Draft 2.0 4.18.25 [subject to revision]

# Empirically Testing the Impact of Pitchers on "Balls Hit in Play"

*Xavier Fünf**


## Abstract

This paper examines a question of long-standing contention in the study of major league baseball: whether pitchers vary in their influence to stifle the impact of balls hit in play. A hierarchical multivariate regression model was constructed that disentangled the effects of pitchers' "balls in play" propensities and team fielding, respectively, on runs allowed per game. Applied to the performances of all American and National League pitchers from 1912-2024, the model suggested that varying run-stifling BIP propensities exist and are highly reliable across pitchers (ICC [1,1] = 0.97 [95% CI: 0.94, 0.98]). The paper derives a metric for measuring the effect of these propensities: BIP-ERA. Pitchers with superior BIP-ERAs typically reduce their total runs allowed per game by 0.30-0.50; the very best have reduced their runs allowed per game by over 1.0, saving their teams 30-40 runs in particular seasons (multiple times in some cases). Of particular consequence in the early decades of the twentieth century, BIP propensities have nonetheless continued to contribute to the success of select pitchers into the twenty-first. The paper examines the impact of BIP-ERA across time and the cost of neglecting it on the calculation of pitcher WARs.


## Introduction

Do major league pitchers vary in their propensity to induce batters to hit into outs? The question of whether such a characteristic exists is of intense theoretical, practical, and historical significance.

Whether inducing low-quality contact is a characteristic of pitching proficiency is at the core of one of the most provocative theses associated with modern baseball analytics: the McCracken conjecture. Propounded over a quarter century ago, the conjecture asserts that "[t]here is little if any difference among major-league pitchers in their ability to prevent hits on balls hit in the field of play"; what happens after a batter makes contact with a pitched ball is determined by a combination of chance and the quality of a teams' fielders (McCracken, 2001**)**. Baseball analysts have for decades disputed this claim without decisively refuting it.

As a result, the best means for measuring pitcher proficiency remain uncertain. Should pitcher value be estimated on the basis of indicia of "fielding-independent pitching" (strikeouts, walks, home runs allowed), as McCracken's thesis implies, or on more inclusive criteria reflecting the outcomes of balls hit in play (BIPs)? The two major competing systems for calculating player WAR ("Wins above Replacement") strike opposing stances over this issue.

The same uncertainty hampers comprehension of how the basic economy of baseball run production has changed over time. Today's elite pitchers can be shown to rely on increasingly high strike out rates to stifle the similarly growing trend among batters to hit home runs. But high strike out rates were markedly less common, or critical to pitcher success, for most of the twentieth century. Without a way to measure the impact that pitchers exert on hitter contact, it is impossible to know whether inducing more readily fielded balls in play helped distinguish elite from mediocre pitchers in previous eras.

---

* Research associate, bbcardstats.com. Direct correspondence to xavier_funf@bbcardstats.com.

The aim of this paper is to help overcome these deficits in understanding. It describes the structure and results of a study aimed at measuring the effect (if any) of pitchers' "BIP propensities"—that is, their tendency to yield balls in play more or less amenable to being turned into outs. Based on evidence of such an effect, the study develops a metric—BIP-ERA—that quantifies the impact of variance in pitchers' BIP propensities on runs allowed per game. It also supplies individual pitcher estimates of season, decade, and career BIP "runs saved," which can then be added to the contribution made by other indicia of pitching proficiency (such as rates of strikeouts, walks, and home runs allowed) to determine a pitcher's independent net effect in suppressing runs.

## Background: a surplus of disagreement, a deficit of data

Despite the period of time that has elapsed since it was made, the McCracken conjecture cannot be said to have been persuasively rebutted. The principal evidence against it consists of a collection of diversely calculated season-to-season correlations for opponent "batting average for balls hit in play" (BABIP), reported both in absolute terms and in relation to pitchers' leagues' and their teams' BABIPs. Focusing on select trends or the performance of particular pitchers, detractors treat one or another set of correlations as sufficient to reject McCracken's position (Pavitt 2025, ch. 8, pp. 20-27).

To date, however, none of McCracken's opponents—or supporters, for that matter—has performed a multivariate regression analysis structured to disentangle individual-pitcher BIP propensities from team fielding. Without the benefit of this form of analysis or its equivalent, it is not genuinely possible to estimate either the reliability of these propensities or their impact—on BABIP, on runs allowed, or on any other outcome of interest. The significance of pitchers' BIP propensities thus remains a matter of theoretical uncertainty.

The unsettled nature of the question is reflected in the prevailing systems for assessing pitcher value. FanGraphs, the producer of one of the two leading WAR frameworks, sides largely with McCracken. It grounds its measure of pitcher value on fielding-independent pitching (FIP)—that is, pitchers' rates of strikeouts, home-runs allowed, walks, and hit batters (Tango, 2004; McCracken, 2001). FanGraph's only minor qualification of its resistance to treating BIP propensities as a source of pitcher value is an ad hoc adjustment for infield pop-ups, which it purports to treat as equivalent to strikeouts (Weinberg, 2017). But as the organization's analysts make clear, "At FanGraphs, our headline WAR number for pitchers is based on FIP" (Clemens, 2022).

In contrast, Baseball Reference, sponsor of the competing system, leans heavily in the opposite direction. It determines the number of runs that a pitcher saved relative to an average pitcher in two steps. First, it calculates the difference between the number of runs he allowed and the number an "average" pitcher would have been expected to surrender against the same opponents. Then, it subtracts the number of runs deemed saved for that pitcher by his supporting fielders—a quantity that equals the proportion of the team's total fielding-runs equivalent to the fraction of balls in play fielded while that pitcher was on the mound (Baseball Reference, undated). Because the pitcher's BIP propensities are implicitly included along with everything else that he is presumed to have done to suppress opponent scoring by that amount, this formula necessarily counts his "ability to generate weak contact" (Smith, 2024, p. 198).

While it is clear that FanGraph's and Baseball Reference's frameworks differ in the degree to which they credit pitchers with control over the outcome of balls in play, it is not possible to extract from either the precise weight their WAR tallies assign to BIP propensities. This feature of their systems stands in marked contrast to the exacting specificity with which they calculate the elements of player offense and fielding value.

The best available empirical assessment of pitching proficiency relies on FIP alone. By itself, FIP explains more variance in runs allowed per game than either Baseball Reference's or FanGraphs' WAR measure (despite the latter's purported central reliance on FIP [Fünf, 2024]). In addition, over the course of the last twenty-five years, FIP, it has been demonstrated, has become progressively more consequential. Measured at the team level, differences in FIP have now consumed nearly all impact on runs scored that differences in teams' fielding proficiency exerted for most of the twentieth century (Fünf, 2025*a*).

The enormity of FIP's explanatory power should be viewed as strong evidence in favor of McCracken's position. Indeed, the trend in the impact of FIP is itself a plausible consequence of the conjecture's influence in improving the decisions of those responsible for selection of major league pitchers.

Yet FIP's ascendency—if it indeed signals the inconsequence of BIP tendencies—generates its own unanswered questions. FIP has always mattered but not nearly to the extent it does today. It seems unlikely that perceptions of the dominance of particular pitchers from the 1920s to 1980s was based on illusion. If those pitchers—many of whom struck out only 4 or 5 batters per 9 innings, as opposed to the 9+ *average* of today's MLB hurdlers—were not differentiating themselves on the basis of their ability to stifle balls in play altogether, then what *other than* their ability to degrade hitter contact can possibly explain their success?

## Data and study design

The study featured in this paper was specifically designed to address the theoretical, practical, and historical questions posed by the McCracken conjecture. The basic analytical strategy was straightforward: to measure the impact of variance in pitchers' BIP properties on runs allowed *controlling* for the effects of *both* non-BIP elements of pitching performance *and* team-fielding proficiency.

***1. Sample and data.*** The sample consisted of all AL/NL pitchers from 1912 to 2024. Data on their performance came from Retrosheet (for all seasons), from the Lahman database (for 1912 to 2023), and from Baseball Reference (for 2024). In total, the sample comprised 47,016 individual season performances recorded by 8,979 individual pitchers.

The study begins with the 1912 season because that is the first one in which BIP data are available from Retrosheet. The types of balls hit in play (i.e., those that didn't result in a home run) were coded on the basis of the "event" field of Retrosheet's play-by-play reports. Along with the pitchers responsible for yielding them, BIPs were classified as grounders, fly balls, pop ups and line drives, and categorized as hit to either the infield or the outfield. Where an event entry did not explicitly indicate a ball-in-play type but specified enough information to support a confident inference, it was coded on that basis (e.g., putout by first baseman assisted by shortstop: infield groundout; unassisted putout by an infielder other than the first baseman: infield pop-up). If devoid of sufficient information to support a classification (e.g., "single"), the BIP was ignored.

Although the Project Scoresheet companion for Retrosheet indicates BIP types, it was decided to code the events from scratch based on evidence of the unreliability of Project Scoresheet data reported by Smith (2024). It was also decided to use Retrosheet reports for seasons after 2015, the point at which Statcast BIP data became available, because in the form published by FanGraphs the Statcast data do not differentiate between BIP types with the degree of specificity reflected in Retrosheet game reports.

As valuable a resource as Retrosheet is, its scoring of BIPs will inevitably be imperfect. Differences in judgments among and, frankly, the quality of individual scorers will compound the inherent difficulty of historically reconstructing play outcomes. Limitations of this nature, however, create noise, not bias; untethered to outcomes of interest, they cannot plausibly be expected to manufacture unreal effects in one direction or another but instead only to attenuate estimations of the size of any relationships that genuinely exist; whether such effects are robust enough to remain detectable nevertheless can be determined only by actual testing. Neither surmises about how data quality might have diluted effect sizes nor statistical remedies for offsetting them are included in the main paper. But certain issues and strategies of that nature are flagged and discussed in the Supplemental Information ("SI").

Pitchers' FIPs were computed on the basis of a regression model of runs allowed on per inning rates of strikeouts, home runs allowed, walks, and hit batters. Independent study had determined that this metric—which will referred to as $FIP_r$ for "FIP regression"—explains a substantially greater share of the variance in individual pitcher runs allowed than does the conventional FIP measure, reported in FanGraphs and Baseball Reference, which uses fixed, theory-derived weights rather than empirically derived ones for fielding-independent pitching outcomes ( (Fünf, 2025*b*). Separate models were fit for each season (SI).

To measure team fielding, the study relied on a composite measure. For seasons between 1912 and 1900, and between 2000 and 2002, that measure uses Total Zone Rating (TZR) (Smith, 2024). Those data were collected from Baseball Reference, which used it as the basis of its fielding WAR metric up until 2002. For the seasons between 1990 and 1999, the study composite measure uses Defensive Efficiency Rating (Smith 2024). DER is substituted for TZR because of the latter's attenuation by unreliable Project Scoresheet data for that time period (Smith 2024). For the seasons after 2002, the composite measure again uses DER. That measure was selected over alternatives (including Baseball Info Solution's Defensive Runs Saved and Statcast's Outs Above Average) because independent study had identified DER to be the team-fielding measure that, after controlling for FIP, explained the highest level of variance in team runs allowed over that period (Fünf, 2025*a*).

The ball-in- play variables included balls in play allowed per inning (BIP_IP) plus four additional ones relating to specific BIP types: infield pop-ups (IFPOP), infield ground balls (IFGB), outfield fly balls (OFFB), and outfield line drives (OFLD), measured as proportions of BIPs allowed. These types constitute over 90% of all BIPs. If pitchers have distinctive BIP propensities, they are likely to be reflected in consistent distributions of these types of balls in play.[1]

Runs allowed per 9 innings pitched—RAPG—was selected as the study outcome variable. Data were analyzed with a model that treated $FIP_r$, team-fielding, BIPs per inning, and rates of BIP types as predictors of pitchers' season RAPGs. Each pitcher's season $FIPr$ was the predicted score of the $FIP_r$ regression model associated with that season (SI). The team-fielding variable (FIELD) was the composite measure just described.

To assure that season-to-season variability did not distort study estimates, data were grouped into 22 five-season bins or periods for the seasons between 1912 to 2016, and one 8-season one for seasons between 2017 and 2024.[2] Preliminary testing suggested that five seasons were the minimal number

[1] The remaining types—outfield ground balls and infield fly balls—are implicitly treated as a reference group against which the effect of the others are assessed.
[2] The periods are (1) 1912-1916; (2) 1917-1921; (3) 1922-1926; (4) 1927-1931; (5) 1932-1936; (6) 1937-1941; (7) 1942-1946; (8) 1947-1951; (9) 1952-1956; (10) 1957-1961; (11) 1962-1966; (12) 1967-1971; (13) 1972-1976; (14) 1977-1981; (15) 1982-1986; (16) 1987-1991; (17) 1992-1996; (18) 1997-2001; (19) 2002-2006; (20) 2007-2011; (21) 2012-2016; (22) 2017-2024. When treated as variables in reported analyses, the periods are identified by their first year only.

needed to assure stable and precise measurement of the predictors. The alternative of using longer periods, reflective of different baseball "eras," was rejected on the ground that five-year bins could be expected to capture era effects with sufficient accuracy without the need to make discretionary choices among era specifications with potentially varying effects on model estimates.

***2. The basic model.*** The model used to analyze the data was a hierarchical or multi-level linear regression. The fixed effect predictors included the 22 periods, $FIP_r$, the team-fielding composite variables, BIP rate, and the four BIP types. This level of the model allowed the impact of BIP propensities on RAPG to be separated from fielding-independent pitching and team fielding within each of the multi-season periods.

A second, random-effect level was added for individual pitchers. A common element of analyses that involve repeated measures of individual characteristics or behavior, this aspect of the model avoided the possible biasing effect of treating multiple-season performances by particular pitchers as independent; assured that unmeasured sources of individual pitcher effects in suppressing runs was partialed out before the impact of BIP propensities was estimated; and created a ready basis for estimating the consistency of such propensities via an Intraclass Correlation Coefficient (Cohen et al., 2003, pp. 536-37, 579; Gelman & Hill, 2006, pp. 241-42), a measure of reliability more discerning than a simple Pearson's $r$ (Koo & Li, 2017; Cicchetti, 1994). This model will be referred to as "the basic model."

Both the basic model and the season-by-season regression models for computing pitchers' $FIP_r$s were weighted on pitchers' innings pitched. It was decided that the study should include all pitchers, regardless of innings, lest the selection of an arbitrary cutoff point confine analysis to more successful pitchers, who could be expected to differ from ones of more modest abilities in ways that might bias the model estimates (in particular by deflating the impact of BIP propensities). Moreover, the shrinkage effect associated with the random-effect component of the model, along the weighting of observations by innings pitched, was expected to mitigate any distortions associated with inclusion of pitchers with very low innings-pitched totals (Efron & Morris, 1977).

***3. Isolating BIP effects.*** The variables included in the basic model enabled implementation of the most important study objective: the disentangling of the full impact of a pitcher's run-suppressing influence from team fielding. Using the basic model, a pitcher's predicted runs allowed per game in any season can be compared to the predicted RAPG of a composite pitcher with the *same* team-fielding score but with an average $FIP_r$, an average BIP per inning rate, and an average set of BIP propensities (i.e., mean rates of infield groundouts, infield pop-ups, outfield fly balls and outfield line drives). The difference between these two reflects the sum total of the pitcher's individual responsibility—independent of his fielding support—for runs allowed.

This impact can be disaggregated into two components: a pitchers' ability to prevent balls hit in play; and his ability to mute the impact of those that are. The first is captured straightforwardly by measuring the influence of a pitcher's $FIP_r$ plus the residual RAPG impact of his BIP rate relative to that of an average pitcher.

The second can be derived from the incremental impact of the mix of BIP types a pitcher allows. Certain BIPS (particularly infield popups and outfield fly balls) can be expected to be associated with fewer runs allowed, and others (particularly outfield line drives) with more. A higher rate of the former, then, can be expected to reduce, and a higher rate of the latter to increase, a pitcher's RAPG. The size of these contributions will necessarily depend on the rate at which a pitcher yields balls in play. This relationship is operationalized by including in the basic model *interaction terms* for BIP_IP and the individual BIP-type variables: such terms measure the extent to which BIP rates magnify favorable or unfavorable mixes of BIPs allowed.

These distinct ball-in-play *suppressive* and ball-in-play *muting* components of a pitcher's influence can be characterized as his *BNIP-ERA* and *BIP-ERA*, respectively. They will be valenced in this study to reflect how many more runs than average a pitcher allows. A low BIP rate naturally reduces a pitcher's BNIP-ERA. The impact of a pitcher's BIP propensities determines the sign of his BIP-ERA: a negative BIP-ERA reflects, in effect, the fraction of expected BIP-associated runs per game a pitcher *recoups* through his capacity to induce low-quality contact; a positive BIP-ERA, in contrast, indicates that a pitcher's BIP propensities create a surplus of runs allowed in relation to his BIP rate. Added together, BNIP-ERA and BIP-ERA form a measure of runs allowed per game that more accurately reflects a pitcher's "fielding independent" impact than does FIP as traditionally measured. Or at least it will *if* BIP propensities have any measureable effect on RAPG.

## Results and analysis

*Basic model performance*

The results of the basic model are reported in Table 1. Because it is it is the sum of the BIP variables and their associated interaction terms that determine a pitcher's BIP propensities, Table 1 reports a separate test of the *joint effect* of IFGB, IFPOP, OFFB, OFLD and their interactions with BIP_IP. This effect was statistically significant in every period.

The incremental contribution of the BIP variables to variance explained can also be computed. The complete model $R^2$ was 0.71, indicating that it explained 71% of the variance in pitchers' RAPGs. When IFGB, IFPOP, OFFB, OFLD, and their associated BIP_IP interaction terms were removed, the remaining variables ($FIP_r$, BIP_IP, and FIELD, along with the study period variables) generated an $R^2$ of 0.62 ($\Delta R^2 = 0.09$ [0.95 CI: 0.02, 0.15]). The particular mix of balls in play yielded by individual pitchers in the sample thus explains 9% of the variance in RAPGs, adding about 15% to the explanatory power of a model that predicts pitcher runs allowed based on $FIP_r$, balls-in-play allowed, and team fielding alone. These results are consistent with the inference that variance in BIP propensities has a discernable impact on runs allowed, or at least have had such an impact over some portion of the seasons included in the model.[3]

The best strategy to clarify the import of the basic model parameters, though, is to *use* the model to create a visual illustration of how changes in the predictors of interest affect the outcome variable (King, Tomz, & Wittenberg, 2000). Figure 1 charts how a propensity to yield a relatively *pitcher-favorable mix* of balls in play as opposed to a relatively *pitcher-unfavorable* one would be expected to affect runs allowed per game, both for pitchers with *moderately low ball-in-play rates* and for pitchers with *moderately high* ones. To simulate "relatively favorable" and "relatively unfavorable" mixes, the model parameters were set, respectively, at one standard deviation above and one below the sample mean rates of *run-stifling BIPs*, and at one standard deviation above and one below the sample mean rates of *run-productive* ones: this equates to roughly 12% versus 6% infield popups; 25% versus 17% outfield fly balls; 44% versus 32% infield grounders; and 17% versus 25% outfield line drives. A moderately high

[3] Interaction terms were also added to the basic model to determine whether these effects varied between the American and National Leagues during the span of time in which only the former had the DH. These terms had a negligible (and statistically nonsignificant) effect. They were thus not included in the final model.

ball in play rate is 3.1 per inning, and a low one 2.5—+1.0 and -1.0 SDs from the sample means. The basic model was used to estimate these impacts separately for each multi-season period.

