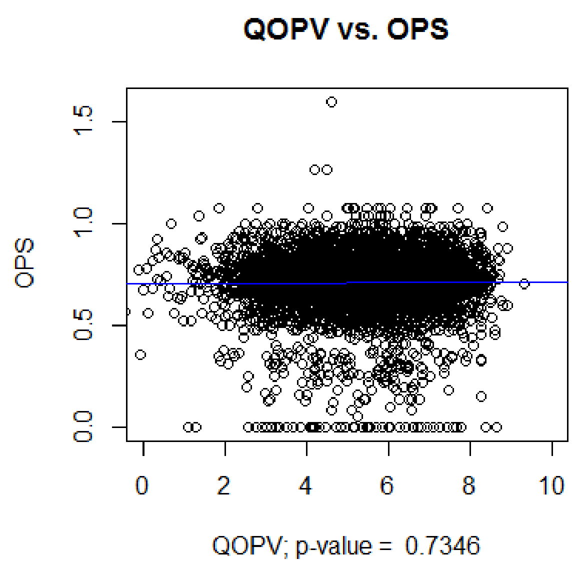

Figure 7. Scatterplot of QOPV vs. OPS for 10,000 pitches during the 2013 season. There is essentially no correlation between QOPV and OPS.

To begin, Table 4 shows the proportion of Bases by groups of QOPV. Notice that the QOPV 8-10 pitches have 6.6% home runs and 74.4% singles. By comparison, QOPV 0-2 pitches have 16.7% home runs and 61.8% singles. Thus, the highest QOPV pitches which were hit achieved the lowest results while the lowest QOPV pitches which were hit achieved the highest results. The same proportions may be viewed more readily in Figure 8. Whether using Table 4 or Figure 8, it can be seen that the proportions follow this consistent pattern for the other QOPV groups as well.

The data of Table 4 was submitted to a chi-squared test of independence (X^2 = 221.6, df=12, p-value < 2×10^{-16}). The results are highly statistically significant, meaning that the pattern of higher QOPV resulting in fewer bases is strong in the population. The same pattern holds, with p-value < 2×10^{-16} for all seasons from 2008 to 2014.

		Number of Bases				
		1	**2**	**3**	**4**	**Sum**
	0-2	0.618 (178)	0.205 (59)	0.010 (3)	0.167 (48)	1.00 (288)
QOPV	**2-4**	0.621 (3222)	0.201 (1042)	0.027 (141)	0.151 (782)	1.00 (5187)
Groups	**4-6**	0.662 (7251)	0.196 (2143)	0.022 (241)	0.120 (1312)	1.00 (10947)

	Number of Bases					
	6-8	0.711 (6412)	0.185 (1671)	0.019 (174)	0.084 (757)	1.00 (9014)
	8-10	0.744 (428)	0.170 (98)	0.019 (11)	0.066 (38)	1.00 (575)

Table 4. Proportion (Count) of bases by groups of QOPV. The proportions in each cell are with respect to the QOPV group by row. For example, the first entry of 0.618 represents 178 singles from pitches with 0-2 QOPV pitches out of a total of 288 0-2 QOPV pitches, 0.618=178/288.

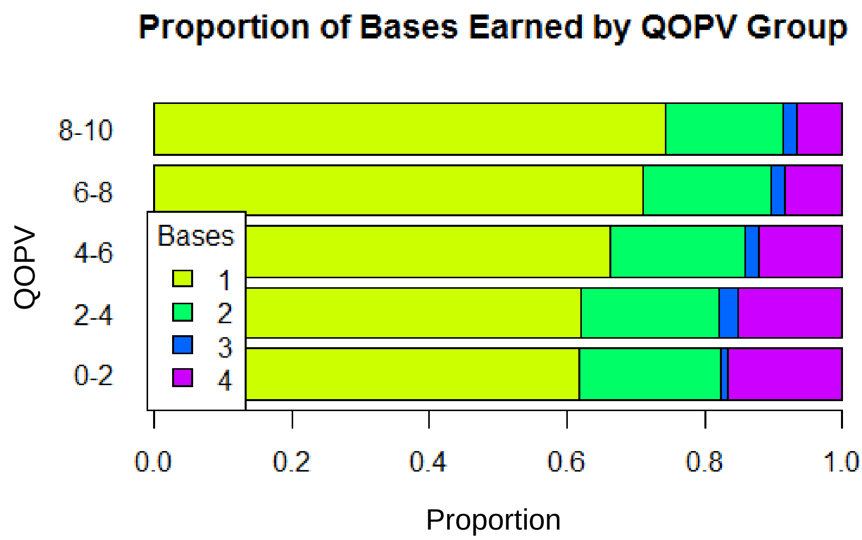

Figure 8. Proportion of bases run by QOPV group. Each bar represents 100%, which is divided between the proportion of hits which resulted in 1,2,3, or 4 bases. The bars are the rows from Table 4. The higher the QOPV, the lower the bases.

