# What is the best season by a pitcher in British top-tier baseball since 2001?

Joe Gray ★ 21 March 2009

In a recent article that I wrote to look at the same question posed in the title but for batters,[1] I settled on the concept of dominance as a means of cross-season comparisons (represented numerically by the z-score), and we shall do the same here for pitchers. But what will be the best choice of statistic to reflect a pitcher's performance? To measure a pitcher's contribution to the team, it is necessary to separate out the effect of the defence.

## Defence-independent pitching statistics

Earned-run average goes some way to removing the effect of fielding quality by not penalizing the pitcher for errors made by the defence; however, it is not truly defence independent because committing errors is only one way in which fielders can disadvantage a pitcher (for instance, a shortstop who makes few errors but has a poor range may offer less support to a pitcher than one with a better range but who makes more errors).

To be truly defence independent, a statistic can use only those events that essentially bypass the fielders,[2] namely over-the-fence home runs, strike-outs, bases on balls, and hit batters. Crucially, this list does not include any hits other than home runs. The strength of considering only these statistics is boosted by research suggesting that:

> *"There is little if any difference among major-league pitchers in their ability to prevent hits on balls hit in the field of play."* [3]

Another way of looking at it is that there is a much better scope for differentiating between pitchers in those plate appearances that end either in a home run or in a pitch not being hit into play.

The expression below is one of several statistics that have been designed to combine all of the raw defence-independent numbers in an informative fashion:[4]

$$\frac{(13 * HR) + (BB + HBP - IBB) - (2 * K)}{Innings\ pitched} + Const$$

In the expression, *Const* represents a league-specific constant that places the number on an ERA-equivalent scale and is usually in the neighbourhood of 3.2.

## Solving a problem with home runs

Home runs receive a heavy weighting in the above expression, but in the relatively short British season there is plenty of room for flukily low or unluckily high totals. The combination of the heavy weighting and this lack of precision leads to the potential for the large proportion of luck contained in the raw data of home runs to cloud the richer information provided by free passes and strike-outs. For this reason, home runs conceded will be dropped from the list of defence-independent events to be considered.

Removing home runs leaves three factors to be combined: unintentional free passes,[5] strike-outs, and innings pitched. Given that there are just two defence-independent pitching events left, we can simply divide one by the other and ignore innings pitched. Doing this means that we do not need to worry about whether free passes and strike-outs are weighted correctly, since the weighting is no longer needed (there is no point weighting the whole of the top or the bottom term of a fraction as it will make no difference to your conclusions when you come to compare different values, which is what we are really interested in). So the expression that we shall use for

defence-independent pitching statistics in the current analysis is as shown below, with strike-outs placed on top of the ratio so that higher numbers are better:

$$\frac{K}{(BB + HBP - IBB)}$$

It contains less information than the expression that we started with, but its increased resilience to the large scope for random error that is characteristic of the short British baseball season justifies the simplification.

## Answering our question

Now that we have decided upon our statistic, it is a fairly straightforward task to compute the dominance scores (z-scores) for each pitcher with at least 40 innings pitched in a season, which is the cut-off used by the Great Britain Baseball Scorers Association (GBBSA) for single-season pitching records.[6] Ranking the findings gives us the top-ten list shown in Table 1, with Simon Pole's 2003 season with the London Warriors sitting in the number-one position. Pole's following two seasons with the Warriors also feature in the top ten, so it is perhaps unsurprising that he leads all pitchers with at least 225 innings pitched (the GBBSA cut-off for career pitching records) in the statistic between 1995 and 2008. What is more remarkable is that Pole also leads the single-season batting dominance list.[1]

| # | Player | Team | Year | Raw | z-Score |
|---|---|---|---|---|---|
| 1 | Simon Pole | London Warriors | 2003 | 6.667 | 2.308 |
| 2 | Dean Stoka | Windsor Bears | 2003 | 6.417 | 2.259 |
| 3 | Cody Cain | London Warriors | 2004 | 4.714 | 2.191 |
| 4 | Dean Stoka | Windsor Bears | 2004 | 4.533 | 2.142 |
| 5 | Ryan Koback | Windsor Bears | 2004 | 4.333 | 2.086 |
| 6 | Brian Essery | London Mets | 2008 | 6.333 | 1.947 |
| 7 | Alex Keprta | London Warriors | 2005 | 3.308 | 1.722 |
| 8 | Simon Pole | London Warriors | 2005 | 3.133 | 1.654 |
| 9 | Simon Pole | London Warriors | 2004 | 2.850 | 1.561 |
| 10 | Adam Lemke | Croydon Pirates | 2006 | 5.500 | 1.553 |