***Fixed effects***

**pre-BIP effect variables**

| | period | | FIPR | | FIELD | | BIP_IP | |
|---|---|---|---|---|---|---|---|---|
| 1912 | | | **0.72** | (15.38) | **-0.20** | (-5.08) | **0.18** | (4.18) |
| 1917 | **0.14** | (2.49) | **0.66** | (16.03) | **-0.16** | (-4.18) | **0.28** | (6.50) |
| 1922 | **0.78** | (11.56) | **0.71** | (18.94) | **-0.26** | (-8.39) | **0.20** | (5.59) |
| 1927 | **0.93** | (11.81) | **0.75** | (18.91) | **-0.22** | (-6.15) | **0.17** | (5.15) |
| 1932 | **0.78** | (9.14) | **0.77** | (23.12) | **-0.14** | (-4.24) | **0.24** | (8.11) |
| 1937 | **0.66** | (6.96) | **0.87** | (27.73) | **-0.30** | (-9.02) | **0.05** | (1.50) |
| 1942 | 0.02 | (0.17) | **0.81** | (22.64) | **-0.33** | (-10.39) | **0.09** | (3.02) |
| 1947 | **0.57** | (5.30) | **0.87** | (31.42) | **-0.24** | (-7.82) | **0.06** | (1.91) |
| 1952 | **0.24** | (2.14) | **0.88** | (28.23) | **-0.19** | (-6.97) | **0.04** | (1.13) |
| 1957 | 0.16 | (1.30) | **0.79** | (29.26) | **-0.15** | (-6.31) | **0.15** | (5.40) |
| 1962 | -0.13 | (-1.07) | **0.78** | (30.12) | **-0.20** | (-7.87) | **0.14** | (4.73) |
| 1967 | **-0.32** | (-2.44) | **0.68** | (30.02) | **-0.14** | (-6.32) | **0.14** | (5.47) |
| 1972 | -0.23 | (-1.72) | **0.69** | (34.47) | **-0.16** | (-6.99) | **0.17** | (6.07) |
| 1977 | 0.05 | (0.36) | **0.70** | (34.38) | **-0.16** | (-8.46) | **0.21** | (9.43) |
| 1982 | 0.12 | (0.88) | **0.80** | (36.11) | **-0.18** | (-9.95) | **0.16** | (5.86) |
| 1987 | 0.09 | (0.65) | **0.78** | (33.24) | **-0.16** | (-8.50) | **0.45** | (14.15) |
| 1992 | **0.37** | (2.56) | **0.90** | (37.05) | **-0.15** | (-7.97) | **0.60** | (23.13) |
| 1997 | **0.44** | (3.02) | **0.92** | (37.51) | **-0.12** | (-7.23) | **0.59** | (18.75) |
| 2002 | 0.25 | (1.68) | **0.94** | (43.30) | **-0.09** | (-5.59) | **0.58** | (22.52) |
| 2007 | -0.01 | (-0.07) | **0.88** | (38.92) | **-0.15** | (-9.31) | **0.56** | (23.07) |
| 2012 | **-0.36** | (-2.40) | **0.93** | (41.68) | **-0.11** | (-6.41) | **0.53** | (23.61) |
| 2017 | -0.08 | (-0.53) | **1.20** | (48.96) | **-0.11** | (-7.30) | **0.59** | (24.23) |

**BIP variables**

| | OFLD | | IFPOP | | IFGB | | OFFB | |
|---|---|---|---|---|---|---|---|---|
| 1912 | **0.30** | (3.08) | -0.09 | (-1.35) | 0.02 | (0.16) | -0.06 | (-0.82) |
| 1917 | **0.20** | (2.79) | **-0.24** | (-3.73) | **-0.39** | (-3.60) | **-0.27** | (-3.75) |
| 1922 | **0.42** | (5.99) | -0.11 | (-1.86) | 0.09 | (0.93) | 0.02 | (0.34) |
| 1927 | **0.22** | (2.77) | **-0.28** | (-3.98) | **-0.29** | (-2.39) | **-0.16** | (-2.03) |
| 1932 | **0.28** | (5.10) | **-0.31** | (-5.32) | **-0.45** | (-4.75) | **-0.32** | (-4.84) |
| 1937 | 0.16 | (1.95) | -0.09 | (-1.72) | -0.11 | (-1.26) | -0.11 | (-1.85) |
| 1942 | **0.33** | (4.04) | -0.05 | (-1.02) | 0.03 | (0.29) | 0.07 | (1.10) |
| 1947 | **0.24** | (4.38) | **-0.16** | (-3.50) | **-0.23** | (-3.20) | **-0.17** | (-3.22) |
| 1952 | **0.13** | (2.21) | **-0.30** | (-5.90) | **-0.38** | (-4.85) | **-0.25** | (-4.60) |
| 1957 | **0.15** | (3.85) | **-0.20** | (-4.74) | **-0.30** | (-4.45) | **-0.20** | (-4.07) |
| 1962 | **0.11** | (2.59) | **-0.16** | (-3.81) | **-0.13** | (-1.97) | -0.08 | (-1.90) |
| 1967 | **0.22** | (5.34) | **-0.14** | (-3.63) | **-0.13** | (-2.09) | **-0.12** | (-2.79) |
| 1972 | **0.13** | (3.15) | **-0.16** | (-4.21) | **-0.21** | (-3.31) | **-0.20** | (-4.10) |
| 1977 | **0.14** | (3.95) | **-0.22** | (-6.57) | **-0.38** | (-6.01) | **-0.29** | (-6.66) |
| 1982 | **0.13** | (2.47) | **-0.12** | (-2.62) | 0.01 | (0.07) | 0.00 | (0.03) |
| 1987 | 0.07 | (1.84) | **-0.11** | (-3.03) | -0.13 | (-1.75) | **-0.10** | (-2.08) |
| 1992 | **0.07** | (2.04) | **-0.20** | (-5.63) | **-0.22** | (-3.28) | **-0.14** | (-3.45) |
| 1997 | **0.13** | (4.62) | **-0.20** | (-7.23) | **-0.31** | (-6.01) | **-0.23** | (-6.65) |
| 2002 | 0.02 | (1.09) | **-0.31** | (-12.21) | **-0.57** | (-13.33) | **-0.33** | (-10.18) |
| 2007 | 0.04 | (1.60) | **-0.32** | (-11.40) | **-0.55** | (-10.69) | **-0.30** | (-7.56) |
| 2012 | -0.01 | (-0.34) | **-0.31** | (-10.54) | **-0.60** | (-11.76) | **-0.36** | (-10.17) |
| 2017 | **0.10** | (4.58) | **-0.24** | (-10.94) | **-0.33** | (-9.23) | **-0.23** | (-9.40) |

**BIP_IP BIP interactions**

| | OFLD | | IFPOP | | IFGB | | OFFB | |
|---|---|---|---|---|---|---|---|---|
| 1912 | **0.19** | (2.57) | 0.00 | (-0.04) | 0.05 | (0.68) | -0.01 | (-0.13) |
| 1917 | **0.10** | (2.47) | -0.03 | (-0.61) | -0.03 | (-0.45) | **-0.10** | (-2.51) |
| 1922 | **0.16** | (3.87) | -0.04 | (-1.14) | 0.06 | (1.10) | 0.07 | (1.85) |
| 1927 | **0.07** | (2.83) | -0.05 | (-1.66) | -0.04 | (-1.16) | -0.05 | (-1.84) |
| 1932 | 0.05 | (1.40) | -0.06 | (-1.49) | -0.11 | (-1.72) | -0.08 | (-1.72) |
| 1937 | **0.13** | (3.81) | -0.04 | (-1.35) | 0.03 | (0.67) | -0.01 | (-0.33) |
| 1942 | **0.16** | (5.00) | -0.01 | (-0.22) | 0.00 | (0.07) | 0.04 | (1.65) |
| 1947 | **0.15** | (6.63) | **-0.07** | (-2.65) | -0.01 | (-0.44) | 0.02 | (1.02) |
| 1952 | 0.03 | (0.81) | **-0.08** | (-2.32) | **-0.22** | (-4.10) | **-0.15** | (-3.92) |
| 1957 | 0.04 | (1.07) | -0.08 | (-1.94) | **-0.12** | (-2.04) | **-0.13** | (-2.39) |
| 1962 | 0.04 | (1.07) | **-0.09** | (-2.40) | **-0.14** | (-2.31) | -0.07 | (-1.63) |
| 1967 | 0.00 | (-0.07) | **-0.11** | (-3.11) | **-0.17** | (-2.76) | **-0.10** | (-2.89) |
| 1972 | 0.03 | (0.66) | **-0.07** | (-2.46) | **-0.18** | (-2.78) | **-0.12** | (-2.40) |
| 1977 | -0.02 | (-0.89) | **-0.08** | (-3.42) | **-0.20** | (-4.66) | **-0.16** | (-4.19) |
| 1982 | -0.01 | (-0.09) | -0.07 | (-1.38) | -0.11 | (-0.99) | -0.07 | (-0.89) |
| 1987 | -0.01 | (-0.43) | **-0.14** | (-3.71) | **-0.18** | (-3.19) | **-0.11** | (-2.90) |
| 1992 | 0.01 | (0.46) | **-0.11** | (-4.06) | **-0.08** | (-1.71) | **-0.08** | (-2.27) |
| 1997 | 0.02 | (0.49) | **-0.13** | (-2.89) | **-0.19** | (-2.41) | **-0.11** | (-2.52) |
| 2002 | 0.02 | (1.04) | **-0.09** | (-2.88) | **-0.19** | (-3.89) | **-0.08** | (-2.31) |
| 2007 | 0.05 | (1.58) | **-0.09** | (-2.62) | **-0.14** | (-2.30) | **-0.07** | (-1.55) |
| 2012 | -0.04 | (-1.80) | **-0.17** | (-6.09) | **-0.33** | (-6.45) | **-0.21** | (-6.74) |
| 2017 | -0.04 | (-1.43) | **-0.16** | (-6.28) | **-0.26** | (-5.68) | **-0.18** | (-5.88) |
| constant | 4.62 | (38.32) | | | | | | |

**BIP joint effects**

| | Joint effects | |
|---|---|---|
| 1912 | **0.09** | 74.44 |
| 1917 | **0.17** | 265.40 |
| 1922 | **0.16** | 232.40 |
| 1927 | **0.18** | 278.15 |
| 1932 | **0.20** | 359.59 |
| 1937 | **0.13** | 157.97 |
| 1942 | **0.14** | 167.94 |
| 1947 | **0.19** | 315.82 |
| 1952 | **0.20** | 341.95 |
| 1957 | **0.18** | 283.27 |
| 1962 | **0.14** | 183.41 |
| 1967 | **0.17** | 251.27 |
| 1972 | **0.16** | 236.63 |
| 1977 | **0.21** | 402.09 |
| 1982 | **0.12** | 131.33 |
| 1987 | **0.14** | 168.87 |
| 1992 | **0.15** | 190.92 |
| 1997 | **0.20** | 360.00 |
| 2002 | **0.25** | 540.71 |
| 2007 | **0.22** | 443.41 |
| 2012 | **0.23** | 470.67 |
| 2017 | **0.24** | 498.98 |

***Random effects***

Individual pitchers

| | | Est. | SE | 95% CI | |
|---|---|---|---|---|---|
| | Variance | 13.29 | 2.87 | 8.71 | 20.28 |
| | Residual | 0.60 | 0.01 | 0.59 | 0.62 |
| *N* | 46,639 | | | | |
| Clusters | 8,942 | | | | |
| $R^2$ | 0.71 | | | | |

**Table 1. Basic model.** Outcome variable is RAPG. Data weighted on innings pitched. Predictors standardized for interpretability and for computational convenience in calculation of BIP-ERAs. MLE coefficients with z-statistics indicated parenthetically. For periods after 1912 (the reference period), coefficients for BIP variables and for BIP_IP and BIP interactions reflect sum of the relevant main effect (i.e., reference period) estimate and relative period-specific effect estimate. "Joint effects" refer to the joint effect of OFLD, IFPOP, IFGB, OFFB, and their interactions with BIP_IP in each period; the effect is reported in terms of Cohen's $f$, along with the joint-effect Wald Test $\chi^2$.

Reported in Figure 1, the estimates confirm that BIP propensities make a difference within the basic model. As one might expect, the difference tends to be larger—by substantial margins in most periods—for pitchers with moderately high BIP rates, since they necessarily are affected more by the types of balls in play that they yield than are pitchers who allow fewer balls to be hit in play to begin with. These trends have fluctuated over time. In particular, for both pitchers with high and those with low BIP rates, the impact of the difference between favorable and unfavorable mixes of BIPs was estimated to escalate sharply in the early 1990s. These results are consistent with the conclusion that BIP propensities are a consequential element of pitching proficiency.

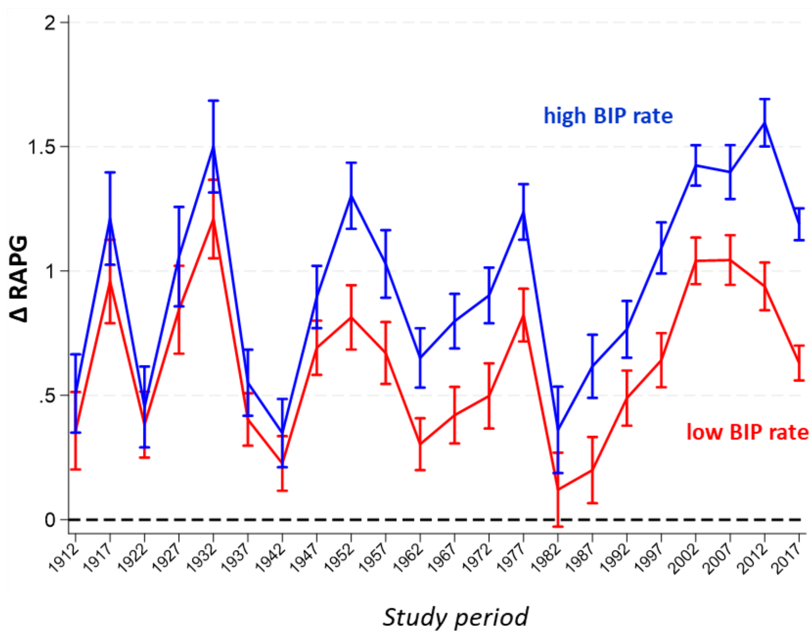



**Figure 1. Model-estimated impact of (hypothetical) BIP propensities, conditional on (hypothetical) BIP rates.** The blue and red lines connect by period the estimated run-per-game impact of yielding a pitcher-unfavorable (-1.0 SD) mix of BIP types versus a favorable mix (+1.0 SD) for pitchers with high (+1.0 SD) and low (-1.0 SD) BIP rates. Error bars are 0.95 CIs.

Obviously, though, these numbers are made up! The illustration *does* convey information about the upshot of the basic model parameters relating to BIP propensities. But because it rests on hypothetical BIP profiles, it *doesn't* tell us how much BIP propensities actually matter. Figuring that out requires applying the model to the varying BIP propensities found in real-life pitchers.

*BIP-ERA*

For that purpose, the model was used to compute sample members' BIP-ERAs. Consistent with the discussion above, each pitcher's BIP-ERA was formed by subtracting from his predicted RAPG an amount equal to the estimated RAPG of a pitcher from the same season with the same $FIP_r$, the same team-fielding value, and the same BIP per inning rate but with BIP propensities equal to the season *means* of the four BIP type variables (IFPOP, IFGB, OFLD, OFFB) and their associated BIP_IP interactions. It thus indicates how many more runs a game than average that pitcher would be expected to yield based solely on his individual influence on the types of balls hit in play, independent of both his ability to prevent balls in play and his fielders' abilities to turn them into outs. Each pitcher's BNIP-ERA, the other component of his predicted RAPG above average, was formed by subtracting from his predicted RAPG an amount equal to that of a pitcher with the same BIP propensities (that is, the same BIP type and associated BIP_IP interaction values) and team-fielding score but with a mean $FIP_r$ and mean BIP rate. It thus reflects how many more runs than average a pitcher allows by virtue of his ability to prevent balls in play.

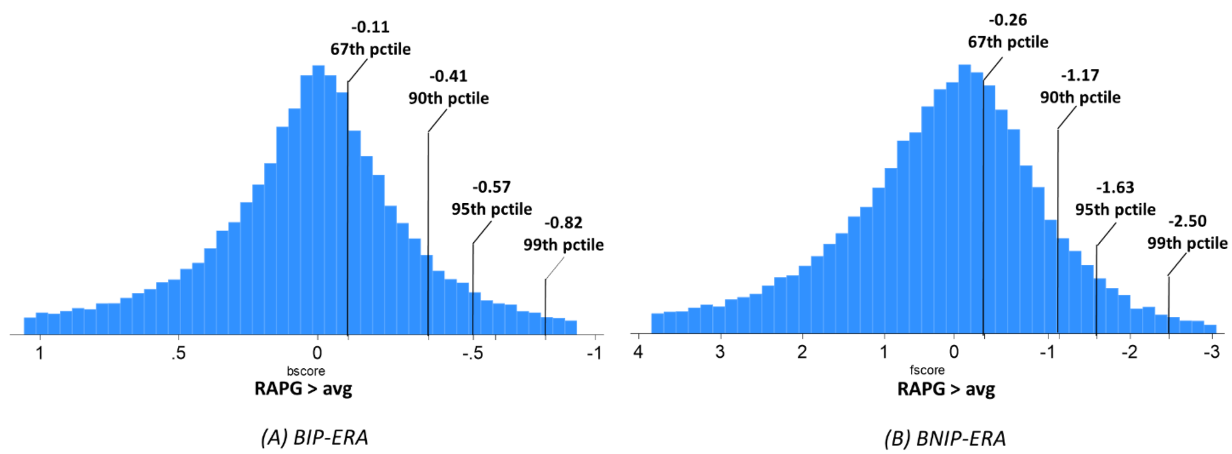



**Figure 2. Sample distributions of BIP-ERAs and BNIP-ERAs.** IQR method used to remove outliers.

BIP-ERAs so calculated were durable and consistent season-to-season. The ICC (1,1), for BIP-ERA was 0.97 (95% CI: 0.94, 0.98). This result indicates that that 97% of the variance in BIP-ERAs observed across pitchers in the sample was attributable to differences in individual pitcher characteristics—a notably high degree of reliability (Koo & Li, 2016; Cicchetti, 1994). The reliability of BNIP-ERAs was also high (ICC [1,1] = 0.93, 95% CI: 0.92, 0.94).

Looking at the sample as a whole (Figure 2), the expected difference for a pitcher with a BIP-ERA at the 67th percentile and one at the 33rd (where a higher percentile denotes a *lower* BIP-ERA) is 0.30 runs allowed per game (95% CI: 0.29, 0.31). The difference for pitchers located at the same points of the distribution for BNIP-ERA is 1.07 runs per game (95% CI: 1.05, 1.09).