Looking at the data differently, we can quantify the relationship between QOPV and Bases. Figure 9 shows a boxplot for each number of bases. The super-imposed regression line covers the median of each box with a negative slope. In particular, starting from 5.58 QOPV, as the number of bases increases by one, the QOPV drops 0.16 points, on average (p-value < 2×10^{-16}). The correlation between QOPA and Bases is as strong as possible (-1 using Spearman's correlation, -0.98 using Pearson's, with QOPAs 5.42, 5.28, 5.19, 4.94 for Bases 1,2,3,4).

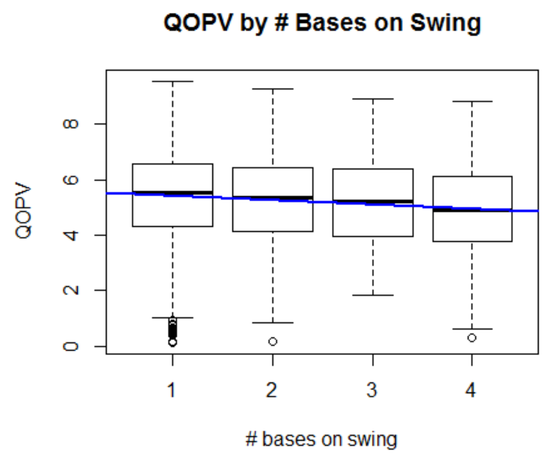

Figure 9. Boxplots of QOPV by Bases. The Linear regression line super-imposed shows a negative trend with very strong correlation between QOPA and bases (Spearman's correlation = -1.00).

Figure 10 is telling. Each row is the current pitches for the 2015 season of a particular designation. The designations are in the conceptual order of Called Strike, hit balls (In Play-Out(s), In Play-Run-No-Out, Fouls), missed balls (Missed Bunt, Swinging Strikes), and out of zone pitches (Balls, Pitchout, Hit by Pitch). The Called Strike has the highest proportion of 8-10 QOPV and the lowest proportion of 0-2 QOPV. Conversely, the Out of Zone Pitches have the lowest proportion of 8-10 QOPV and the highest proportion of 0-2 QOPV. The other categories and proportions lie in between these extremes. The table differences are highly statistically significant (chi-square test of independence p-value < 2×10^{-16}, see Table 5). The same pattern holds for each season from 2008 to 2014.

	OutZone	SwStrike	MisBunt	Foul	IPRuns	IPOut	ClStrike
0-2	23171	1404	11	1147	312	564	372
2-4	38477	7820	53	15162	5657	9946	9510
4-6	49611	17776	150	35638	11969	23128	28476
6-8	24329	11561	119	25917	9971	18837	33570
8-10	1156	464	7	1488	635	1290	2750

	OutZone	SwStrike	MisBunt	Foul	IPRuns	IPOut	ClStrike
0-2	0.169	0.036	0.032	0.014	0.011	0.010	0.005
2-4	0.281	0.200	0.156	0.191	0.198	0.185	0.127
4-6	0.363	0.456	0.441	0.449	0.419	0.430	0.381
6-8	0.178	0.296	0.350	0.327	0.349	0.350	0.450
8-10	0.008	0.012	0.021	0.019	0.022	0.024	0.037

Table 5. Chi-square Frequencies and Proportions for QOPV vs. Pitch Designations.