**Table 1.** The top-ten most dominant single-season pitching performances, assessed using the z-score of a defence-independent pitching stat, in British top-tier baseball between 2001 and 2008.

## Effect of the quality of other pitchers on dominance

Dominance was chosen as the means of correcting for inter-season differences at the start of this article for consistency with the sister article on batting.[1] There is a potential problem with this approach, though, in that it corrects not just for the difficulty of the pitching environment but also, unfortunately, for the quality of other pitchers.[7]

To illustrate this, first imagine a season with an average set of batters, and a group of pitchers who are all average except for one stand-out pitcher. Then imagine that in the next season the batters are the same average players from the season before, but three average pitchers are all replaced with pitchers who are all well above average, but not as good as that stand-out pitcher from the first season. Even if that stand-out pitcher registers an identical set of numbers in the second season, the z-score for this pitcher will probably be lower than that registered in the first season, even though the pitching environment was no more difficult. The z-score would be lower because the overall quality of pitching was better.

A means of potentially side-stepping this issue is to rank the pitchers once more, but this time using just the raw statistic. This tests how sensitive our initial findings are to a potentially flawed assumption in making cross-season comparisons. In general, such tests are known as sensitivity analyses. If it is found that the result still holds when the assumption is removed or changed, then more confidence can be placed on the initial finding.

Fortunately, producing raw rankings still places Pole's 2003 season at the top of the list, so we can now more confidently conclude that this is the right answer to the question posed in the title. Still, it is worth noting that Brian Essery's 2008 season with the London Mets would jump from sixth place to third place, while Adam Lemke's 2006 performance for the Croydon Pirates would move up from tenth to fourth.

## Conclusion

By using defence-independent pitching statistics, but discounting home runs conceded because of the scope for flukily low or unluckily high totals in the

relatively short British season, we have found that Simon Pole's 2003 season with the London Warriors was both the most dominant and the best unadjusted season since 2001 for a pitcher who threw at least 40 innings.

To comment on this article, please visit
http://www.baseballgb.co.uk/?p=2122

## Notes

1 See http://www.baseballgb.co.uk/?p=2120.

2 These raw statistics do not totally bypass fielders: strike-outs, for instance, rely in part on the way that the catcher sets up.

3 This quote is taken from Voros McCracken's article on defence-independent pitching statistics, which constitutes essential sabermetric reading: http://baseballprospectus.com/article.php?articleid=878. It might be that McCracken over-stated the reality in the article (for a critique, see http://www.diamond-mind.com/articles/ipavg2.htm), but there is little doubt that the ability of pitchers to prevent hits on batted balls in play is considerably less than it was previously thought to be.

4 This is taken from *The Hardball Times'* glossary (http://www.hardballtimes.com/main/statpages/glossary), where the concept is credited to Tangotiger (http://www.tangotiger.net/). Note that only unintentional walks are considered, as the intentional walk is not a particularly helpful means of assessing a pitcher's talent.

5 Unintentional free passes includes all walks that were not intentional as well as all hit batters (some of the latter may have been intentional events, of course, but no trace of malice is left in the statistical summary of a game).

6 Technical note: before calculating the z-scores, the pitching statistic chosen must be logged to normalize its distribution.

7 The same weakness does apply to using the concept of dominance with batting statistics, but to a lesser extent since the overall curve of pitching performances is based on far few players than the overall curve of batting performances, meaning that an individual pitcher can have a much bigger impact on the curve than an individual batter.