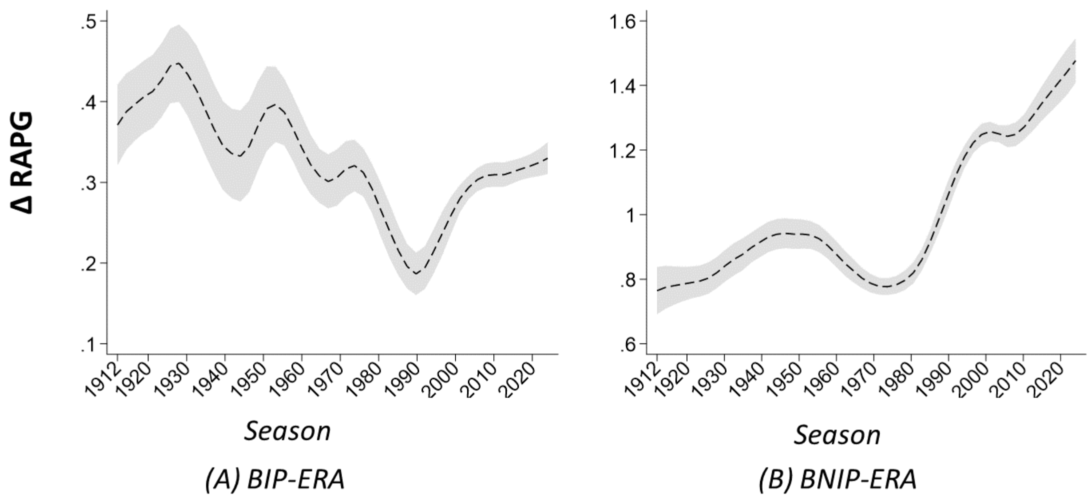



**Figure 3. RAPG effect for pitchers at 67th and 33rd percentiles of BIP-ERA and BNIP-ERA.** Panels reflect estimates of how many more runs per game a pitcher at 33rd percentile allowed than one at the 67th percentile of BIP-ERA and of BNIP-ERA, respectively, from 1912 to 2024. Estimates were made by subtracting the values of the metric in question at the indicated points in the distribution for every season. Polynomial smoothing used for presentation.

Over time, the RAPG impact of BIP propensities has fluctuated significantly (Figure 3.A). Whereas the 67th-33rd percentile interval represented 0.45 runs per game in the late 1920s, it began to decline in the 1960s and had dropped to approximately 0.20 by the early 1990s. In recent decades , it has increased dramatically, and has reached levels higher than at any point since the mid-1950s.

These shifting magnitudes, however, paint a potentially misleading picture. Period effects will be influenced not only by the genuine impact of BIP propensities but also by changes in runs scored per game: as run production increases or decreases, the consequences of any constraint on runs will likely increase or decrease as well. Variance in BIP-ERAs can be expected to change, too, as run scoring fluctuates. These influences continuously alter the scale in which BIP-ERA is measuring performance at different times (Schell, 1999, 2005). The relationship between BNIP-ERA and runs allowed per game, it is worth observing, follows the same pattern as BIP-ERA Figure 3.B)—confounding any attempt to assess the relative impact of these components of performance (which are correlated at $r = 0.33$, sample wide) from their raw scores alone. The nature of some of the scaling effects relevant to BIP-ERA and BNIP-ERA is discussed further in the SI (see SI Figure 1 and SI Figure 2, in particular).

One way to lift the veil of scaling distortions is to measure the independent contributions of BIP-ERA and BNIP-ERA to variance in RAPG over time (Figure 4). The effect of BNIP-ERA has always been larger but has ballooned in recent decades, a period in which differences in FIP are known to have swallowed up the effect that differences in team fielding have traditionally made (Fünf, 2025). BIP-ERA's contribution to variance explained has changed less dramatically in absolute terms but has grown less important in *relative* ones as BNIP-ERA has become even more decisive.

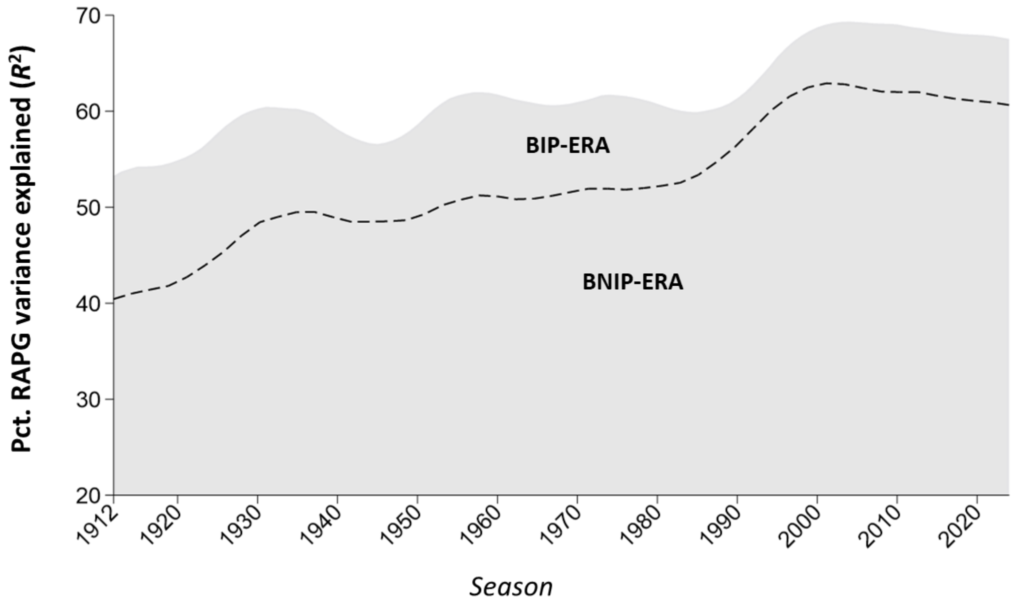



**Figure 4. Impact of BIP-ERA and non-BIP propensities on RAPG.** Derived from season-by-season regressions of RAPG on BNIP-ERA and BIP-ERA (SI). Polynomial smoothing used for presentation.

To form a valid perspective on the influence of BIP propensities at various times in AL/NL history, it is necessary to adjust BIP-ERA to reflect the run-scoring environments of the pitchers whose performances it measures. This objective can be achieved by standardizing the scale used to compute BIP-ERA across seasons, a technique Schell (1999, 2005) has applied to make hitting measures commensurable across different major league eras.

For this purpose, the basic model was re-fit using *season-standardized runs* (SI Table 1). This variant of the model estimates the impact of pitcher BIP propensities (and BNIP ones) in terms of standard deviations above or below the RAPG mean—a BIP-ERA z-score—for the season in question. The z-score scale is, by design, uniform across the history of the American and National Leagues.

For ease of interpretation, we can assign BIP-ERA z-scores a run value (Schell, 1999, 2005). It makes sense to treat each unit on this scale as equivalent to 1.5 runs, because that is the median standard deviation in season-specific RAPGs from 1912 to 2024. When multiplied by 1.5, a pitcher's BIP-ERA z-score indicates how many more "standard runs" per game than average a pitcher allowed by virtue of his unique BIP propensities. This metric will be denoted BIP-ERA$_s$.

Figure 5 looks at how standardized measures of BIP-ERA and BNIP-ERA compared to raw ones, over time. The 1920s remain the period in which differences in BIP propensities had the biggest effect. Standardization enlarges the consequence of BIP-ERAs for the early 1960s to early 1970s, a period when run scoring was depressed relative to other eras. There is also a modest deflation of the importance of BIP propensities among pitchers post-2000: after rebounding from the 1990s depression, this aspect of pitching stabilizes at a level of influence below the one that existed at earlier points in the twentieth century. This result stands in marked contrast to the effect of standardization on BNIP-ERA, differences in which continue to exceed those observed at any other point in AL/NL history (Figure 5.B). Putting the pitching performances of different eras on a common scale, then, reinforces the conclusion that BIP propensities have declined in importance relative to those associated with striking batters out, avoiding walks, and suppressing home runs.

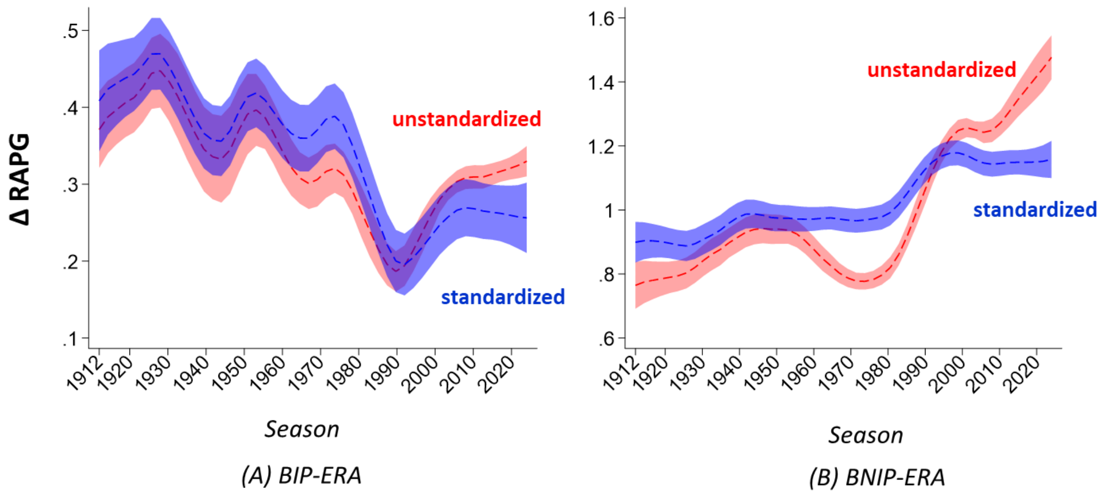



**Figure 5. RAPG impacts: unstandardized and standardized runs.** Dashed lines reflect difference in expected runs allowed per game for pitcher at 67th and 33rd percentiles in distribution of BIP-ERA and of BNIP-ERA$_s$, respectively. For both raw and standard runs. Bands reflect 0.95 CI. Polynomial smoothing used for presentation.

Individual player estimates will be presented in standardized runs in order to promote comprehension of how the shifting impact of BIP propensities as major league baseball has evolved. It should be pointed out, though, that the clarifying function of standardization does not imply that the differences in raw runs are unreal. They count no more or less just because their relationship to individual performances is distorted by non-skill related shifts in game dynamics.

It would be a large mistake, in particular, to infer that because BIP runs saved are *relatively* less important than they were in earlier times, they are today no longer of consequence. The basic model implies that they have large on-the-field impacts, as Figure 3.A attests. Indeed, for the purpose of assessing active pitchers in relation to one another, raw runs will be of greater interest than the standardized ones.

*Individual player estimates*

This section evaluates individual performances principally in terms of BIP runs saved. Such estimates will be reported in standardized runs.

Table 2 identifies the all-time best *season* BIP standard runs saved performances. The total number of season standard runs avoided by virtue of a pitcher's BIP propensities can be calculated by multiplying his BIP-ERA$_s$ by his innings pitched and dividing by -9.

The list comprises the top 58 season best marks—effectively the top 49, since the final 9 are tied. The list displays a decided skew toward the earliest decades in AL/NL history. Thirty-three of the seasons occurred before 1930. Only three were recorded after 1950 and only one after 2000.

Runs saved range from 23 to 51. Using the rule of thumb that a 10-run differential equates to 1 win (e.g., Thorn and Palmer, 2015), these pitchers can be understood to have contributed between 2 1/2 and 5 wins to their teams' season records based on their BIP propensities. Because this total is computed relative to the average pitcher, WAR would be slightly higher, in the neighborhood of 3 to 7. To determine a pitcher's *net* season WAR, of course, would require account as well for his BNIP runs allowed, which would either boost or deflate his value depending on whether it is above or below average.

| rank | Player | Season | IP | BIP-$ERA_s$ | runs saved | rank | Player | Season | IP | BIP-$ERA_s$ | runs saved |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Stan Coveleski | 1917 | 298 | -1.53 | 51 | 30 | Waite Hoyt | 1927 | 256 | -0.89 | 25 |
| 2 | Carl Mays | 1921 | 337 | -1.16 | 44 | 31 | Curt Davis | 1944 | 194 | -1.17 | 25 |
| 3 | Jesse Barnes | 1919 | 296 | -1.25 | 43 | 32 | Larry Benton | 1931 | 204 | -1.11 | 25 |
| 4 | Jim Bagby | 1917 | 321 | -1.11 | 40 | 33 | Whit Wyatt | 1941 | 288 | -0.79 | 25 |
| 5 | Lefty Gomez | 1934 | 282 | -1.14 | 39 | 34 | Joe Niekro | 1979 | 264 | -0.86 | 25 |
| 6 | Stan Coveleski | 1918 | 311 | -1.03 | 38 | 35 | Jesse Petty | 1926 | 276 | -0.82 | 25 |
| 7 | Kirby Higbe | 1941 | 298 | -1.05 | 38 | 36 | Dazzy Vance | 1928 | 280 | -0.80 | 25 |
| 8 | Bob Shawkey | 1922 | 300 | -1.01 | 37 | 37 | Ray Benge | 1934 | 227 | -0.97 | 25 |
| 9 | Watson Clark | 1935 | 207 | -1.44 | 37 | 38 | Monte Pearson | 1936 | 223 | -0.99 | 24 |
| 10 | Doug McWeeny | 1928 | 244 | -1.20 | 36 | 39 | Larry Benton | 1928 | 310 | -0.71 | 24 |
| 11 | Carl Hubbell | 1936 | 304 | -0.95 | 35 | 40 | Vern Kennedy | 1936 | 274 | -0.80 | 24 |
| 12 | Jeff Pfeffer | 1914 | 315 | -0.90 | 35 | 41 | George Pipgras | 1928 | 301 | -0.72 | 24 |
| 13 | Wilbur Cooper | 1919 | 287 | -0.98 | 33 | 42 | Burleigh Grimes | 1920 | 304 | -0.71 | 24 |
| 14 | Bob Shawkey | 1916 | 277 | -1.01 | 33 | 43 | Stan Coveleski | 1920 | 315 | -0.69 | 24 |
| 15 | Ray Caldwell | 1915 | 305 | -0.90 | 33 | 44 | Whit Wyatt | 1942 | 217 | -0.99 | 24 |
| 16 | Ed Klepfer | 1917 | 213 | -1.27 | 33 | 45 | Eddie Cicotte | 1917 | 347 | -0.62 | 24 |
| 17 | Urban Shocker | 1926 | 258 | -1.03 | 32 | 46 | Lon Warneke | 1932 | 277 | -0.78 | 24 |
| 18 | Jim Bagby | 1920 | 340 | -0.75 | 32 | 47 | Jeff Tesreau | 1915 | 306 | -0.70 | 24 |
| 19 | Carl Hubbell | 1932 | 284 | -0.89 | 32 | 48 | Waite Hoyt | 1928 | 273 | -0.78 | 24 |
| 20 | Jeff Tesreau | 1914 | 322 | -0.78 | 31 | 49 | Curt Davis | 1943 | 164 | -1.28 | 23 |
| 21 | Ken Raffensberger | 1949 | 284 | -0.88 | 31 | 50 | Rube Marquard | 1913 | 288 | -0.73 | 23 |
| 22 | Phil Niekro | 1979 | 342 | -0.73 | 31 | 51 | Bill Lee | 1936 | 259 | -0.80 | 23 |
| 23 | Hal Gregg | 1944 | 198 | -1.24 | 31 | 52 | Harry Perkowski | 1953 | 193 | -1.07 | 23 |
| 24 | Stan Coveleski | 1922 | 277 | -0.87 | 31 | 53 | Jack Warhop | 1914 | 217 | -0.95 | 23 |
| 25 | Waite Hoyt | 1921 | 282 | -0.84 | 31 | 54 | Dutch Ruether | 1919 | 243 | -0.85 | 23 |
| 26 | Freddie Fitzsimmons | 1928 | 261 | -0.91 | 31 | 55 | Carl Hubbell | 1935 | 303 | -0.68 | 23 |
| 27 | Curt Davis | 1942 | 206 | -1.13 | 30 | 56 | John Cumberland | 1971 | 185 | -1.11 | 23 |
| 28 | Derek Lowe | 2002 | 220 | -1.05 | 30 | 57 | Hal Schumacher | 1933 | 259 | -0.79 | 23 |
| 29 | Red Faber | 1921 | 331 | -0.70 | 30 | 58 | Jeff Pfeffer | 1916 | 329 | -0.62 | 23 |

**Table 2. BIP season standard runs saved.** Derived from the standard runs version of the basic model (SI Table 1), "runs saved" reflects standard runs saved per 9 innings as a result of the pitchers' BIP-$ERA_s$

The BIP-ERA$_s$s range from -1.53 to -0.62. The average is -0.93.

Many pitchers appear on the list multiple times. These include Hall of Fame member Stan Coveleski, with four top 50 performances; Hall of Famers Carl Hubbell and Waite Hoyt, along with Curt Davis three a piece; and Jim Bagby, Larry Benton, Jeff Pfeffer, Bob Shawkey, Jeff Tesreau, and Whit Wyatt, all two. The sample overall includes over 7,500 season performances among pitchers who threw enough innings to qualify for an ERA title. That 9 pitchers alone account for 23 of the top 58 performances in BIP runs saved is certainly due in large part to the disproportionate impact of BIP propensities in the early twentieth century. But it also reinforces the conclusion that superior BIP propensities are a *consistent* characteristic of certain pitchers.

Table 3 identifies the leaders in BIP runs saved for select decades. These rankings illustrate that the contribution of BIP propensities made to pitcher success has not been limited to the earliest portions of AL/NL history. Pitchers from the 1970s, including Catfish Hunter (103 standard BIP runs saved), Phil Niekro (102), and Jim Palmer (90), for example, would all have ranked near the top during the BIP high-water mark decades from the 1920s to 1940s. Led by Derek Lowe (87) and Tim Wakefield (86), pitchers from the 2000s would also have ranked near the top in earlier decades—despite fewer innings pitched. Post-2020 pitchers will not match the total BIP standard runs saved of their counterparts from those decades; but on a per-inning basis, their BIP-stifling capacities are comparably potent.

Table 4 identifies pitchers in the top 50 for total *career* BIP standard runs saved. That total is determined by summing the number of BIP standard runs saved for every season a pitcher played. Just as one can use a pitcher's BIP-ERA$_s$ to calculate how many more or fewer standard runs he allowed by virtue of his BIP propensities, one can use his BNIP-ERA$_s$ to estimate how many more or fewer standard runs per game he would have yielded in a particular season relative to one with an average FIP$_r$ and an average BIP rate. Summing *these* two runs-saved totals over all the seasons that a pitcher played generates an estimated total career "standard runs saved." The proportions attributable to his BNIP-ERA$_s$ and BIP-ERA$_s$, respectively, reflect the pitcher's abilities to stifle contact and to soften it when it occurs. The Table includes the percentage of total standard runs saved attributable to the pitchers' BIP-ERA$_s$s,

These players turned in consistently strong BIP performances over the course of their careers. The average career BIP-ERA$_s$ for list members is -0.31, which corresponds to the 85th percentile for all season BIP-ERA$_s$ recorded by pitchers over AL/NL history; a pitcher who compiles a performance metric score equivalent to the top 15% for every season pitched over an extended career evinces a high degree of proficiency in this aspect of pitching.

The contribution that suppressing quality contact made to the success of these pitchers is manifest. For 28 of them—including Hall of Famers Catfish Hunter, Phil Niekro, Burleigh Grimes, Stan Coveleski, and Lefty Gomez—BIP runs saved made up half or more of their career standard runs saved total. Indeed, eight member the list, the results suggest, relied on their BIP propensities to offset a *higher than average* total of BNIP runs allowed. For three pitchers—those with negative BIP runs-saved percentages—the facility to stifle quality contact is estimated to have reduced the margin by which their career standard runs allowed *exceeded* the number expected for an average major league pitcher during their seasons of play.