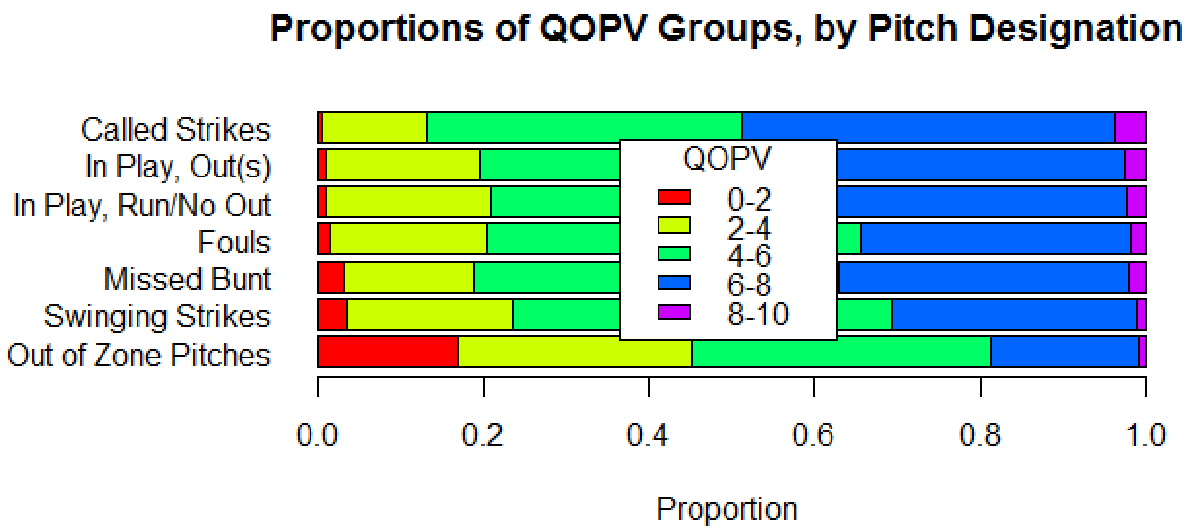

Figure 10. Barplot of proportions of QOPV by pitch designation for 2015. The 8-10 QOPV has the highest proportion of Called Strikes and the lowest proportion of Out of Zone Pitches (Balls, Pitchout, Hit by Pitch). Conversely, the 0-2 QOPV has the highest proportion Out of Zone Pitches and the lowest proportion of Called Strikes.

3.4 Using QOPA to Retrodict and Predict ERA & FIP

For our final line of approach, we will consider performance prediction. There are numerous ways and means which QOP could be used to predict pitcher performance, whether for one season to another or for a future game, or even for a single future pitch. We have not yet begun to explore the depth of the possibilities of this direction in our research. For the purposes of this paper, because my aim is strictly to demonstrate the viability of the statistic, I limit the study to a simple analysis. In this section I explore using QOPA for one season to **retrodict**[6] whether a pitcher's ERA will go down in the subsequent season. The data is the same as the previous section where ERA was used, which is about 2/3 of MLB pitchers per season[7]. The data run from 2008 to 2013, giving the ability to use performance from 2008 to 2012 to retrodict performance from 2009 to 2013, which is five seasons. The ERA data are in Table 6. An example of the full statistics for one of the retrodictions from 2010 to 2011 is in Table 8.

For the ERA data, Table 6 shows that, averaged over all seasons, for players with an ERA of 4.5 or higher in one season, if they also have a high QOPA (>4, 4.5, 5, or 5.5), then we can retrodict with around 75.9 to 83.4% accuracy that their ERA will go down in the subsequent season. The higher the QOPA, the higher the accuracy, although the fewer the players which meet this criteria. The same kinds of observations can be made for the FIP data in **Error! Reference source not found.**. Note that the For players with a FIP of 2 or higher in one season, if they also have a modest QOPA (>3, 3.5, 4, 4.5, or 5), then we can retrodict with 77.2 to 96.8% accuracy that their FIP will go down in the subsequent season. If predictions were made from the 2015 season to the 2016 season, I would expect the same results. The margin of error for such predictions is from 3 to 10%.

