Never Swat an Infield Fly

by Dave Studeman

June 20, 2004

Hopefully, you've noticed that we collect a lot of batted ball information at The Hardball Times. I have to admit that I'm fascinated by the information, and I constantly look up the number of line drives so-and-so has hit, or what's-his-name has given up, just to see what it implies about who's-his-face. Now if only I could remember names ...

We're building on the work of many other people in this area, including the **original, provocative formation of DIPS** by Voros McCracken, and **his toned-down second look**, as well as **Tom Tippett's study**. I think the best synopsis of how to think like a DIP comes from Voros' second article:

> 1. The amount that MLB pitchers differ with regards to allowing hits on balls in the field of play, is much less than had been previously assumed. Good pitchers are good pitchers due to their ability to prevent walks and homers and get strikeouts in some sort of combination of those three.
>
> 2. The differences that do exist between pitchers in this regard are small enough so that if you completely ignore them, you still get a very good picture of the pitcher's overall abilities to prevent runs and contribute to winning baseball games.
>
> 3. That said, the small differences do appear to be statistically significant if generally not very relevant.

How we separate blame and credit between pitching and fielding is one of the most interesting aspects of baseball research these days. The issue is critical to total player evaluation systems, such as **Win Shares**, that strive to assign total credit and blame for a team's performance. Even if pitchers don't vary a lot in their ability to influence hits on balls in play, significant differences do exist between pitchers within any given year. And it is important to understand the cause of these differences if you

want to evaluate the contribution of pitchers and fielders.

Let me give an example. The Expos have been very good at preventing runs this year. Their **FIP is slightly below average, but their DER is the best in the league**. This would normally lead you to believe that their fielding is most responsible for that good run prevention record.

However, Expo pitchers have only given up lines drives at a .157 rate, which is way low — more than twenty points below average. So their pitchers do deserve a lot of credit for preventing those runs. Now, Expo pitchers probably won't keep yielding line drives at such a low rate for the rest of the year, but that's sort of irrelevant for understanding what's happened so far this year. If you were to evaluate the relative contribution of the Expos' pitchers and fielders, you'd want to know this sort of thing.

Baseball statistics companies, such as Stats Inc. and **Baseball Info Solutions**, have been collecting play-by-play data for many years now. Among the bits of data they collect are specific information on each batted ball, including the type of ball that was hit and where it landed. This has enabled baseball analysts, such as **Mike Emeigh** and **Mitchel Lichtman** (also known as MGL) to dive further into the pitching/fielding contribution issue.

It was actually an article from MGL called **DIPS Revisited**, as well as Mike Emeigh's comments on **this thread at Baseball Primer**, that got us thinking about collecting batted ball data for The Hardball Times. So we've been collecting groundball, flyball and line drive data this season, and we took an **in-depth look at the data a month and a half ago**. Some of our findings were:

- 79% of flyballs are fielded for outs, 72% of groundballs are fielded for outs, but only 26% of line drives are fielded for outs.
- Batters differ in their ability to hit line drives — at least more than pitchers differ in their ability to prevent them.
- 90% of home runs are flyballs, and the remaining 10% are line drives.
- Line drives, flyballs and walks help the offense the most, while groundballs and strikeouts are the worst things for an offense.

However, if you read MGL's article, you'll see that there is one other type of batted ball we need to track: infield flies. This was also a key insight of Michael Humphries' cutting-edge fielding study, Defense Regression Analysis (which was published at Baseball Primer during the offseason, though it doesn't seem to be available now). There are two reasons infield flies are important:

1. They are caught for an out 97% of the time.
2. Pitchers actually seem to have different levels of skill of inducing them.

So I'm happy to announce that our friends at Baseball Info Solutions are now sending us the number of infield flies allowed by each pitcher, and we are now reporting that info in our **American League** and **National League** pitching stats.

More from The Hardball Times

A Hardball Times Update
by RJ McDaniel
Goodbye for now.

What does this new data tell us? Well, let me start with a simple table, which lays out the percent of time each batted ball occurs, how often it is converted into an out, and how often it is hit for a home run. This data is based on 2004 data, through games of June 17.

Type	Percent	Out%	HR%
Groundballs	45%	72%	0%
OF Flyballs	30%	75%	12%
Line Drives	19%	26%	2%
IF Flyballs	6%	97%	0%

When you separate infield flies from outfield flies, you get some interesting results. For instance, the out percentage of outfield flies is only somewhat higher than groundballs — it is the infield fly that makes a flyball more desirable from a pitching perspective. Also, a non-caught outfield fly has about an even chance of being a home run or in-park hit.