**1920s**

| | pitcher | IP | BIP-$ERA_s$ | BIP RS |
|---|---|---|---|---|
| 1 | Waite Hoyt | 2346 | -0.53 | 138 |
| 2 | Urban Shocker | 2149 | -0.43 | 104 |
| 3 | Bob Shawkey | 1613 | -0.54 | 96 |
| 4 | Jesse Petty | 1128 | -0.67 | 84 |
| 5 | Burleigh Grimes | 2798 | -0.24 | 74 |
| 6 | Herb Pennock | 2313 | -0.29 | 73 |
| 7 | Dazzy Vance | 2054 | -0.30 | 69 |
| 8 | Art Nehf | 1720 | -0.35 | 67 |
| 9 | Freddie Fitzsimmons | 1021 | -0.56 | 64 |
| 9 | Stan Coveleski | 1934 | -0.28 | 60 |

**1930s**

| | pitcher | IP | BIP-$ERA_s$ | BIP RS |
|---|---|---|---|---|
| 1 | Red Ruffing | 2439 | -0.50 | 135 |
| 2 | Carl Hubbell | 2597 | -0.41 | 119 |
| 2 | Lefty Gomez | 2235 | -0.48 | 119 |
| 4 | Hal Schumacher | 1737 | -0.40 | 77 |
| 5 | Monte Pearson | 1296 | -0.42 | 60 |
| 6 | Johnny Murphy | 679 | -0.76 | 58 |
| 7 | Van Mungo | 1715 | -0.27 | 52 |
| 8 | Watson Clark | 1174 | -0.37 | 48 |
| 9 | Roy Parmelee | 1113 | -0.36 | 45 |
| 10 | Freddie Fitzsimmons | 1938 | -0.19 | 42 |

**1940s**

| | pitcher | IP | BIP-$ERA_s$ | BIP RS |
|---|---|---|---|---|
| 1 | Curt Davis | 1061 | -1.01 | 119 |
| 2 | Kirby Higbe | 1693 | -0.55 | 104 |
| 3 | Whit Wyatt | 1015 | -0.78 | 88 |
| 4 | Les Webber | 432 | -1.00 | 48 |
| 5 | Ed Head | 465 | -0.89 | 46 |
| 6 | Warren Spahn | 990 | -0.38 | 42 |
| 7 | Hal Gregg | 785 | -0.47 | 41 |
| 8 | Rube Melton | 704 | -0.51 | 40 |
| 9 | Bill Voiselle | 1322 | -0.26 | 38 |
| 9 | Larry Jansen | 785 | -0.39 | 34 |

**1970s**

| | pitcher | IP | BIP-$ERA_s$ | BIP RS |
|---|---|---|---|---|
| 1 | Catfish Hunter | 2399 | -0.39 | 103 |
| 1 | Phil Niekro | 2881 | -0.32 | 101 |
| 3 | Jim Palmer | 2745 | -0.30 | 90 |
| 4 | Don Wilson | 1125 | -0.48 | 60 |
| 5 | Steve Renko | 1846 | -0.28 | 57 |
| 6 | Luis Tiant | 2063 | -0.21 | 47 |
| 7 | Joaquin Andujar | 636 | -0.61 | 43 |
| 8 | Carl Morton | 1619 | -0.22 | 40 |
| 9 | Ross Grimsley | 1863 | -0.19 | 39 |
| 9 | Ken Forsch | 1271 | -0.27 | 38 |

**2000s**

| | pitcher | IP | BIP-$ERA_s$ | BIP RS |
|---|---|---|---|---|
| 1 | Derek Lowe | 1834 | -0.43 | 87 |
| 2 | Tim Wakefield | 1747 | -0.44 | 86 |
| 3 | Tim Hudson | 1923 | -0.32 | 68 |
| 4 | Mark Buehrle | 2061 | -0.24 | 54 |
| 5 | Barry Zito | 1999 | -0.23 | 51 |
| 5 | Roy Halladay | 1883 | -0.24 | 50 |
| 7 | Jose Contreras | 1084 | -0.28 | 34 |
| 8 | Freddy Garcia | 1571 | -0.18 | 32 |
| 8 | Carlos Zambrano | 1551 | -0.19 | 32 |
| 8 | Greg Maddux | 1940 | -0.15 | 32 |

**2020s**

| | pitcher | IP | BIP-$ERA_s$ | BIP RS |
|---|---|---|---|---|
| 1 | JP Sears | 423 | -0.38 | 18 |
| 1 | Corbin Burnes | 757 | -0.21 | 18 |
| 3 | Zack Wheeler | 758 | -0.18 | 16 |
| 4 | George Kirby | 512 | -0.25 | 14 |
| 4 | Kutter Crawford | 392 | -0.32 | 14 |
| 6 | Tyler Rogers | 301 | -0.38 | 13 |
| 6 | Joe Ryan | 470 | -0.24 | 13 |
| 8 | Jameson Taillon | 641 | -0.17 | 12 |
| 8 | Framber Valdez | 710 | -0.15 | 12 |
| 9 | Matt Waldron | 188 | -0.53 | 11 |
| 9 | Adrian Houser | 425 | -0.23 | 11 |
| 9 | Tarik Skubal | 539 | -0.18 | 11 |

**Table 3. Select Decade BIP runs saved leaders.** Derived from basic model (SI Table 1). "BIP RS" formed by summing season BIP runs saved for indicated decade. "BIP %" is percent of runs saved attributable to BIP-$ERA_s$. The Supplemental Information (SI Table 6) reports additional data on decade leaders (measured in raw runs saved).

| rank | Player | BIP-ERA$_s$ | BIP RS | BIP RS % |
|---|---|---|---|---|
| 1 | Carl Hubbell | -0.37 | 148 | 28% |
| 2 | Phil Niekro | -0.24 | 147 | 53% |
| 3 | Red Ruffing | -0.30 | 143 | 43% |
| 4 | Freddie Fitzsimmons | -0.38 | 135 | 60% |
| 5 | Catfish Hunter | -0.35 | 134 | 81% |
| 6 | Warren Spahn | -0.22 | 131 | 26% |
| 7 | Stan Coveleski | -0.38 | 130 | 50% |
| 8 | Bob Shawkey | -0.39 | 129 | 64% |
| 9 | Lefty Gomez | -0.42 | 116 | 50% |
| 9 | Jim Palmer | -0.26 | 116 | 46% |
| 11 | Waite Hoyt | -0.27 | 111 | 34% |
| 11 | Tim Wakefield | -0.30 | 108 | < 0% |
| 13 | Curt Davis | -0.42 | 107 | 41% |
| 14 | Urban Shocker | -0.34 | 102 | 39% |
| 15 | Tim Hudson | -0.29 | 99 | 49% |
| 16 | Kirby Higbe | -0.44 | 96 | > 100% |
| 17 | Whit Wyatt | -0.48 | 94 | 49% |
| 18 | Hal Schumacher | -0.33 | 91 | 76% |
| 19 | Derek Lowe | -0.30 | 88 | 62% |
| 19 | Burleigh Grimes | -0.19 | 88 | 56% |
| 19 | Ray Fisher | -0.52 | 85 | 91% |
| 22 | Jeff Tesreau | -0.44 | 83 | 81% |
| 22 | Ray Caldwell | -0.38 | 83 | > 100% |
| 24 | Ken Raffensberger | -0.34 | 81 | 36% |
| 25 | Art Nehf | -0.27 | 80 | 45% |
| 26 | Dazzy Vance | -0.24 | 79 | 16% |
| 26 | Jesse Barnes | -0.28 | 79 | 38% |
| 28 | Jesse Petty | -0.55 | 74 | 56% |
| 29 | Jeff Pfeffer | -0.27 | 73 | 59% |
| 30 | Rube Marquard | -0.23 | 71 | 31% |
| 31 | Phil Douglas | -0.37 | 70 | 53% |
| 32 | Slim Sallee | -0.29 | 68 | 48% |
| 32 | George Pipgras | -0.41 | 67 | 65% |
| 32 | Luis Tiant | -0.17 | 67 | 32% |
| 32 | Al Demaree | -0.41 | 64 | 90% |
| 32 | Steve Renko | -0.23 | 64 | > 100% |
| 32 | Dutch Ruether | -0.27 | 64 | 90% |
| 32 | Mark Buehrle | -0.17 | 64 | < 0% |
| 39 | Jim Bagby | -0.31 | 63 | > 100% |
| 40 | Fred Toney | -0.26 | 61 | 68% |
| 41 | Lew Burdette | -0.17 | 59 | 45% |
| 42 | Jack Coombs | -0.53 | 57 | < 0% |
| 43 | Johnny Murphy | -0.49 | 56 | 98% |
| 43 | Juan Marichal | -0.14 | 56 | 16% |
| 43 | Ken Forsch | -0.23 | 56 | 38% |
| 46 | Monte Pearson | -0.35 | 55 | > 100% |
| 46 | Roy Halladay | -0.18 | 55 | 18% |
| 46 | Watson Clark | -0.28 | 55 | 29% |
| 49 | Don Wilson | -0.28 | 54 | 34% |
| 50 | Don Sutton | -0.09 | 52 | 12% |

**Table 4. BIP career standard runs saved.** Derived from standardized-runs basic model (SI Table 1). Career BIP-ERA$_s$ reflects IP-weighted average over seasons played; BIP RS refers to career BIP runs saved, determined by sum of season BIP standard runs saved over course of career. BIP RS % refers to percentage of career runs saved due to BIP-ERA$_s$, calculated in relation to sum of BIP and BNIP standard season runs saved over course of career; "> 100%" indicates that BIP runs saved exceeded the number of BNIP runs saved, "< 0%" that the pitcher's total runs saved were negative on net, and a negative % that BIP runs allowed reduced positive career runs saved on net.

| rank | Player | BIP-ERA$_s$ | BIP RS | Total RS | BIP RS % |
|---|---|---|---|---|---|
| 1 | Randy Johnson | 0.02 | -7 | 880 | -1% |
| 2 | Roger Clemens | -0.02 | 13 | 872 | 2% |
| 3 | Walter Johnson | -0.04 | 23 | 743 | 3% |
| 4 | Nolan Ryan | 0.05 | -29 | 721 | -4% |
| 5 | Pedro Martinez | -0.03 | 9 | 684 | 1% |
| 6 | Greg Maddux | -0.09 | 49 | 586 | 8% |
| 7 | Pete Alexander | -0.07 | 38 | 552 | 7% |
| 8 | Curt Schilling | 0.03 | -11 | 545 | -2% |
| 9 | John Smoltz | -0.01 | 5 | 529 | 1% |
| 10 | Clayton Kershaw | -0.11 | 33 | 529 | 6% |
| 11 | Carl Hubbell | -0.37 | 148 | 525 | 28% |
| 12 | Warren Spahn | -0.22 | 131 | 512 | 26% |
| 13 | Dazzy Vance | -0.24 | 79 | 508 | 16% |
| 14 | Lefty Grove | -0.04 | 16 | 480 | 3% |
| 15 | Don Sutton | -0.09 | 52 | 450 | 12% |
| 16 | Steve Carlton | 0.05 | -31 | 448 | -7% |
| 17 | Tom Seaver | 0.07 | -36 | 435 | -8% |
| 18 | Gaylord Perry | 0.06 | -33 | 421 | -8% |
| 19 | Max Scherzer | 0.01 | -3 | 415 | -1% |
| 20 | Sandy Koufax | -0.08 | 21 | 396 | 5% |
| 21 | David Cone | 0.02 | -7 | 360 | -2% |
| 22 | Bert Blyleven | 0.14 | -78 | 353 | -22% |
| 23 | Mike Mussina | 0.04 | -17 | 351 | -5% |
| 24 | Justin Verlander | -0.01 | 5 | 350 | 1% |
| 25 | Kevin Brown | -0.06 | 23 | 345 | 7% |
| 26 | Dennis Eckersley | -0.11 | 39 | 342 | 11% |
| 27 | Juan Marichal | -0.14 | 56 | 342 | 16% |
| 28 | Red Ruffing | -0.30 | 143 | 328 | 43% |
| 29 | Waite Hoyt | -0.27 | 111 | 327 | 34% |
| 30 | Dizzy Dean | 0.06 | -14 | 324 | -4% |
| 31 | Bob Gibson | 0.09 | -39 | 322 | -12% |
| 32 | Johan Santana | -0.10 | 22 | 320 | 7% |
| 33 | Dutch Leonard | -0.01 | 4 | 320 | 1% |
| 34 | Paul Derringer | 0.20 | -81 | 320 | -25% |
| 35 | Chris Sale | 0.01 | -1 | 317 | 0% |
| 36 | Dwight Gooden | 0.01 | -4 | 313 | -1% |
| 37 | Steve Rogers | -0.11 | 36 | 310 | 12% |
| 38 | Roy Halladay | -0.18 | 55 | 306 | 18% |
| 39 | Mariano Rivera | -0.28 | 39 | 296 | 13% |
| 40 | Fergie Jenkins | -0.05 | 23 | 291 | 8% |
| 41 | Claude Passeau | -0.04 | 12 | 285 | 4% |
| 42 | Bret Saberhagen | 0.01 | -4 | 283 | -1% |
| 43 | J. R. Richard | -0.10 | 18 | 280 | 6% |
| 44 | Don Drysdale | 0.05 | -20 | 277 | -7% |
| 45 | Phil Niekro | -0.24 | 147 | 276 | 53% |
| 46 | Jacob deGrom | -0.03 | 4 | 275 | 1% |
| 47 | Hoyt Wilhelm | -0.17 | 43 | 274 | 16% |
| 48 | Herb Pennock | -0.06 | 24 | 268 | 9% |
| 49 | Billy Wagner | 0.02 | -2 | 266 | -1% |
| 50 | Robin Roberts | 0.08 | -41 | 266 | -15% |

**Table 5. Career standard runs saved, BIP contribution.** Derived from basic model standardized (SI Table 1). Career BIP-ERA$_s$ reflects IP-weighted average over seasons played; career BIP RS refers to career BIP runs saved, determined by sum of season BIP standard runs saved over course of career. BIP RS % refers to percentage of runs saved due to BIP-ERA$_s$, calculated in relation to sum of BIP and BNIP standard season runs saved over course of career; negative % indicates that BIP runs allowed reduced career runs saved on net..

The lists feature a diverse collection of hurlers. Pitchers from the pre-war period, including career leaders Hubbell and Coveleski, still dominate, occupying roughly half the slots. But they are joined by pitchers from a variety of other eras, including the late 1940s (Ken Raffensburger and Kirby Higbe), 1950s (Warren Spahn and teammate Lew Burdette), the 1960s and 1970s (Catfish Hunter, Juan Marichal Phil Niekro, Jim Palmer, Steve Renko, and Luis Tiant). Five pitchers who played into the twenty-first century are on the list.

The pitchers also display an interesting variety of styles. Knuckleballers Wakefield and Niekro, both top 10 finishers, stand out. So do "grandfathered" spitballers Coveleski and Grimes. Hubbell was known for his screwball. Spahn also relied on a screwball (or possibly a circle-change) by the mid-1950s. As his career progressed, Hunter relied less on his fastball than on an array of pitches that hitters found it difficult to adjust to (James & Neyer, 2008). BIP propensities seem, not surprisingly, to be a weapon included principally in the arsenal of pitchers lacking overpowering speed.

Jim Palmer leaned heavily on a better-than-average fastball for his first decade in the big leagues but thereafter employed a deceptive mix of speeds. Commentators have wondered whether Palmer's exceptionally low career ERA of 2.86 should be credited to Brooks Robinson, Paul Blair, Mark Belanger, Bobby Grich, and the other great fielders who played for the Orioles from the late 1960s to mid-1970s (Smith, 2008, 2024). With a career -0.25 BIP-ERA$_s$ (equivalent to the 82nd percentile season average) and with 46% of his career standard runs saved due to his BIP ones, Palmer apparels to have honestly earned his own low earned run average.

The twenty-first century pitchers, in particular, are distinguished by their soft deliveries. Neither Hudson (61% career runs saved attributable to BIP-ERA$_s$), Lowe (79%), nor Buehrle (>100%) possessed overwhelming speed (James & Neyer, 2008). Wakefield (< 0%) threw a fastball slower than most pitchers' changeups. Halladay (25%) became an effective pitcher only after he deemphasized his four-seam fastball in favor of slower cutter.

The number of career standard runs saved is also worthy of note. According to basic model estimates, Hubbell saved 148 career runs by virtue of his BIP propensities, Niekro 147, and Ruffing 143. These totals equate to somewhere in the vicinity of 21 WAR. Fitzsimmons (135), Hunter (134), Spahn (131), and Coveleski (130), are not so far behind, racking up runs-save totals equivalent to 15-18 WAR These numbers reflect respectable quantums of pitcher value.

Nevertheless, the totals become modest—well under 10 WAR—as one proceeds to the bottom of the list. The decade-leader totals, too, are very modest for pitchers in the lower positions.

It bears emphasis, too, that the results featured in Tables 2-4 reflect the very best of the best-ever BIP season, decade and career performances. Pitchers not on these lists who ranked higher than average in stifling quality contact—even considerably higher than average—realized much smaller numbers of runs saved. In addition, throughout AL/NL history, and over particular decades in particular, most of the best pitchers have succeeded without the benefit of superior BIP propensities.

This point can be illustrated by examining the contribution BIP propensities made to the performances of pitchers with the highest career runs saved in total. Table 5 identifies the top 50, based on summing the season totals for both BIP and BNIP standard runs saved over the course of their playing time. The contribution of BIP standard runs saved ranges from 53% (Niekro) to -25% (Derringer); because all members of the list necessarily have positive totals for runs saved, negative BIP-contribution percentages denote BIP runs allowed in excess of the ones that would have been yielded by an average pitcher with the same fielding support. The mean contribution of BIP runs saved is 5%. Career BIP-ERA$_s$s range from -0.37 (Hubbell) to 0.20 (again Derringer), with a mean of -0.05. The list, then, is effectively BIP neutral on net. The majority of career leaders did not rely on BIP propensities to a meaningful degree. At one and the same point in major league history, there were leading pitchers who relied heaving on their BIP propensities for success and others who relied on them not at all, or who even overcame their inferior BIP propensities by virtue of their superior BNIP ones: Red Ruffing (44%) and Lefty Grove (3%) in the 1930s, for example; or Warren Spahn (26%) and Robin Roberts (-15%) in the 1950s; Phil Niekro (53%) and Bert Blyleven (-22%) in the 1970s; and Roy Halladay (18%) and Pedro Martinez (1%) in the late 1990s and 2000s.

The story across the board, then, is pretty much the same. BIP propensities matter. For some pitchers—over a variety of eras in baseball—they have mattered a lot. But on the whole, this aspect of pitching proficiency has played only a minor role in relation to the traditional elements of fielding-independent pitching (strikeouts, walks, and home runs allowed).

*Knuckleballers*

Because they have been cited in the McCracken debate as generating poor quality contact, knuckleballers were separately examined. Table 6 reports BIP standard runs saved and other pertinent information for recognized knuckleball pitchers (Neyer, unknown), listed in order of BIP runs saved.