QOPA	ERA	08-09		09-10		10-11		11-12		12-13		Overall	
		n	p	n	p	n	p	n	p	n	p	n	p
4	4.5	164	0.713	156	0.737	124	0.766	108	0.796	122	0.811	674	0.759
4.5	4.5	102	0.706	87	0.782	68	0.809	60	0.817	84	0.821	401	0.781
5	4.5	20	0.850	16	0.750	13	0.846	12	0.750	24	0.875	85	0.824
5.5	4.5	2	1.000	0	NA	0	NA	1	1.000	3	0.667	6	0.834
4	5	118	0.788	115	0.765	81	0.753	78	0.833	92	0.837	484	0.793
4.5	5	74	0.730	61	0.787	46	0.804	45	0.822	60	0.867	286	0.797
5	5	14	0.786	7	0.714	**10**	**1.000**	9	0.778	15	0.933	55	0.855
5.5	5	0	NA	0	NA	0	NA	1	1.000	2	1.000	3	1.000

Table 6. Retrodicting next year's ERA with this year's QOPA and ERA. n is the number of pitchers that meet the QOPA and ERA criteria; p is the proportion of the n pitchers whose ERA was lower the following year. The first row of the table reads as follows: For the 2008 MLB season, there were 164 pitchers with an ERA of 4.5 or higher who also had a QOPA of 4 or higher. Of these 164 pitchers, 71.3% had their ERA go down in 2009. The 09-10 columns read similarly with the 2009 season data retrodicting the 2010 results, and so on. To illustrate, the data behind QOPA>5, ERA>5 for 2010-2011 (numbers in red) is shown in Table 8. Note: As described above, the number of pitchers (n) is given out of about 2/3 of all MLB pitchers, NOT all MLB pitchers.

	08-09	09-10	10-11	11-12	12-13	Overall
ERA	0.555	0.501	0.495	0.581	0.513	0.530

Table 7. Baseline ERA drop percentages. The first entry, 0.555, means that 55.5% of all pitchers' ERAs from 2008 dropped in the 2009 season. The other columns are interpreted similarly.

In order to evaluate the retrodiction rates in Table 6, we need to know the baseline rate. The overall rate is 53.0% of all pitcher's ERAs (see Table 7), which is a little better than fifty-fifty[8]. Thus, using QOPA to retrodict ERA substantially increases the success rate. Such predictions could be done simply, without human bias, and could be used to identify undervalued potential.

Upon inspection, the statistics in Table 6 make rational sense. All of the overall statistics are monotonic. This means that the results are consistent, implying that the statistics are reliable and would be likely to have a consistent prediction record. More importantly, it also implies that the underlying QOPA is consistent, which is the primary thing we are trying to show. From a logical standpoint, it makes sense that if pitchers are pitching high quality pitches during a season, but this goes unrecognized for some reason in their ERA, that on average their ERA should come into alignment with their true pitching ability which was previously observed in their QOPA.

	2010						2011					
Pitcher	**QOPA**	**ERA**	**SR**	**WR**	**HitR**	**n**	**QOPA**	**ERA**	**SR**	**WR**	**HitR**	**n**
Aaron Harang	5.47	5.32	0.163	0.075	0.276	2045	5.00	3.64	0.172	0.081	0.243	3103
Blake Wood	5.13	5.07	0.141	0.1	0.245	881	5.11	3.75	0.205	0.106	0.218	1218
Chad Qualls	5.25	5.57	0.165	0.066	0.264	1087	5.13	3.51	0.141	0.065	0.239	1237
Chris Jakubauskas	5.26	27	0	0	0.5	12	5.19	5.72	0.157	0.088	0.281	1297
Felipe Paulino	5.15	5.11	0.202	0.112	0.231	1614	4.78	4.11	0.224	0.09	0.232	2410
Kyle Lohse	5.01	6.55	0.125	0.081	0.299	1704	5.03	3.39	0.143	0.054	0.23	3033
Nick Blackburn	5.28	5.42	0.098	0.058	0.28	2275	5.38	4.49	0.113	0.081	0.273	2445
Sandy Rosario	5.29	54	0	0.083	0.75	41	4.57	2.45	0.111	0.111	0.278	65
Tyson Ross	5.05	5.49	0.189	0.118	0.231	706	5.24	2.75	0.166	0.09	0.228	538
Waldis Joaquin	5.45	9.64	0.074	0.259	0.222	135	5.92	4.26	0.111	0.111	0.222	101

Table 8. Example of pitching statistics for the 2010 retrodicting 2011 results. This table shows the ten players from 2010 who had QOPA>5 and ERA>5, along with their performance in 2011. SR = strikeout rate; WR = walk rate; HitR = hit rate; n = number of pitches for the season. For this year, the retrodiction accuracy was 100%, as all ERAs went down, most by a large margin.