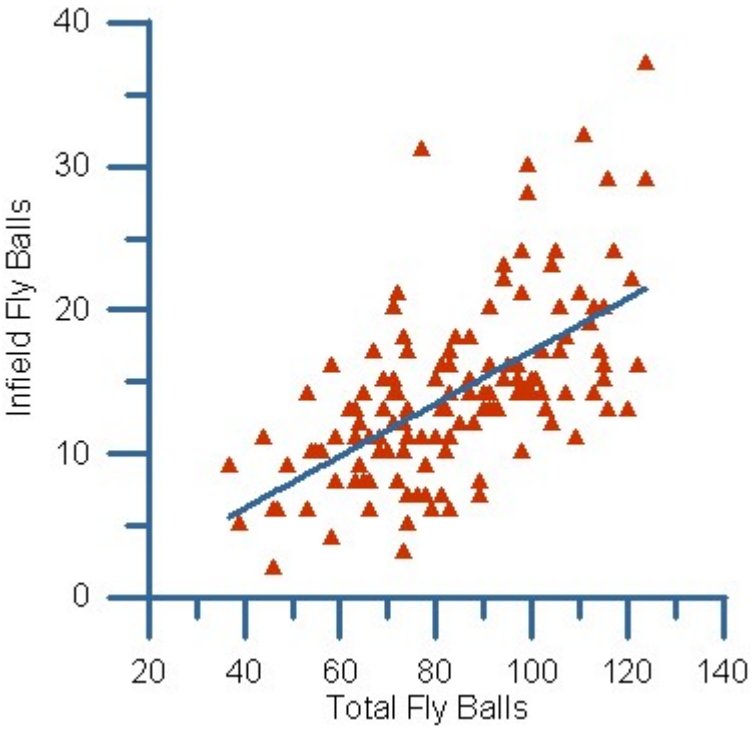
To get a better handle on what these stats mean, I calculated a run impact for each

type of batted ball by assigning a value for each out (-.27), home run (1.4) and hit allowed (educated guesses that vary by type of ball). The source of these values is the second part of Tangotiger's seminal analysis, **How Runs Are Really Created**.

Batted Ball	Run Impact
IF Flyballs	-0.25
Groundballs	-0.06
OF Flyballs	0.06
Line Drives	0.36

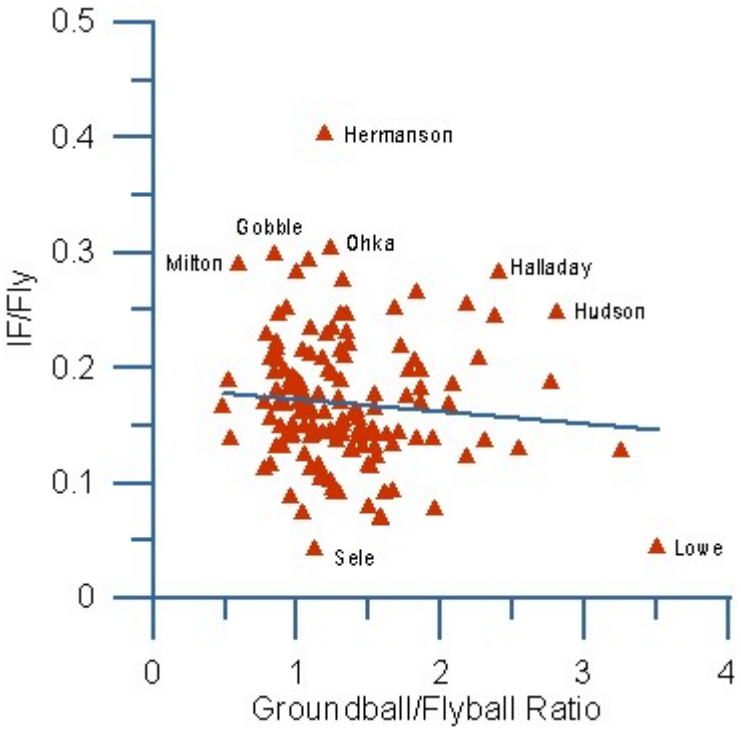
As you can see, hitters want to hit line drives, and pitchers want to induce infield flies. The difference is big.

We've already discussed the control that hitters have over line drives. But which pitchers are most likely to induce infield flies? Here's a graph of all pitchers who have faced at least 200 batters this year (sample size of 141 pitchers), with the total number of flyballs allowed and infield flyballs allowed:



There's no surprise here. Flyball pitchers allow more infield flies. For this reason, MGL reasoned that the most important measure here is the **percent of flyballs that are infield flies**, instead of just total infield flyballs. We label this measure IF/Fly in our data.

So let's change the graph a little bit. I'll put each pitcher's groundball/flyball ratio on the "X" axis (so that groundball pitchers are on the right and flyball pitchers are on the left) and IF/Fly on the "Y" axis. I'll also label a few of the outliers:



Now, the line slopes down slightly, which means that flyball pitchers are slightly more likely to induce infield flies than groundball pitchers. But there are some wild outliers on this graph — especially Dustin Hermanson. Let's take a closer look at their specific stats.