Niekro and Wakefield, the basic model suggests, fully merit recognition as pitchers able to blunt the impact of BIPs. As indicated previously, they are both among the top in career BIP standard runs saved.

| rank | Player | IP | BIP-ERA$_s$ | BIP RS | BIP % |
|---|---|---|---|---|---|
| 1 | Phil Niekro | 5404 | -0.24 | 147 | 53% |
| 2 | Tim Wakefield | 3226 | -0.30 | 108 | < 0% |
| 3 | Charlie Hough | 3801 | -0.10 | 43 | < 0% |
| 4 | Hoyt Wilhelm | 2254 | -0.17 | 43 | 16% |
| 5 | Bob Purkey | 2115 | -0.17 | 40 | > 100% |
| 6 | R. A. Dickey | 2074 | -0.17 | 38 | < 0% |
| 7 | Eddie Rommel | 2556 | -0.09 | 27 | 75% |
| 8 | Eddie Fisher | 1539 | -0.15 | 25 | 83% |
| 9 | Steve Sparks | 1320 | -0.14 | 21 | < 0% |
| 10 | Tom Candiotti | 2725 | -0.05 | 14 | 15% |
| 11 | Steven Wright | 348 | -0.33 | 13 | < 0% |
| 12 | Matt Waldron | 188 | -0.53 | 11 | < 0% |
| 13 | Dutch Leonard | 3218 | -0.01 | 4 | 1% |
| 14 | Al Papai | 240 | -0.06 | 2 | < 0% |
| 15 | Jared Fernandez | 109 | -0.12 | 1 | < 0% |
| 16 | Wally Burnette | 263 | -0.04 | 1 | 16% |
| 17 | Eddie Gamboa | 13 | -0.33 | 0 | 51% |
| 18 | Eddie Cicotte | 2368 | 0.00 | -1 | 0% |
| 19 | Charlie Haeger | 83 | 0.11 | -1 | 7% |
| 20 | Wilbur Wood | 2684 | 0.05 | -13 | - 29% |
| 21 | Johnny Niggeling | 1251 | 0.11 | -16 | > 100% |
| 22 | Mickey Haefner | 1467 | 0.12 | -19 | < 0% |
| 23 | Roger Wolff | 1025 | 0.18 | -21 | < 0% |

**Table 6. Knuckleball pitchers.** Derived from Basic Model, standard runs (SI Table 1). BIP-ERA$_s$ is career weighted average of season BIP-ERA$_s$ (Schell, 1999, 2005); BIP RS is career runs saved by virtue of BIP-ERA$_s$; and "BIP%" is percentage of career runs saved due to BIP runs saved. "> 100%" indicates that BIP runs saved exceeded the number of BNIP runs saved, "< 0%" that the pitcher's total runs saved were negative on net negative, and a negative % that BIP runs allowed reduced positive career runs saved on net.

There were a number of other knuckleballers whose ability to mute the impact of balls hit in play can be viewed as instrumental to their success. These include Eddie Rommel, likely the first prominent AL/NL pitcher to rely on the knuckleball, who enjoyed short-term success as a pitcher for the Athletics in the 1920s: according to the basic model, Rommel owes 75% of his career standard runs saved to his BIP propensities. Nineteen sixties pitchers Bob Purkey, who enjoyed short-lived success as a starter, and Eddie Fisher, who lasted slightly longer as a relief specialist, also make effective use of BIP propensities (> 100% and 83% contributions to standard runs saved, respectively). Charlie Hough also relied on BIP runs saved to compensate for greater than average number of BNIP runs allowed.

Hoyt Wilhelm ranks fourth among knuckleballers for BIP runs saved. His -0.17 BIP-ERA$_s$ is equivalent to a 74th percentile season average—very respectable but well behind the marks of Wakefield (-0.30; 87th percentile) and Niekro (-0.24; 80th). BIP runs saved make up only 16% of his career total. While markedly superior to those of an average pitcher, Wilhelm's BIP propensities might still be viewed as making a more modest contribution to his success than might have been suspected.

Indeed, at least some successful knuckleballers have enjoyed successful careers without relying on superior BIP propensities. Dutch Leonard, first in knuckleball pitcher total career standard runs saved, owed 1% to his BIP propensities. Eddie Cicotte, fourth in knuckleball pitcher total career standard runs saved, gained nothing from his BIP propensities, and 1970s knuckleball star Wilbur Wood gave back a portion of his total career standard runs saved by virtue of his below average BIP propensities.

Employing a logistic regression analysis, it was found that relying principally on the knuckleball made it 7 percentage points more likely that a pitcher would record a season BIP-ERA$_s$ at or above the 90th percentile ( (0.95 CI: 2%, 12%); a knuckleballer was predicted to be 14 percentage points more likely to record a season BIP-ERA$_s$ at or above the 75th percentile (0.95 CI: 7%, 21%).[4] Given the relatively small number of knuckleball pitchers over AL/NL history, these numbers are likely being driven almost entirely by Niekro and Wakefield. Nevertheless, the basic model supports the conclusion that pitchers in this class are more proficient than average in stifling the impact of balls in play.

*BABIP*

Participants in the debate over the McCracken conjecture have focused on BABIP. The obvious reason for doing so is that any influence of pitcher BIP propensities on runs allowed is necessarily mediated by the effect of such propensities on base hits. Still, the impact of BABIP on runs scored has never been precisely measured; indeed, it has undoubtedly varied substantially over time as BABIP has fluctuated, both in absolute terms and in relation to other influences on runs (Figure 6). Accordingly, even if one succeeded in disentangling the effect of pitching from fielding on BABIP, one would not be able to quantify the effect of BIP propensities on runs allowed—the true outcome of interest. For that reason, this study has focused on estimating the impact of BIP propensities on RAPG *directly*, rather than indirectly through the opaque lens of BABIP.

[4] These same analysis was performed using unstandardized runs. It was found that being a knuckleballer was associated with a 5 percentage point greater likelihood (0.95 CI: 1%, 10%) of recording a BIP-ERA at or above 90th percentile and an 13 percentage greater likelihood (0.95 CI: 7%, 20%) of recording a BIP-ERA at or above the 75th percentile.

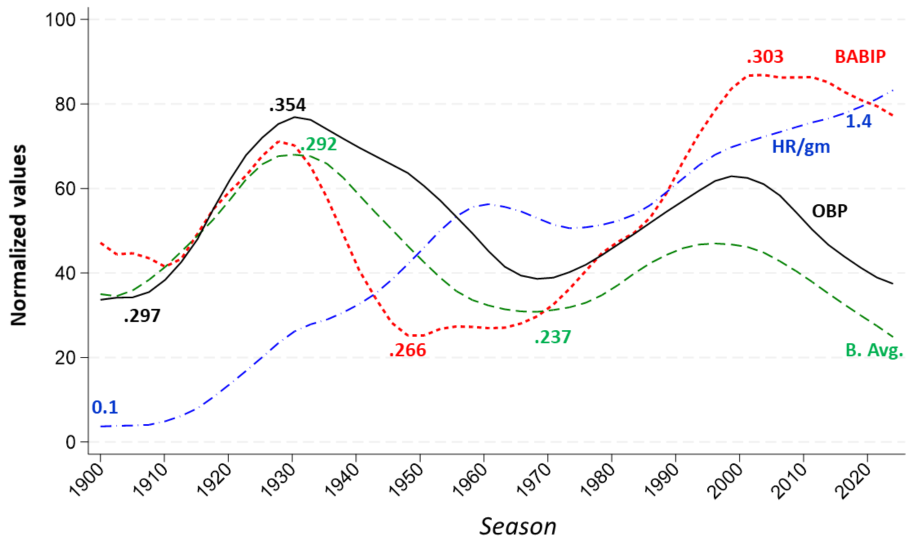



**Figure 6. Trends in elements of run production.** Values for annual MLB averages (AL/NL only), normalized on 0-100 scale with high and low values noted for reference. Data sources: Lahman database; Baseball Reference. Polynomial smoothing for presentation.

Nevertheless, analyses of the effect of BIP propensities on BABIP were also conducted. Envisioned as a validation of BIP-ERA, a separate regression model was constructed to test the power BIP-ERA itself in predicting BABIP, again controlling for team fielding. The results suggested that, over the course of AL/NL history as a whole each 0.10 change in BIP-ERA is associated with a 4.5-point change in BABIP. Accordingly, a pitcher with a BIP-ERA of -0.45—equivalent to the level Carl Hubbell achieve between 1930 and 1939, or what Derek Lowe and Tim Wakefield did from 2000 and 2009 (SI Table 6)—could be expected to reduce opposing hitters' BABIP by about 20 points.

*BIP-ERA and pitcher WAR*

The principal objective of evaluating discrete elements of pitching performance is to estimate the impact that individual pitchers' have on their teams' capacity to win games. WAR aggregates these estimations. As indicated, the leading WAR frameworks—Baseball Reference's and FanGraph's—reflect opposing views on the responsibility of BIP propensities for pitcher success. Neither, however, attempts to measure the impact of those propensities. By virtue of this omission, it stands to reason that both are likely to misestimate pitcher value to some degree.

The basic model confirms this surmise (Figure 7). With BIP-ERA added, FanGraphs WAR would explain 34% of RAPG variance over the course of AL/NL history; without it, it explains only 19%. Adding BIP-ERA or its equivalent would thus increase the RAPG explanatory power of FanGraphs pitcher WAR by nearly 80%. Not surprisingly, the impact is substantially smaller for Baseball Reference's pitcher WAR assessments, which, as discussed, are intended to credit pitchers for their BIP outcomes (Smith 2024). Still, the effect is not trivial: adding BIP-ERA to Baseball Reference's WAR scores improves RAPG explanatory power by about a third—from 33% to 44% variance explained (SI Table 7). It seems reasonable to conclude that both frameworks would be strengthened by the incorporation of BIP-ERA or an equivalent measure of BIP propensities. (For comparison, the addition of BIP-ERA to BNIP-ERA, the component of the basic model that predicts pitcher performance independent of BIP propensities, increases RAPG variance explained from 54% to 63%, approximately a 17% boost in power.)

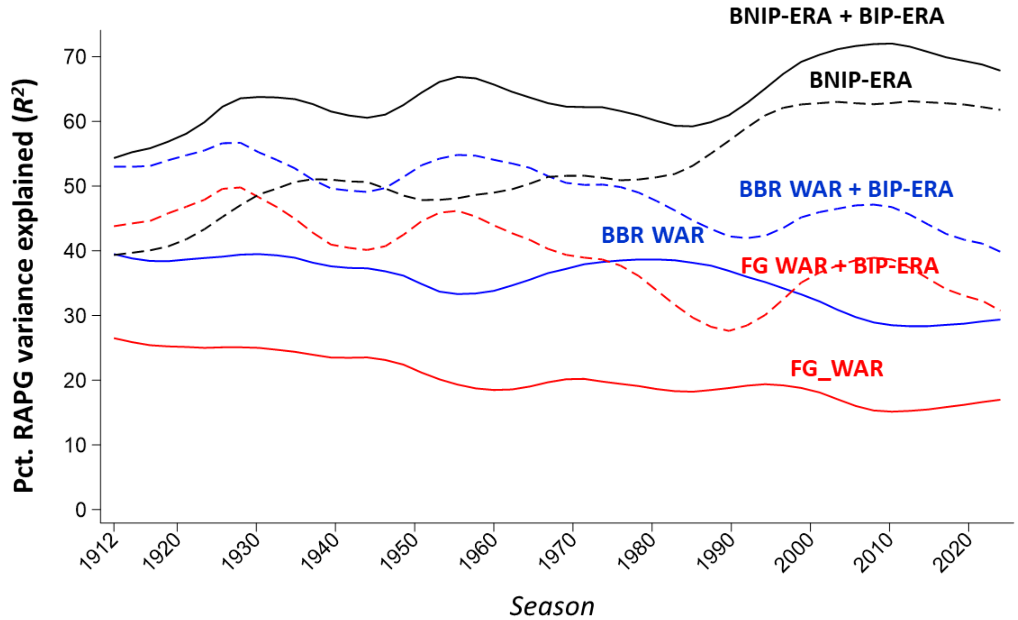



**Figure 7. Explanatory impact of neglecting BIP-ERA for calculation of pitcher WAR (AL/NL only).** Derived from season-by-season regression models of runs allowed per game on the indicated measures (SI). Polynomial smoothing used for presentation.

## Conclusions

This paper has presented the results of an empirical study of the impact of pitchers on balls in play. It is now possible to take stock of what that inquiry has revealed.

*The strange career of BIP propensities*

The most basic finding is that pitcher BIP propensities are real. The ability (or lack thereof) to stifle quality contact is a stable and highly reliable feature of individual pitchers. Measured over the course of AL/NL history, individual pitcher characteristics explain over 95% of the variance in the recurring mixes of ball-in-play outcomes—grounders, infield pop ups, outfield fly balls, and line drives—that can be demonstrated to affect run production.

The study also suggests that this characteristic has had an important impact. Pitchers who displayed the highest proficiency in stifling quality contact can be expected to avoid between 0.30 to 0.50 runs per game over the course of their careers. In individual seasons (particularly before the advent of early removal of starting pitchers), certain hurlers have achieved BIP-ERAs of -1.0 or better, saving their teams 30, 40, and even 50 runs by virtue of their BIP propensities. The failure of existing WAR frameworks to include elements for measuring the impact of BIP propensities has demonstrably impaired their assessments of pitcher value.

Nevertheless, the overall contribution of BIP propensities is modest. Pitchers' BNIP proficiencies—their ability to avoid balls in play while simultaneously limiting home runs and walks—have always mattered more. Over recent decades, moreover the contribution of BIP propensities have shrunk dramatically relative to BNIP ones. Much like differences in the abilities of fielders to turn batted balls into outs (Fünf, 2025*a*), differences in the facility of pitchers to induce easily fielded batted balls have been robbed of a substantial portion of their significance by the growing hegemony of the strikeout/home run showdown.

It would be a mistake to conclude, though, that the data reveal a story of BIP propensities that is perfectly straightforward and linear. As surmised, many early century pitches posted exceptional BIP-ERAs. But even in the middle of the twentieth century, there were just as many more pitchers who succeeded despite their BIP propensities than pitchers who succeeded because of them (Table 5). Despite the evidence it amasses on the impact of pitchers' ability to dampen quality contact, then, this study did not convincingly reveal it to be the "missing proficiency" that distinguished superior from inferior pitchers in the era before strikeout rates became the predominant factor.

What is more, the data presented here complicate rather than clarify the picture of what differentiates the best and worst pitchers today. BIP propensities might have grown less consequential over time (Figure 4); yet a nontrivial number of soft-throwers have still managed to parley a skill for inducing harmless infield popups and lazy outfield flies into successful twenty-first century baseball careers. Indeed, even if not in relative terms, the significance of BIP propensities has actually increased in *absolute* ones in recent decades (Figure 3.A.; SI). Like John Cleese, run-stifling BIP propensities are not dead yet.

*The McCracken conjecture*

The conjecture that furnished the principal impetus for this study was McCracken's: that the outcomes of balls hit in play was entirely a matter of fielding and chance rather than any material "difference among major-league pitchers in their ability to prevent hits on balls hit in the field of play" (McCracken 2001). Again, his thesis continues to divide baseball analysts.

So who should be declared the victor in the BIP debate? McCracken and his supporters or his many detractors?

The question is ill-formed. Provocative, impassioned disagreements of this sort are not contests to determine who possesses superior insight. Rather, they are the signatures of *shared* deficits in knowledge, which systematic empirical inquiry alone can dispel.

Leo Durocher might say that in baseball it's not how you play the game but only whether you win or lose that counts. But in the empirical study of baseball, who is "right" and who "wrong" is irrelevant. The only thing that matters is *measurement*.

## Bibliography


Baseball Reference, Pitching War Calculations and Details (undated). Available at https://www.baseball-reference.com/about/war_explained_pitch.shtml (undated

Cicchetti, D. & Sparrow, S.A. Developing Criteria for Establishing Interrater Reliability of Specific Items: Applications to Assessment of Adaptive Behavior. *American journal of mental deficiency* **86**, 127-137 (1981).

Clemens, B. The case against FIP. (2022). Available at https://blogs.fangraphs.com/the-case-against-a-case-against-fip/.

Cohen, J., Cohen, P., West, S.G. & Aiken, L.S. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (L. Erlbaum Associates, Mahwah, N.J., 2003).

Devore, J.L. *Probability and Statistics for Engineering and the Sciences* (Thomson/Brooks/Cole, 2008).

Efron, B. & Morris, C. Stein's paradox in statistics. *Scientific American* **236**, 119-127 (1977).

Fünf, X. Fielding before and after baseball's great transformation. Working paper. (2025*a*). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5067539.

Fünf, X. Flipping out over FIP. Bbcardstats.com (Nov. 27, 2024). Available at https://bbcard-stats.com/flipping-out-over-fip/.

Fünf, X. What's wrong with FIP and how to fix it. Bbcardstats.com (Feb. 19, 2025*b*). Available at https://bbcardstats.com/whats-wrong-with-fip-and-how-to-fix-it/

Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, 2007).

James, B. & Neyer, R. *The Neyer/James Guide to Pitchers: An Historical Compendium of Pitching, Pitchers, and Pitches* (Touchstone, 2008).

King, G., Tomz, M. & Wittenberg, J. Making the most of statistical analyses: Improving interpretation and presentation. *American journal of political science*, 347-361 (2000).

Koo, T.K. & Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **15**, 155-163 (2016).

McCracken, V. Pitching and Defense: How much Control Do Hurlers Have. *Baseball Prospectus* (2001). Available at https://web.archive.org/web/20060613054445/http://baseballprospectus.com/article.php?articleid=878.

Neyer, R. The (Mostly) Complete List of Knuckleball Pitchers. (Oddball-mall.com) (date unknown). Available at https://web.archive.org/web/20090329041844/http://www.oddball-mall.com/knuckleball/list.htm..

Pavitt, C. *Sabermetric Research Literature Review* (unpublished, 2025), available at https://charlie-pavitt.home.blog/.

Schell, M.J. *Baseball's All-Time Best Hitters: How Statistics Can Level the Playing Field* (Princeton University Press, 1999).

Schell, M.J. *Baseball's All-time Best Sluggers: Adjusted Batting Performance from Strikeouts to Home Runs* (Princeton University Press, 2005).

Smith, S. Measuring defense for players back to 1956 (Part 2). *Hardball Times* (2008). Available at https://tht.fangraphs.com/measuring-defense-for-players-back-to-1956-part-2/.

Smith, S. *War in Pieces* (Indpnt., North Haven, Conn., 2024).

Thorn J, Palmer P and Law K (2015) The Hidden Game of Baseball. University of Chicago Press.

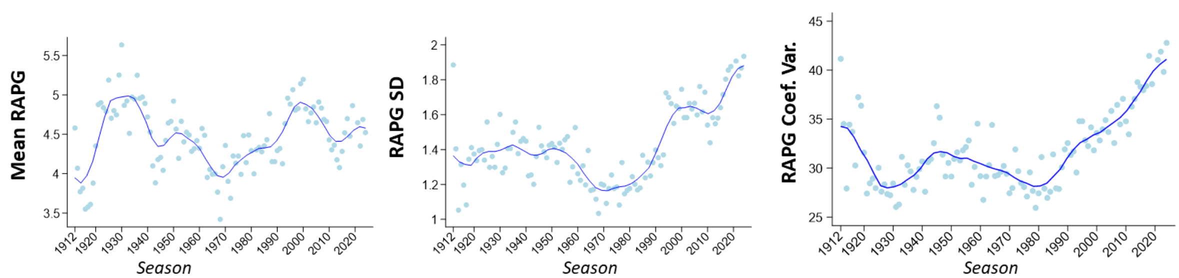
Weinberg, N. Calculating Pitcher WAR, A Complete Example. (FanGraphs, 2017). Available at https://library.fangraphs.com/calculating-pitcher-war-a-complete-example/.