4. Conclusion

Quality of Pitch (QOP) is a new patent pending pitching metric which holds promise for quantifying pitches. QOPV was shown to match very well with all pitch designations, including Called Strikes having significantly higher QOPV than all other pitch designations. Even so, since there is one value per pitch, it only weakly correlates with conventional statistics. However, when compared with direct results data – number of bases earned per hit – the statistic is strongly correlated. In addition, QOPA was shown to effectively and consistently retrodict high QOPA pitchers' ERA from one season to another. Cumulatively, this provides multiple lines of evidence that QOP is doing what it claims to do, namely provide an accurate measurement of the quality of a single pitch. It does so in a novel way and therefore offers information not contained in the conventional statistics.

Throughout this article, I have compared QOP *between pitchers*. It was designed to do this for scouting and fan purposes and we have here offered evidence for its viability. Once QOP is established as a trustworthy statistic, it could be added to the FIPS (or perhaps even Batter Independent Statistics – BIPS?) and used that way. However, the more prominent purpose in the design of QOP was to compare QOPV *within a single pitcher* – for player development and injury prevention. Single pitcher comparison offers more variation due to pitches (e.g. examining video to discover why three curveballs pitched had QOPVs of 4.7, 5.7, and 6.7) and less variation due to other factors (e.g. physical condition and experience differences).

We have attempted to provide evidence of the validity of QOP exclusively using *between pitcher* approaches, because it is what we were most asked by MLB teams. The preceding was the path my research took in response to the inquiries. I offer it to the analytics community as a first effort in characterizing the relationship between conventional statistics and QOP in an effort to establish its validity. Some negative results were reported, not because they helped my case, but because we are searching for a characterization and I want to place a yellow "Caution" sign in front of the road of direct comparison to conventional statistics. Moving ahead, I look forward to the necessary feedback on this research which will help refine my understanding and suggest further avenues for exploration. One such avenue, which is both difficult and delicate, was hinted at in our March 2015 SABR presentation. That avenue is *within pitcher* analysis. We showed Josh Beckett's QOPV was in measurable decline preceding his 2014 hip injury (not detectable with MPH).[9] Extracting patterns to retrodict other known injuries has not been as easy to find, but it is one of the areas we are working on with this fertile new statistic.

---

[1] See presentation RP2: http://sabr.org/latest/2015-sabr-analytics-conference-research-presentations . The slides may be viewed at http://www.qopbaseball.com/mlb.html .

[2] It is a separate matter as to the choice for how to combine these properties in the QOP model. Since the model is proprietary, that is a conversation we would like to have with interested MLB analysts.

[3] www.r-project.org

[4] My source for conventional statistics is Sean Lahman's (http://www.seanlahman.com/baseball-archive/) database in R, which at the time of writing only has conventional statistics through 2013 (https://cran.r-project.org/web/packages/Lahman/index.html). The name formats between PITCHf/x data and the Lahman data vary somewhat. Using regular expressions, I was able to match about 2/3 of the pitchers and batters. The discrepancies are typically names with abbreviations, like A.J. Burnett. Therefore, in all results shown using conventional statistics, the players that did not make the list are missing for presumably random reasons. The R scripts used to generate the results of this paper will be posted on our QOP website, http://www.qopbaseball.com/, along with QOPA data, following the end of the 2015 baseball season. Note, however, that the QOPVs will not publicly available.

[5] 10,202 pitches, to be exact. I initially selected 11,000 random pitches, and then excluded those with missing trajectory data from PITCHf/x. Different random samples yielded the same results. When the balls and other pitches not swung at were removed, 5,595 pitches remained in the analysis.

[6] A retrodiction is a prediction, but it is based on data which has already occurred, not data which is yet to occur. For example, I used 2008 data to "predict" what happened in 2009 – and then I checked whether it was accurate or not. In order to distinguish between whether the data was known or unknown, statisticians use the term retrodiction.

[7] See Endnote 4.

[8] Why are the ERA and FIP rates not at 50%? It may be due to a few one or more of the following factors: (i) low ERA pitchers retiring, who may otherwise contribute an ERA increase (remove negative results), (ii) high ERA rookie

pitchers entering the dataset who improve their ERA (add positive results) . Both factors would push the percentages above the random rate of 50%.
[9] See slides 25 & 26 of the presentation at http://www.qopbaseball.com/mlb.html