First, let's look at the overall league averages for the stats. Specifically, I'm going to list Runs Allowed Per Game (RA), the percent of batted balls that are line drives (LD%), the ratio of Groundballs/Flyballs (G/F), the proportion of total flyballs that are infield flies (IF/Fly), the percent of batted balls that are converted into outs by fielders (DER), and the three basic FIP rate stats.

	RA	LD%	G/F	IF/Fly	DER	K/9	BB/9	HR/9
American	4.92	.178	1.21	.176	.689	6.3	3.5	1.1
National	4.66	.182	1.30	.176	.698	6.6	3.4	1.1

You probably knew that the American League is the higher scoring league. But the two leagues also differ a bit in their batted balls. The National is more of a groundball league, with a slightly higher LD%. As a result, the league DER is higher (line drives drive DER down, but groundballs drive it up; park factors are also

important). Interestingly, the basic rate stats are virtually even, particularly IF/Fly!

Okay, let's look at the stats of some of our outliers, to see what we can see:

Player	Team	RA	LD%	G/F	IF/Fly	DER	K/9	BB/9
Hermanson D.	SFG	4.62	.123	1.19	.403	.702	6.4	2.6
Ohka T.	MON	3.77	.115	1.23	.303	.709	4.3	2.0
Gobble J.	KC	5.17	.157	.85	.298	.752	2.2	2.3
Milton E.	PHI	4.73	.225	.59	.288	.699	7.2	4.0
Halladay R.	TOR	4.25	.117	2.41	.282	.703	6.6	2.7
Hudson T.	OAK	3.04	.162	2.81	.247	.707	4.3	1.9

Wow. **Dustin Hermanson** is allowing line drives at a very low rate, and his IF/Fly rate is a staggering 40%. Yet he's allowed more than the average number of home runs per inning pitched, and about the average number of runs per game. One problem he's having is that he's not pitching well with runners in scoring position. But maybe his fielders are letting him down a bit, too.

Tomo Ohka is really having a fine season all around, or at least he was before his injury. **Eric Milton** is an extreme flyball pitcher, and he's also giving up line drives at an alarming rate. That won/loss record is very misleading. And look at **Roy Halladay**: all his batted ball stats are great. He's a groundball pitcher who has given up line drives at a very low rate and infield flies at a very high rate. Yet his DER is only .703. Fielders, you think?

Here are some of the outliers from the bottom of the graph.

Player	Team	RA	LD%	G/F	IF/Fly	DER	K/9	BB/9
Sele A.	ANA	3.72	.152	1.12	.041	.717	4.5	3.4
Lowe D.	BOS	6.65	.173	3.50	.043	.677	4.1	4.4
Thomson J.	ATL	5.14	.194	1.49	.079	.661	6.3	2.3
Pavano C.	FLO	2.81	.173	1.23	.101	.766	5.3	2.0
Loaiza E.	CHW	4.18	.165	1.18	.108	.739	5.2	2.7

It's hard to see why **Pavano** has such a high DER. His line drive and GB/FB ratios are about league average and his IF/Fly ratio is low. Either he's got some great fielders, or he's been lucky. Meanwhile, **John Thomson** is at the other end of the spectrum; he seems to deserve that low DER despite being a groundball pitcher.

Aaron Sele is a flyball pitcher who has not been inducing a lot of infield flies. That line drive percentage helps his performance. And it looks like you can put **Esteban Loaiza** in the lucky/good fielders bucket.

Okay, enough with the individual pitchers. Let me try to summarize what I think we know.

- There are definitely groundball and flyball pitchers, and there seem to be some pitchers who are somewhat better at inducing infield flies. They tend to be flyball pitchers.

- Line drives have a big impact on the game, and pitchers may differ quite a bit in the number of line drives they've allowed in a given year. However, most major league pitchers don't seem to have different levels of innate ability to stop line drives from being hit.

- Hitters, on the other hand, do have different line drive hitting abilities.

- If you want to establish credit and blame for allowing or preventing runs, batted ball types can tell you a lot.

Our next step will be to craft a formula for predicting DER and ERA based on batted ball types, park factors and FIP. Y'all come back now.

References & Resources

As always, our thanks to **Baseball Info Solutions** for their data. As a reminder, the definition of a batted ball type depends on the trajectory of the batted ball. Groundballs hit the ground before leaving the infield, line drives come off the bat on a relatively flat trajectory and flyballs follow an upward path off the bat. Infield flies are defined as flyballs that land in the infield (or would have landed in the infield if not caught). If an infielder runs into the outfield to catch a fly, it is still considered an outfield fly, because it would have landed there.

And don't forget that line drives are not necessarily hit harder than flyballs or groundballs. Batted ball type is not a function of how hard the ball is hit.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Comments are closed.