## Supplemental Information

## RAPG variation

One of the most pronounced and surprising features of the impact of BIP propensities is their pronounced upward trajectory in the last 30 years. This effect is muted by standardization of runs per game, but only partially (Figure 5.A).

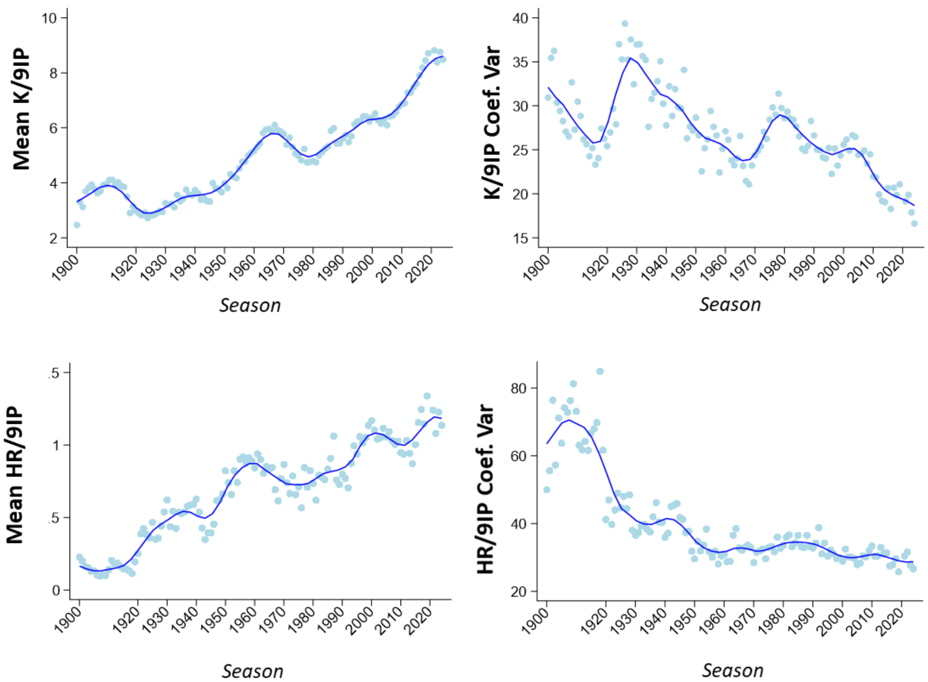
The greater differentiation between best and worst performers is not a modeling artifact; rather it is a demonstrable consequence of a real-world change in both the magnitude and the variability of runs allowed. SI Figure 1 shows that both have increased considerably in recent decades. The change in variability is particularly mysterious. It is not attributable merely to an increase in runs allowed per game: even when measured by the coefficient of variation (which effectively standardizes changes in standard deviations among measurements with diverse scales [Chernick & Friis 2003]), the sharp spike in the variance persists.




One might reasonably expect changes in pitcher usage to explain this pattern. In the twentieth century, teams relied on a smaller number of pitchers, who worked many more innings. Today's starters start less often and are pulled much earlier. Relievers tend to face only a few batters per outing. By reducing the volume of individual pitching, these changes could plausibly result in more variance.

But this surmise doesn't seem correct. If one looks at other indicia of pitching proficiency, one does not observe the variance spiral associated with runs allowed. Rates of strikeouts and home runs have exploded in the last two decades of baseball. Yet *variance* in these elements of pitching performance have diminished (SI Figure 2), consistent with an expected and more widely observed general effect toward decreasing variance in sporting performance over time (Gould, 1996; Schell, 1999, 2005).

The increase in runs allowed and in particular individual variance in the same reflects a truly profound development in game dynamics. It is reflected in greater performance spreads not only in BIP-ERA but also in BNIP-ERA (Figure 3.B) and in FIP traditionally understood (Fünf, 2025). This feature in the evolution of the sport demands additional empirical investigation.

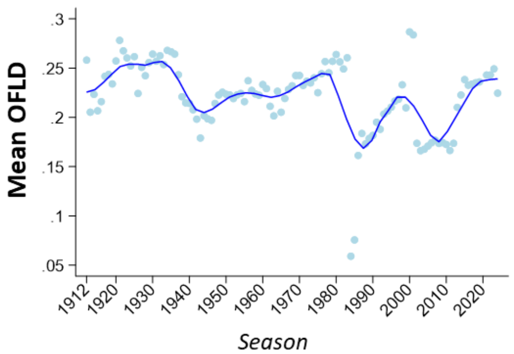


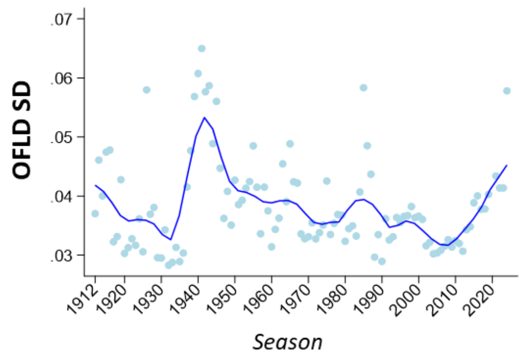

## Retrosheet data: BIP rates and variances

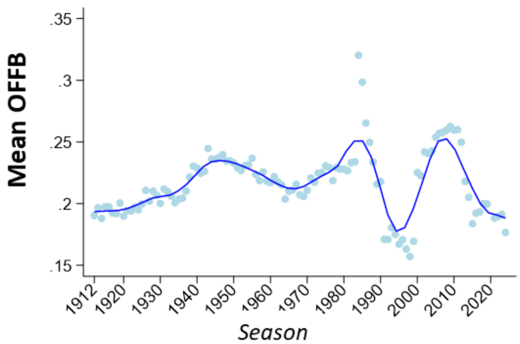
Ball-in-play data used in the study were derived from Retrosheet game reports. Reconstructed from various historical sources and coded by volunteers, the reports will inevitably vary in quality.

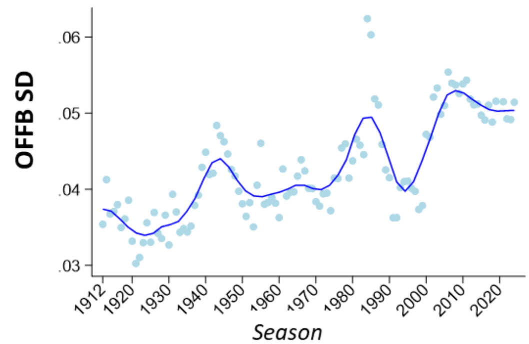
Poor data and coding errors are essentially random and are thus not likely to bias estimates but rather to attenuate them. BNIP-ERA estimates will be less affected by BIP data quality because only one of the five elements they comprise—strikeouts, home runs, walks, hit batters, and balls in play allowed (all measured on a per-inning basis)—depend on BIP data at all and even then do not depend on BIP types. BIP-ERA estimates are much more vulnerable to data quality attenuation. It is, of course, impossible to determine the extent of such attenuation, but it is safe to say that as a result of it, the impact of differences in pitcher BIP propensities reported in the paper is almost certainly understated relative to its true size.

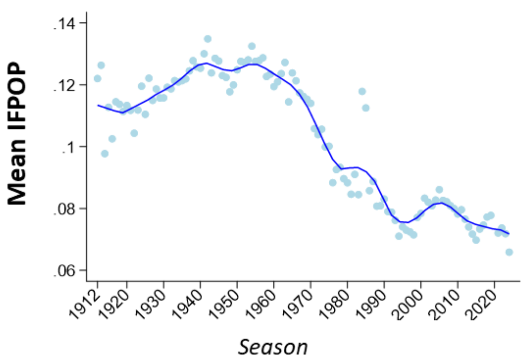
One Retrosheet data anomaly is worthy of note, however. SI Figure 3 reports the weighted means and standard deviations for the BIP variables, IFPOP, IFGB, OFFB, and OFLD. There is an problem for the 1984 and 1985 seasons. It is most obvious in relation to outfield line drives: not only are the means for that form of BIP extremely low in relation to other seasons; but the standard deviations are extremely high. The mean rates for the other BIPs, closer inspection reveals, are higher than in nearby seasons, and also display high SDs. These are the signatures of coding error.

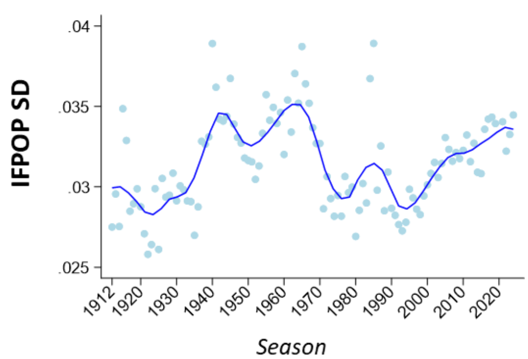

Mean OFLD
.3
.25
.2
.15
.1
.05
1912
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
Season

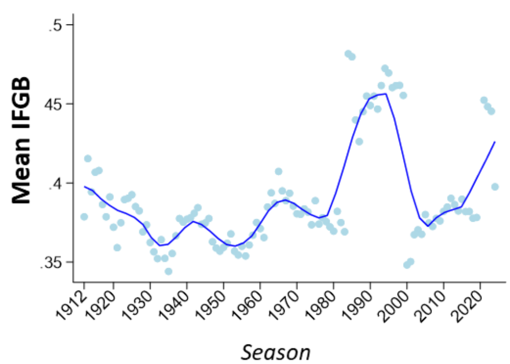

OFLD SD
.07
.06
.05
.04
.03
1912
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
Season

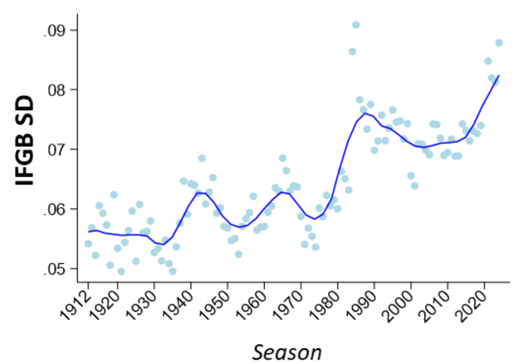

Mean OFFB
.35
.3
.25
.2
.15
1912
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
Season


OFFB SD
.06
.05
.04
.03
1912
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
Season


Mean IFPOP
.14
.12
.1
.08
.06
1912
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
Season


IFPOP SD
.04
.035
.03
.025
1912
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
Season


Mean IFGB
.5
45
.4
.35
1912
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
Season


IFGB SD
.09
08
07
.06
.05
1912
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
Season

The error is almost certainly associated with the Retrosheet scoring of those seasons. The same automated coding algorithm was used to parse the event field data for those seasons as was used for all other seasons, none of which displayed this anomaly. Moreover, the same pattern—low line-drive rates and high standard deviations relative to other seasons—was detected in the companion Project Scoresheet coded balls in play. The likelihood that both my coding and the coding of the Project Scoresheet happened to experience simultaneous errors in these two particular seasons is obviously much smaller than the likelihood that the raw data were miscoded.

No remedial action was taken in the analysis or reporting of the data in the paper. The most obvious effect of the this data problem, however, is the attenuation of any BIP propensity results for those two seasons. It can also be anticipated that the coding error associated with balls in play for those seasons affected the model estimates of the BIP variables for the 1982-86 period as a whole.

Such an effect is corroborated by Figures 3-5, which show differences in BIP-ERA drop precipitously at points corresponding to the 1982-1986 period. In all likelihood, the *true* effect of BIP propensities for the affected seasons is something closer to the ones observed in the 1977-1981 and 1987-1992 periods, which are comparable to one another in size.

Researchers might reasonably decide to re-estimate the BIP parameters for the remaining seasons in the 1982-1986 bin to improve their precision and form more accurate assessments of individual player performances in the unaffected seasons. Ultimately, the best solution would be a rescoring of the primary data on which Retrosheet game reports for the 1984 and 1985 seasons are based; indeed, that is the only step that can generate valid individual performance estimates for those seasons.

The standard model results also suggest that BIP propensities displayed only a minor effect throughout the 1990s. The rates of BIP types and their associated standard deviations, however, do not suggest that this pattern is a result of any defect in the Retrosheet data. It appears to be a genuine effect in game dynamics, the origins of which are unclear, particularly given the rebound of the importance of BIP propensities in the twenty-first century.

### Single-season FIP$_r$ models

The basic model uses pitchers' season FIP$_r$s as one of its non-BIP predictors of runs allowed per game. Sample members FIP$_r$s were computed based on season-specific models that regressed runs allowed per game on the innings-pitched rates of strikeouts, walks, home runs allowed, and hit batters. The models used data weighted on innings pitched.

Because it would be infeasible to reproduce and nearly impossible to comprehend the model outputs in a conventional table, they are instead reported in a downloadable excel file. The model for each season is presented on a separate row. As reflected in the column headings, each row reports (1) b_k, the raw OLS beta coefficient for strikeouts per inning; (2) the associated beta coefficient t-statistic; (3) b_bb, the raw OLS beta coefficient for walks per inning; (3) the associated beta coefficient t-statistic; (4) b_hr, the raw OLS beta coefficient for home runs allowed per inning; (5) the associated beta coefficient t-statistic; (6) b_hbp, the raw OLS beta coefficient for hit batters per inning; (7) the associated beta coefficient t-statistic; (8) the model constant; (9) the associated t-statistic; and (10) R2, the model $R^2$.

### Single-season regression models BIP-ERA, BNIP-ERA

The paper graphically reports the results of season-by-season models used for the purpose of illustrating the incremental contribution of BIP-ERA to variance in runs allowed per game over the expanse of AL/NL history (Figure 4, Figure 7). Two models were conducted for each season from 1912 to

2024: in the first, RAPG was regressed on BNIP-ERA (*model 1*); and in the second RAPG was regressed in the second on BNIP-ERA and BIP-ERA (*model 2*) in order to determine the incremental $R^2$ contribution. The models used data weighted on innings pitched.

Like the single-season $FIP_r$ models (and for the same reasons), the model outputs are reported in a downloadable excel file. The model for each season is presented on a separate row. As reflected in the column headings, each row reports (1) b_nip.1, the raw OLS beta coefficient for BNIP-ERA for model 1; (2) the associated beta coefficient t-statistic; (3) cons.1, the model 1 constant; (4) the associated t-statistic; (5) R2.1, the model 1 $R^2$; (6) b_bnip.2, the raw OLS beta coefficient for BNIP-ERA in model 2; (7) the associated t-statistic; (8) b_bipera, the raw beta coefficient for BIP-ERA in model 2; (9) the associated t-statistic; (10) R2.2, the model 2 $R^2$; (11) cons.2, the model 2 constant; (12) the associated t-statistic; and (12) R2i.bipera, the incremental $R^2$ associated with the addition of BNIP-ERA to model 2.

## Standardized runs

The standard-run variant of the basic model is reported in SI Figure 1.

## Unstandardized runs: individual performance estimates

Raw-run equivalents to the season-best BIP runs saved list, the career-best BIP runs saved list, the career runs saved list, and the knuckleball-pitcher BIP runs saved list (Table 2-4 & Table 6) are reported below in SI Table 2-5).

## WAR

The paper reports the explanatory-power loss for FanGraph's and Baseball Reference's respective pitcher WAR measures attributable to their neglect of pitcher BIP propensities. The regression models used to calculate those deficits is reported in SI Table 7.

In addition, the paper graphically reports the results of season-by-season models used to illustrate this impact over the expanse of AL/NL history (Figure 7). For each WAR measure, two models were conducted for each season from 1912 to 2024; in the first, RAPG was regressed on the relevant set of individual pitcher WAR tallies (*model 1*); and in the second, RAPG was regressed on the relevant pitcher WAR tallies and the same pitchers' BIP-ERAs in order to determine BIP-ERA's incremental $R^2$ contribution. The models were weighted based on innings pitched.

These models are reported in a downloadable excel file. The model for each season is presented on a separate row. Each row reports relevant raw OLS beta coefficients, associated t-statistics, and model $R^2$s. They are denoted as follows, consistent with the column headings:

*1. FanGraphs.* (1) b_fg.1, the raw OLS beta coefficient for FanGraphs WAR for model 1; (2) the associated beta coefficient t-statistic; (3) cons.1, the model 1 constant; (4) the associated t-statistic; (5) R2.1, the model 1 $R^2$; (6) b_fg.2, the raw OLS beta coefficient for FanGraphs WAR in model 2; (7) the associated t-statistic; (8) b_bipera, the raw beta coefficient for BIP-ERA in model 2; (9) the associated t-statistic; (10) R2.2, the model 2 $R^2$; (11) cons.2, the model 2 constant; (12) the associated t-statistic; and (12) R2i.bipera, the incremental $R^2$ associated with the addition of BNIP-ERA to model 2.

*2. Baseball reference.* (1) b_bbr.1, the raw OLS beta coefficient for Baseball Reference WAR for model 1; (2) the associated beta coefficient t-statistic; (3) cons.1, the model 1 constant; (4) the associated t-statistic; (5) R2.1, the model 1 $R^2$; (6) b_bbr.2, the raw OLS beta coefficient for Baseball Reference WAR in model 2; (7) the associated t-statistic; (8) b_bipera, the raw beta coefficient for BIP-ERA in model 2; (9) the associated t-statistic; (10) R2.2, the model 2 $R^2$; (11) cons.2, the model 2 constant; (12) the associated t-statistic; and (12) R2i.bipera, the incremental $R^2$ associated with the addition of BNIP-ERA to model 2.

***Fixed effects***

**pre-BIP effect variables**

| | period | | FIPR | | FIELD | | BIP_IP | |
|---|---|---|---|---|---|---|---|---|
| 1912 | | | **0.57** | (17.13) | **-0.14** | (-5.78) | **0.14** | (5.03) |
| 1917 | 0.01 | (0.34) | **0.51** | (20.41) | **-0.13** | (-5.34) | **0.20** | (7.90) |
| 1922 | -0.02 | (-0.55) | **0.52** | (20.74) | **-0.19** | (-8.13) | **0.12** | (5.25) |
| 1927 | 0.00 | (0.09) | **0.54** | (21.16) | **-0.16** | (-6.86) | **0.11** | (5.06) |
| 1932 | 0.05 | (0.99) | **0.55** | (24.91) | **-0.09** | (-4.52) | **0.16** | (8.29) |
| 1937 | 0.05 | (0.82) | **0.63** | (29.23) | **-0.21** | (-9.92) | **0.04** | (1.59) |
| 1942 | 0.10 | (1.59) | **0.61** | (22.75) | **-0.24** | (-10.41) | **0.07** | (2.98) |
| 1947 | 0.12 | (1.74) | **0.62** | (31.54) | **-0.18** | (-8.07) | **0.04** | (1.94) |
| 1952 | 0.06 | (0.80) | **0.62** | (28.30) | **-0.13** | (-6.88) | **0.02** | (0.91) |
| 1957 | 0.07 | (0.84) | **0.61** | (29.85) | **-0.13** | (-7.12) | **0.13** | (6.14) |
| 1962 | 0.06 | (0.72) | **0.63** | (31.90) | **-0.16** | (-8.54) | **0.12** | (5.36) |
| 1967 | 0.06 | (0.67) | **0.60** | (32.95) | **-0.12** | (-7.26) | **0.12** | (6.03) |
| 1972 | 0.08 | (0.86) | **0.59** | (37.06) | **-0.14** | (-7.84) | **0.14** | (6.41) |
| 1977 | 0.10 | (1.07) | **0.58** | (35.62) | **-0.14** | (-9.70) | **0.17** | (9.68) |
| 1982 | 0.11 | (1.19) | **0.65** | (36.77) | **-0.15** | (-10.23) | **0.13** | (6.03) |
| 1987 | 0.09 | (0.95) | **0.59** | (35.70) | **-0.12** | (-8.97) | **0.33** | (14.85) |
| 1992 | 0.07 | (0.69) | **0.59** | (40.17) | **-0.10** | (-8.90) | **0.39** | (23.82) |
| 1997 | 0.02 | (0.18) | **0.57** | (38.71) | **-0.08** | (-7.39) | **0.37** | (19.37) |
| 2002 | 0.01 | (0.13) | **0.58** | (44.01) | **-0.06** | (-5.71) | **0.36** | (22.55) |
| 2007 | -0.05 | (-0.50) | **0.55** | (40.99) | **-0.09** | (-9.96) | **0.35** | (23.72) |
| 2012 | -0.11 | (-1.01) | **0.57** | (41.92) | **-0.07** | (-7.00) | **0.32** | (23.84) |
| 2017 | -0.14 | (-1.33) | **0.62** | (49.47) | **-0.06** | (-7.36) | **0.31** | (24.29) |

**BIP variables**

| | OFLD | | IFPOP | | IFGB | | OFFB | |
|---|---|---|---|---|---|---|---|---|
| 1912 | **0.17** | (2.43) | -0.08 | (-1.68) | -0.06 | (-0.69) | -0.08 | (-1.37) |
| 1917 | **0.17** | (3.72) | **-0.16** | (-3.82) | **-0.25** | (-3.44) | **-0.19** | (-3.86) |
| 1922 | **0.29** | (5.53) | **-0.10** | (-2.33) | **0.03** | (0.39) | 0.00 | (-0.05) |
| 1927 | **0.14** | (2.79) | **-0.22** | (-4.76) | **-0.24** | (-3.08) | **-0.14** | (-2.74) |
| 1932 | **0.20** | (5.60) | **-0.20** | (-5.19) | **-0.29** | (-4.47) | **-0.21** | (-4.67) |
| 1937 | **0.12** | (1.98) | -0.07 | (-1.83) | -0.09 | (-1.28) | -0.08 | (-1.81) |
| 1942 | **0.25** | (4.20) | -0.03 | (-0.86) | 0.03 | (0.40) | 0.05 | (1.12) |
| 1947 | **0.16** | (4.29) | -0.12 | (-3.48) | **-0.16** | (-3.31) | **-0.12** | (-3.33) |
| 1952 | **0.10** | (2.26) | **-0.21** | (-6.01) | **-0.27** | (-4.91) | **-0.17** | (-4.61) |
| 1957 | **0.12** | (4.11) | **-0.15** | (-4.83) | **-0.22** | (-4.52) | **-0.14** | (-4.08) |
| 1962 | **0.08** | (2.52) | **-0.14** | (-4.44) | **-0.12** | (-2.53) | **-0.08** | (-2.35) |
| 1967 | **0.20** | (6.19) | **-0.12** | (-3.81) | **-0.11** | (-2.05) | **-0.10** | (-2.98) |
| 1972 | **0.11** | (3.45) | **-0.14** | (-4.57) | **-0.18** | (-3.69) | **-0.17** | (-4.46) |
| 1977 | **0.12** | (4.23) | **-0.18** | (-6.82) | **-0.31** | (-6.30) | **-0.24** | (-6.95) |
| 1982 | **0.12** | (2.64) | **-0.09** | (-2.39) | 0.02 | (0.29) | 0.01 | (0.26) |
| 1987 | 0.05 | (1.76) | **-0.09** | (-3.12) | **-0.10** | (-1.97) | **-0.08** | (-2.36) |
| 1992 | 0.04 | (1.75) | **-0.13** | (-6.03) | **-0.15** | (-3.55) | **-0.09** | (-3.76) |
| 1997 | **0.06** | (3.42) | **-0.14** | (-8.08) | **-0.22** | (-7.19) | **-0.16** | (-7.69) |
| 2002 | 0.02 | (1.32) | **-0.19** | (-12.01) | **-0.34** | (-13.19) | **-0.19** | (-9.84) |
| 2007 | 0.03 | (1.67) | **-0.20** | (-11.53) | **-0.34** | (-11.13) | **-0.19** | (-7.89) |
| 2012 | 0.00 | (-0.30) | **-0.19** | (-10.59) | **-0.38** | (-11.92) | **-0.22** | (-10.18) |
| 2017 | **0.07** | (5.82) | **-0.11** | (-10.04) | **-0.14** | (-7.67) | **-0.10** | (-7.71) |

**BIP_IP BIP interactions**

| | OFLD | | IFPOP | | IFGB | | OFFB | |
|---|---|---|---|---|---|---|---|---|
| 1912 | 0.08 | (1.93) | 0.00 | (0.00) | 0.00 | (0.06) | -0.02 | (-0.47) |
| 1917 | **0.08** | (2.60) | -0.01 | (-0.22) | -0.02 | (-0.52) | **-0.06** | (-2.01) |
| 1922 | **0.10** | (3.43) | -0.05 | (-1.73) | 0.01 | (0.16) | 0.03 | (1.22) |
| 1927 | **0.04** | (2.65) | -0.04 | (-1.71) | -0.04 | (-1.55) | **-0.0**4 | (-2.13) |
| 1932 | 0.04 | (1.81) | -0.03 | (-1.10) | -0.06 | (-1.56) | -0.04 | (-1.52) |
| 1937 | **0.09** | (3.74) | -0.03 | (-1.35) | 0.02 | (0.51) | -0.01 | (-0.49) |
| 1942 | **0.12** | (5.30) | 0.01 | (0.42) | 0.01 | (0.54) | 0.03 | (1.60) |
| 1947 | **0.10** | (6.39) | **-0.05** | (-2.73) | -0.01 | (-0.51) | 0.02 | (1.14) |
| 1952 | 0.02 | (0.81) | **-0.06** | (-2.30) | **-0.15** | (-3.95) | **-0.10** | (-3.60) |
| 1957 | 0.04 | (1.72) | -0.04 | (-1.64) | -0.07 | (-1.66) | **-0.08** | (-2.31) |
| 1962 | 0.03 | (0.86) | -0.08 | (-2.83) | **-0.13** | (-2.79) | -0.06 | (-1.92) |
| 1967 | -0.01 | (-0.29) | **-0.11** | (-3.77) | **-0.16** | (-3.09) | **-0.10** | (-3.38) |
| 1972 | 0.03 | (0.89) | **-0.05** | (-2.29) | **-0.13** | (-2.48) | **-0.09** | (-2.20) |
| 1977 | -0.01 | (-0.76) | **-0.06** | (-3.15) | **-0.16** | (-4.66) | **-0.12** | (-4.20) |
| 1982 | 0.01 | (0.18) | -0.04 | (-1.12) | -0.07 | (-0.73) | -0.04 | (-0.61) |
| 1987 | -0.01 | (-0.43) | **-0.10** | (-3.72) | **-0.12** | (-3.11) | **-0.08** | (-2.88) |
| 1992 | 0.00 | (0.15) | **-0.07** | (-4.25) | -0.05 | (-1.81) | **-0.05** | (-2.49) |
| 1997 | 0.00 | (0.22) | **-0.08** | (-3.06) | **-0.12** | (-2.57) | **-0.07** | (-2.74) |
| 2002 | 0.01 | (1.07) | **-0.06** | (-2.95) | **-0.12** | (-3.94) | **-0.05** | (-2.32) |
| 2007 | 0.03 | (1.56) | **-0.06** | (-2.98) | **-0.08** | (-2.49) | -0.04 | (-1.64) |
| 2012 | -0.02 | (-1.56) | **-0.10** | (-6.02) | **-0.20** | (-6.45) | -0.12 | (-6.63) |
| 2017 | -0.02 | (-1.37) | **-0.08** | (-6.11) | **-0.14** | (-5.49) | **-0.10** | (-5.67) |

| | | |
|---|---|---|
| constant | 0.15 | (1.90) |

**BIP joint effects**

| | Joint effects | |
|---|---|---|
| 1912 | **0.09** | 77.25 |
| 1917 | **0.18** | 300.75 |
| 1922 | **0.17** | 251.75 |
| 1927 | **0.18** | 278.63 |
| 1932 | **0.20** | 363.59 |
| 1937 | **0.13** | 163.25 |
| 1942 | **0.14** | 175.71 |
| 1947 | **0.19** | 313.56 |
| 1952 | **0.19** | 338.37 |
| 1957 | **0.18** | 297.26 |
| 1962 | **0.15** | 197.85 |
| 1967 | **0.19** | 316.85 |
| 1972 | **0.17** | 248.44 |
| 1977 | **0.22** | 427.22 |
| 1982 | **0.12** | 135.59 |
| 1987 | **0.14** | 173.23 |
| 1992 | **0.15** | 199.27 |
| 1997 | **0.20** | 355.22 |
| 2002 | **0.24** | 536.80 |
| 2007 | **0.23** | 471.41 |
| 2012 | **0.24** | 497.69 |
| 2017 | **0.23** | 485.06 |

***Random effects***

Individual pitchers

| | | Est. | SE | 95% CI | |
|---|---|---|---|---|---|
| | Variance | 5.06 | 0.87 | 3.61 | 7.10 |
| | Residual | 0.27 | 0.00 | 0.26 | 0.27 |
| *N* | 46,639 | | | | |
| Clusters | 8,942 | | | | |
| $R^2$ | 0.68 | | | | |

**SI Table 1. Basic model, standard runs**. Outcome variable is z_RAPG. Data weighted on innings pitched. Predictors standardized for interpretability and for computational convenience in calculation of BIP-ERA$_s$. MLE coefficients with z-statistics indicated parenthetically. For periods after 1912 (the reference period), coefficients for BIP variables and for BIP_IP and BIP interactions reflect sum of main effect (i.e., reference period) estimate and period-specific effect estimate. "Joint effects" refers to the joint effect of OFLD, IFPOP, IFGB, OFFB, and their interactions with BIP_IP in each period; the effect reported as Cohen's *f* with the joint-effect Wald Test $\chi^2$ reported parenthetically. Model $R^2$ determined by correlation of predicted and observed RAPG values (Devore, 2008, p. 510). Bolded coefficients and Cohen's *f* denote significant at $p < 0.05$.

| rank | Player | Season | IP | BIP-ERA | runs saved | rank | Player | Season | IP | BIP-ERA | runs saved |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Stan Coveleski | 1917 | 298 | -1.35 | 45 | 26 | Vern Kennedy | 1936 | 274 | -0.79 | 24 |
| 2 | Carl Mays | 1921 | 337 | -1.02 | 38 | 26 | Freddie Fitzsimmons | 1928 | 261 | -0.82 | 24 |
| 3 | Jeff Pfeffer | 1914 | 315 | -1.05 | 37 | 26 | Monte Pearson | 1936 | 223 | -0.96 | 24 |
| 4 | Jesse Barnes | 1919 | 296 | -1.11 | 37 | 26 | Curt Davis | 1942 | 206 | -1.03 | 24 |
| 5 | Jim Bagby | 1917 | 321 | -1.00 | 36 | 32 | Waite Hoyt | 1921 | 282 | -0.75 | 23 |
| 6 | Lefty Gomez | 1934 | 282 | -1.09 | 34 | 32 | Red Faber | 1921 | 331 | -0.64 | 23 |
| 7 | Watson Clark | 1935 | 207 | -1.38 | 32 | 32 | Lon Warneke | 1932 | 277 | -0.76 | 23 |
| 8 | Kirby Higbe | 1941 | 298 | -0.95 | 31 | 32 | Dazzy Vance | 1928 | 280 | -0.74 | 23 |
| 8 | Carl Hubbell | 1936 | 304 | -0.93 | 31 | 32 | Curt Davis | 1944 | 194 | -1.06 | 23 |
| 8 | Stan Coveleski | 1918 | 311 | -0.89 | 31 | 32 | Tim Wakefield | 2004 | 188 | -1.09 | 23 |
| 11 | Bob Shawkey | 1916 | 277 | -0.98 | 30 | 32 | Whit Wyatt | 1941 | 288 | -0.71 | 23 |
| 11 | Bob Shawkey | 1922 | 300 | -0.89 | 30 | 32 | Hal Schumacher | 1933 | 259 | -0.79 | 23 |
| 13 | Doug McWeeny | 1928 | 244 | -1.08 | 29 | 32 | Jesse Petty | 1926 | 276 | -0.74 | 23 |
| 12 | Ray Caldwell | 1915 | 305 | -0.84 | 29 | 32 | Larry Benton | 1931 | 204 | -1.00 | 23 |
| 15 | Wilbur Cooper | 1919 | 287 | -0.88 | 28 | 32 | Waite Hoyt | 1927 | 256 | -0.79 | 23 |
| 15 | Derek Lowe | 2002 | 220 | -1.15 | 28 | 32 | Carl Hubbell | 1935 | 303 | -0.67 | 23 |
| 15 | Carl Hubbell | 1932 | 284 | -0.88 | 28 | 44 | Bill Lee | 1936 | 259 | -0.78 | 22 |
| 18 | Urban Shocker | 1926 | 258 | -0.93 | 27 | 44 | Phil Niekro | 1979 | 342 | -0.58 | 22 |
| 18 | Ed Klepfer | 1917 | 213 | -1.12 | 27 | 44 | George Pipgras | 1928 | 301 | -0.66 | 22 |
| 20 | Jim Bagby | 1920 | 340 | -0.70 | 26 | 44 | Larry Benton | 1928 | 310 | -0.63 | 22 |
| 21 | Ken Raffensberger | 1949 | 284 | -0.83 | 26 | 44 | Jeff Pfeffer | 1916 | 329 | -0.60 | 22 |
| 22 | Jeff Tesreau | 1914 | 322 | -0.71 | 25 | 44 | Whit Wyatt | 1942 | 217 | -0.90 | 22 |
| 22 | Hal Gregg | 1944 | 198 | -1.15 | 25 | 44 | Curt Davis | 1943 | 164 | -1.19 | 22 |
| 22 | Jack Warhop | 1914 | 217 | -1.05 | 25 | 44 | JP Sears | 2024 | 180 | -1.08 | 22 |
| 22 | Ray Benge | 1934 | 227 | -0.97 | 25 | 44 | Ed Reulbach | 1914 | 256 | -0.76 | 22 |
| 26 | Stan Coveleski | 1922 | 277 | -0.79 | 24 | 44 | Stan Coveleski | 1920 | 315 | -0.62 | 22 |
| 26 | Rube Marquard | 1913 | 288 | -0.75 | 24 | 44 | Hal Schumacher | 1935 | 262 | -0.74 | 22 |

**SI Table 2. BIP season runs saved.** Derived from basic model (Table 1), BIP-ERA reflects estimated BIP runs saved per 9 innings.

.

| rank | Player | $BIP\text{-}ERA_s$ | RS | BIP RS % | rank | Player | $BIP\text{-}ERA_s$ | RS | BIP RS % |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Carl Hubbell | -0.36 | 143 | 29% | 26 | Ray Fisher | -0.42 | 70 | 92% |
| 2 | Red Ruffing | -0.27 | 132 | 43% | 27 | Mark Buehrle | -0.19 | 69 | < 0% |
| 3 | Freddie Fitzsimmons | -0.35 | 124 | 60% | 28 | Rube Marquard | -0.22 | 68 | 33% |
| 4 | Bob Shawkey | -0.36 | 118 | 64% | 28 | Jesse Barnes | -0.24 | 68 | 38% |
| 4 | Tim Wakefield | -0.33 | 118 | < 0% | 29 | Jesse Petty | -0.50 | 67 | 55% |
| 4 | Warren Spahn | -0.20 | 118 | 25% | 29 | Slim Sallee | -0.28 | 67 | 52% |
| 7 | Stan Coveleski | -0.34 | 115 | 51% | 31 | Al Demaree | -0.42 | 66 | 93% |
| 8 | Phil Niekro | -0.19 | 113 | 54% | 32 | Jim Bagby | -0.31 | 62 | > 100% |
| 9 | Lefty Gomez | -0.39 | 109 | 50% | 32 | Phil Douglas | -0.33 | 62 | 54% |
| 10 | Catfish Hunter | -0.28 | 106 | 83% | 34 | George Pipgras | -0.37 | 61 | 64% |
| 11 | Tim Hudson | -0.30 | 103 | 48% | 35 | Roy Halladay | -0.19 | 58 | 18% |
| 11 | Waite Hoyt | -0.24 | 100 | 34% | 35 | Jack Coombs | -0.53 | 57 | 1649% |
| 13 | Curt Davis | -0.38 | 98 | 41% | 37 | Luis Tiant | -0.15 | 56 | 34% |
| 14 | Derek Lowe | -0.31 | 93 | 62% | 37 | Dutch Ruether | -0.24 | 56 | 90% |
| 15 | Urban Shocker | -0.31 | 92 | 39% | 39 | CC Sabathia | -0.14 | 54 | 20% |
| 15 | Jim Palmer | -0.21 | 92 | 46% | 39 | Fred Toney | -0.23 | 54 | 71% |
| 17 | Hal Schumacher | -0.32 | 89 | 76% | 41 | Johnny Murphy | -0.46 | 53 | 98% |
| 18 | Kirby Higbe | -0.40 | 87 | > 100% | 41 | Greg Maddux | -0.10 | 53 | 9% |
| 19 | Whit Wyatt | -0.43 | 85 | 49% | 41 | Lew Burdette | -0.15 | 53 | 43% |
| 20 | Ray Caldwell | -0.37 | 81 | > 100% | 41 | Monte Pearson | -0.33 | 52 | > 100% |
| 21 | Jeff Pfeffer | -0.29 | 78 | 65% | 45 | Watson Clark | -0.26 | 51 | 30% |
| 21 | Ken Raffensberger | -0.32 | 78 | 37% | 45 | Barry Zito | -0.18 | 51 | > 100% |
| 23 | Jeff Tesreau | -0.41 | 76 | 82% | 47 | Brad Radke | -0.18 | 50 | > 100% |
| 24 | Burleigh Grimes | -0.15 | 72 | 57% | 47 | Jered Weaver | -0.22 | 50 | 53% |
| 25 | Dazzy Vance | -0.21 | 71 | 15% | 49 | Van Mungo | -0.21 | 49 | 31% |
| 26 | Art Nehf | -0.23 | 70 | 46% | 49 | Steve Renko | -0.18 | 49 | > 100% |

**SI Table 3. BIP career runs saved.** Derived from basic model (Table 1). Career BIP-ERA reflects IP-weighted average over seasons played; career BIP RS refers to career BIP runs saved, determined by sum of season BIP standard runs saved over course of career. BIP RS % refers to percentage of career runs saved due to BIP-ERA, calculated in relation to sum of BIP and BNIP standard season runs saved over course of career; "> 100%" indicates that BIP runs saved exceeded the number of BNIP runs saved, "< 0%" that the pitcher's total runs saved were negative on net, and a negative % that BIP runs allowed reduced positive career runs saved on net.

| rank | Player | | BIP-ERA | BIP RS | Total RS | BIP RS % | rank | Player | | BIP-ERA | BIP RS | Total RS | BIP RS % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Randy | Johnson | 0.02 | -8 | 933 | -1% | 26 | Roy | Halladay | -0.19 | 58 | 327 | 18% |
| 2 | Roger | Clemens | -0.02 | 12 | 857 | 1% | 27 | Gaylord | Perry | 0.05 | -29 | 327 | -9% |
| 3 | Pedro | Martinez | -0.03 | 9 | 730 | 1% | 28 | Sandy | Koufax | -0.07 | 18 | 325 | 6% |
| 4 | Walter | Johnson | -0.01 | 7 | 634 | 1% | 29 | Mariano | Rivera | -0.29 | 42 | 316 | 13% |
| 5 | Nolan | Ryan | 0.04 | -25 | 610 | -4% | 30 | Red | Ruffing | -0.27 | 132 | 307 | 43% |
| 6 | Greg | Maddux | -0.10 | 53 | 608 | 9% | 31 | Dizzy | Dean | 0.07 | -16 | 304 | -5% |
| 7 | Clayton | Kershaw | -0.13 | 38 | 597 | 6% | 32 | Gerrit | Cole | 0.03 | -6 | 301 | -2% |
| 8 | Curt | Schilling | 0.03 | -12 | 580 | -2% | 33 | Dennis | Eckersley | -0.08 | 30 | 299 | 10% |
| 9 | John | Smoltz | -0.01 | 5 | 553 | 1% | 34 | Waite | Hoyt | -0.24 | 100 | 298 | 34% |
| 10 | Carl | Hubbell | -0.36 | 143 | 497 | 29% | 35 | Dutch | Leonard | -0.02 | 6 | 296 | 2% |
| 11 | Max | Scherzer | 0.00 | -1 | 489 | 0% | 36 | Paul | Derringer | 0.19 | -77 | 291 | -27% |
| 12 | Pete | Alexander | -0.07 | 35 | 480 | 7% | 37 | Billy | Wagner | 0.02 | -2 | 286 | -1% |
| 13 | Warren | Spahn | -0.20 | 118 | 473 | 25% | 38 | CC | Sabathia | -0.14 | 54 | 274 | 20% |
| 14 | Dazzy | Vance | -0.21 | 71 | 472 | 15% | 39 | Juan | Marichal | -0.11 | 45 | 273 | 16% |
| 15 | Lefty | Grove | -0.04 | 16 | 448 | 4% | 40 | Bert | Blyleven | 0.12 | -64 | 273 | -23% |
| 16 | Justin | Verlander | -0.02 | 9 | 406 | 2% | 41 | Dwight | Gooden | 0.01 | -4 | 262 | -2% |
| 17 | Chris | Sale | 0.00 | -1 | 375 | 0% | 42 | Bret | Saberhagen | 0.01 | -3 | 260 | -1% |
| 18 | Kevin | Brown | -0.07 | 26 | 372 | 7% | 43 | Claude | Passeau | -0.03 | 10 | 260 | 4% |
| 19 | Mike | Mussina | 0.05 | -20 | 371 | -5% | 44 | Felix | Hernandez | -0.03 | 8 | 259 | 3% |
| 20 | Don | Sutton | -0.08 | 45 | 356 | 13% | 45 | Trevor | Hoffman | -0.04 | 5 | 256 | 2% |
| 21 | Steve | Carlton | 0.04 | -23 | 356 | -6% | 46 | Bob | Gibson | 0.07 | -29 | 252 | -11% |
| 22 | David | Cone | 0.03 | -9 | 348 | -3% | 47 | Robin | Roberts | 0.07 | -37 | 251 | -15% |
| 23 | Johan | Santana | -0.11 | 25 | 346 | 7% | 48 | Stephen | Strasburg | 0.06 | -10 | 250 | -4% |
| 24 | Tom | Seaver | 0.05 | -26 | 339 | -8% | 49 | Jake | Peavy | -0.06 | 16 | 249 | 6% |
| 25 | Jacob | deGrom | -0.03 | 5 | 336 | 2% | 50 | Steve | Rogers | -0.09 | 28 | 249 | 11% |

**SI Table 4. Career runs saved, BIP contribution.** Derived from basic model standardized (Table 1). Career BIP-ERA reflects IP-weighted average over seasons played; career BIP RS refers to career BIP runs saved, determined by sum of season BIP standard runs saved over course of career; BIP RS % refers to percentage of runs saved due to BIP-ERA, calculated in relation to sum of BIP and BNIP standard season runs saved over course of career; negative % indicates that BIP runs allowed reduced career runs saved on net.

| rank | Player | IP | BIP-ERA | BIP RS | BP % |
|---|---|---|---|---|---|
| 1 | Tim Wakefield | 3226 | -0.39 | 139 | < 0% |
| 2 | Phil Niekro | 5404 | -0.19 | 115 | 54% |
| 3 | R. A. Dickey | 2074 | -0.33 | 77 | < 0% |
| 4 | Bob Purkey | 2115 | -0.16 | 38 | > 100% |
| 5 | Charlie Hough | 3801 | -0.08 | 35 | -24% |
| 6 | Hoyt Wilhelm | 2254 | -0.14 | 34 | 15% |
| 7 | Eddie Rommel | 2556 | -0.09 | 25 | 80% |
| 8 | Steven Wright | 348 | -0.56 | 22 | < 0% |
| 9 | Tom Candiotti | 2725 | -0.06 | 19 | 23% |
| 10 | Eddie Fisher | 1539 | -0.11 | 19 | 86% |
| 11 | Matt Waldron | 188 | -0.75 | 16 | < 0% |
| 12 | Dutch Leonard | 3218 | -0.01 | 4 | 1% |
| 13 | Steve Sparks | 1043 | -0.03 | 4 | -4% |
| 14 | Al Papai | 240 | -0.06 | 2 | -14% |
| 15 | Wally Burnette | 263 | -0.06 | 2 | 24% |
| 16 | Charlie Haeger | 83 | -0.08 | 1 | -5% |
| 17 | Eddie Gamboa | 13 | -0.44 | 1 | 61% |
| 18 | Jared Fernandez | 109 | 0.08 | -1 | 9% |
| 19 | Eddie Cicotte | 2368 | 0.04 | -10 | -6% |
| 20 | Wilbur Wood | 2684 | 0.04 | -13 | -32% |
| 21 | Johnny Niggeling | 1251 | 0.11 | -16 | > 100% |
| 22 | Roger Wolff | 1025 | 0.16 | -19 | < 0% |
| 23 | Mickey Haefner | 1467 | 0.13 | -20 | 78% |

**SI Table 5. Knuckleball pitchers.** Derived from Basic Model (Table 1). Career BIP-ERA is career weighted average of season BIP-ERAs (Schell, 1999, 2005). “Career BIP RS” is career sum of the pitcher’s BIP runs saved. “BIP RS %” is the percentage of the pitcher’s career total runs saved attributable to runs saved by virtue of his BIP propensities; “> 100%” indicates that BIP runs saved exceeded the number of BNIP runs saved, “< 0%” that the pitcher’s total runs saved were negative on net negative, and a negative % that BIP runs allowed reduced positive career runs saved on net.

**1912-19**

| | pitcher | IP | BIP-ERA | RS |
|---|---|---|---|---|
| 1 | Rube Marquard | 1780 | -0.44 | 87 |
| 2 | Jeff Pfeffer | 1502 | -0.47 | 79 |
| 3 | Jeff Tesreau | 1679 | -0.41 | 76 |
| 4 | Ray Caldwell | 1583 | -0.39 | 68 |
| 5 | Slim Sallee | 1885 | -0.32 | 67 |
| 5 | Al Demaree | 1424 | -0.42 | 66 |
| 7 | Stan Coveleski | 1148 | -0.49 | 63 |
| 8 | Jack Coombs | 955 | -0.52 | 55 |
| 9 | Ray Fisher | 1291 | -0.36 | 51 |
| 10 | Jim Bagby | 1123 | -0.35 | 44 |
| 10 | Jack Warhop | 680 | -0.58 | 44 |

**1920s**

| | pitcher | IP | BIP-ERA | RS |
|---|---|---|---|---|
| 1 | Waite Hoyt | 2346 | -0.48 | 124 |
| 2 | Urban Shocker | 2149 | -0.39 | 93 |
| 3 | Bob Shawkey | 1613 | -0.48 | 85 |
| 4 | Jesse Petty | 1128 | -0.61 | 76 |
| 5 | Herb Pennock | 2313 | -0.26 | 68 |
| 6 | Dazzy Vance | 2054 | -0.28 | 63 |
| 7 | Burleigh Grimes | 2798 | -0.20 | 62 |
| 8 | Art Nehf | 1720 | -0.31 | 59 |
| 9 | Fred Fitzsimmons | 1021 | -0.50 | 57 |
| 10 | Stan Coveleski | 1934 | -0.25 | 53 |

**1930s**

| | pitcher | IP | BIP-ERA | RS |
|---|---|---|---|---|
| 1 | Red Ruffing | 2439 | -0.47 | 127 |
| 2 | Carl Hubbell | 2597 | -0.40 | 116 |
| 3 | Lefty Gomez | 2235 | -0.45 | 111 |
| 4 | Hal Schumacher | 1737 | -0.40 | 77 |
| 5 | Monte Pearson | 1296 | -0.40 | 57 |
| 6 | Johnny Murphy | 679 | -0.72 | 54 |
| 7 | Van Mungo | 1715 | -0.27 | 51 |
| 8 | Watson Clark | 1174 | -0.35 | 46 |
| 9 | Roy Parmelee | 1113 | -0.35 | 44 |
| 10 | Fred Fitzsimmons | 1938 | -0.18 | 40 |

**1940s**

| | pitcher | IP | BIP-ERA | RS |
|---|---|---|---|---|
| 1 | Curt Davis | 1061 | -0.93 | 109 |
| 2 | Kirby Higbe | 1693 | -0.50 | 95 |
| 3 | Whit Wyatt | 1015 | -0.70 | 79 |
| 4 | Les Webber | 432 | -0.93 | 44 |
| 5 | Ed Head | 465 | -0.81 | 42 |
| 6 | Warren Spahn | 990 | -0.35 | 39 |
| 7 | Hal Gregg | 785 | -0.43 | 38 |
| 8 | Bill Voiselle | 1322 | -0.25 | 36 |
| 9 | Rube Melton | 704 | -0.46 | 36 |
| 10 | Larry Jansen | 785 | -0.37 | 32 |

**1950s**

| | pitcher | IP | BIP-ERA | RS |
|---|---|---|---|---|
| 1 | Warren Spahn | 2823 | -0.22 | 69 |
| 2 | Ken Raffensberger | 919 | -0.63 | 64 |
| 3 | Lew Burdette | 1864 | -0.21 | 43 |
| 4 | Harry Perkowski | 667 | -0.54 | 40 |
| 5 | Bob Rush | 2047 | -0.16 | 37 |
| 5 | Warren Hacker | 1097 | -0.30 | 37 |
| 7 | Joe Nuxhall | 1340 | -0.24 | 36 |
| 8 | Frank Smith | 496 | -0.55 | 30 |
| 9 | Max Surkont | 1098 | -0.23 | 28 |
| 10 | Bubba Church | 713 | -0.33 | 26 |
| 10 | Ewell Blackwell | 635 | -0.37 | 26 |

**1960s**

| | pitcher | IP | BIP-ERA | RS |
|---|---|---|---|---|
| 1 | Juan Marichal | 2550 | -0.12 | 34 |
| 2 | Lew Krausse | 758 | -0.31 | 27 |
| 3 | Joe Horlen | 1608 | -0.15 | 26 |
| 3 | Denny McLain | 1502 | -0.16 | 26 |
| 5 | Catfish Hunter | 1050 | -0.21 | 25 |
| 6 | Eddie Fisher | 1028 | -0.20 | 23 |
| 8 | Phil Niekro | 888 | -0.22 | 22 |
| 9 | Luis Tiant | 1200 | -0.16 | 21 |
| 9 | Dick Hall | 980 | -0.18 | 20 |
| 10 | Bobby Bolin | 1282 | -0.14 | 20 |

**1970s**

| | pitcher | IP | BIP-ERA | RS |
|---|---|---|---|---|
| 1 | Catfish Hunter | 2399 | -0.30 | 81 |
| 1 | Phil Niekro | 2881 | -0.25 | 79 |
| 3 | Jim Palmer | 2745 | -0.23 | 71 |
| 4 | Don Wilson | 1125 | -0.38 | 47 |
| 5 | Steve Renko | 1846 | -0.21 | 43 |
| 6 | Luis Tiant | 2063 | -0.17 | 40 |
| 7 | Joaquin Andujar | 636 | -0.49 | 35 |
| 8 | Carl Morton | 1619 | -0.17 | 31 |
| 9 | Ross Grimsley | 1863 | -0.14 | 30 |
| 9 | Ken Forsch | 1271 | -0.21 | 30 |

**1980s**

| | pitcher | IP | BIP-ERA | RS |
|---|---|---|---|---|
| 1 | Don Sutton | 1766 | -0.19 | 37 |
| 2 | Eric Show | 1497 | -0.16 | 27 |
| 3 | Scott McGregor | 1604 | -0.15 | 26 |
| 3 | Joe Price | 841 | -0.28 | 26 |
| 5 | Mario Soto | 1614 | -0.14 | 24 |
| 5 | Steve McCatty | 968 | -0.23 | 24 |
| 7 | Mike Norris | 712 | -0.30 | 23 |
| 8 | Dennis Eckersley | 1594 | -0.11 | 20 |
| 8 | Tom Browning | 1211 | -0.15 | 20 |
| 8 | Jeff Reardon | 872 | -0.20 | 20 |
| 8 | Matt Keough | 773 | -0.23 | 20 |

**1990s**

| | pitcher | IP | BIP-ERA | RS |
|---|---|---|---|---|
| 1 | Greg | 2395 | -0.10 | 25 |
| 2 | David Wells | 1897 | -0.11 | 23 |
| 3 | Fernando Valenzuela | 785 | -0.23 | 20 |
| 4 | Joey Hamilton | 1033 | -0.17 | 19 |
| 5 | Jason Schmidt | 736 | -0.22 | 18 |
| 5 | Brad Radke | 1085 | -0.15 | 18 |
| 7 | Tom Candiotti | 1760 | -0.09 | 17 |
| 7 | Pat Hentgen | 1556 | -0.10 | 17 |
| 8 | Al Leiter | 1181 | -0.12 | 15 |
| 8 | Andy Ashby | 1343 | -0.10 | 15 |
| 8 | Dave Fleming | 610 | -0.22 | 15 |

**2000s**

| | pitcher | IP | BIP-ERA | RS |
|---|---|---|---|---|
| 1 | Derek Lowe | 1834 | -0.43 | 87 |
| 2 | Tim Wakefield | 1747 | -0.44 | 86 |
| 3 | Tim Hudson | 1923 | -0.32 | 68 |
| 4 | Mark Buehrle | 2061 | -0.24 | 54 |
| 5 | Barry Zito | 1999 | -0.23 | 51 |
| 6 | Roy Halladay | 1883 | -0.24 | 50 |
| 7 | Jose Contreras | 1084 | -0.28 | 34 |
| 8 | Freddy Garcia | 1571 | -0.18 | 32 |
| 8 | Carlos Zambrano | 1551 | -0.19 | 32 |
| 8 | Greg Maddux | 1940 | -0.15 | 32 |
| 10 | Brandon Webb | 1320 | -0.21 | 31 |
| 10 | Jamie Moyer | 1980 | -0.14 | 31 |

**2010s**

| | pitcher | IP | BIP-ERA | RS |
|---|---|---|---|---|
| 1 | Marco Estrada | 1231 | -0.31 | 42 |
| 2 | R. A. Dickey | 1631 | -0.23 | 42 |
| 3 | Jered Weaver | 1396 | -0.25 | 39 |
| 4 | Dallas Keuchel | 1302 | -0.23 | 34 |
| 5 | Hiroki Kuroda | 1018 | -0.27 | 31 |
| 5 | Clayton Kershaw | 1996 | -0.13 | 29 |
| 7 | Tim Hudson | 1067 | -0.25 | 29 |
| 8 | Jared Hughes | 519 | -0.46 | 26 |
| 9 | Hector Santiago | 921 | -0.25 | 26 |
| 10 | CC Sabathia | 1688 | -0.14 | 26 |

**2020s**

| | pitcher | IP | BIP-ERA | RS |
|---|---|---|---|---|
| 1 | JP Sears | 423 | -0.52 | 24 |
| 2 | Corbin Burnes | 757 | -0.27 | 23 |
| 3 | George Kirby | 512 | -0.35 | 20 |
| 4 | Zack Wheeler | 758 | -0.24 | 20 |
| 5 | Kutter Crawford | 392 | -0.45 | 19 |
| 6 | Joe Ryan | 470 | -0.33 | 17 |
| 7 | Jameson Taillon | 641 | -0.24 | 17 |
| 8 | Tyler Rogers | 301 | -0.49 | 16 |
| 9 | Framber Valdez | 710 | -0.20 | 16 |
| 10 | Matt Waldron | 188 | -0.73 | 15 |
| 11 | Adrian Houser | 425 | -0.23 | 11 |
| 12 | Tarik Skubal | 539 | -0.18 | 11 |

**SI Table 6. Decade BIP runs saved leaders.** Derived from basic model (Table 1). "BIP RS" formed by summing season BIP runs saved for indicated decade.

| | *Model 1.A* | | *Model 1.B* | | *Model 2.A* | | *Model 2.B* | | *Model 3.A* | | *Model 3.B* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FGWAR | **-0.24** | (-103.8) | **-0.22** | (-101.15) | | | | | | | | |
| BBRWAR | | | | | **-0.26** | (-106.61) | **-0.23** | (-97.93) | | | | |
| z_BNIP_ERA | | | | | | | | | **0.73** | (126.48) | **0.66** | (129.45) |
| z_BIP_ERA | | | **0.41** | (45.61) | | | **0.35** | (86.39) | | | **0.30** | (37.88) |
| cons | **0.44** | (70.96) | **0.40** | (71.34) | **0.48** | (91.86) | **0.42** | (35.62) | 0.00 | (-0.07) | 0.00 | (0.03) |
| *N* | 43,026 | | 42,980 | | 43,026 | | 42,980 | | 43,013 | | 42,989 | |
| $R^2$ | 0.19 | | 0.36 | | 0.33 | | 0.45 | | 0.54 | | 0.63 | |
| $\Delta R^2$ | | | **0.17** | | | | **0.12** | | | | **0.09** | |

**SI Table 7. Incremental explanatory power added by BIP propensities.** *N*'s are individual pitcher seasons, 1912-2024. Outcome variable is z_RAPG. OLS coefficients, *t*-statistics noted parenthetically. Along with outcome variable, BNIP- and BIP-ERA are standardized by season to remove the effect of inter-season scaling variability unrelated to the impact of the predictors on the outcome variable (Schell, 1999, 2005). WAR variables are not standardized because they are by design made to reflect a common scale across seasons (wins above replacement). Data weighted by innings pitched. Bolded predictors and $\Delta R^2$s are significant at $p < 0.05$

## Supplemental Information Bibliography


Chernick, M.R. & Friis, R.H. Introductory Biostatistics for the Health Sciences: Modern Applications Including Bootstrap (Wiley, 2003).

Devore, J.L. *Probability and Statistics for Engineering and the Sciences* (Thomson/Brooks/Cole, 2008).

Gould, S.J. Full house : the spread of excellence from Plato to Darwin (Harmony Books, New York, 1996).

Fünf, X. Gould’s Revenge. Bbcardstats.com (Jan. 18, 2025). Available at https://bbcardstats.com/goulds-revenge/.

Schell, M.J. Baseball's All-Time Best Hitters: How Statistics Can Level the Playing Field (Princeton University Press, 1999).

Schell, M.J. Baseball's All-time Best Sluggers: Adjusted Batting Performance from Strikeouts to Home Runs (Princeton University Press, 2005).