How Do Pitchers Age?

by Phil Birnbaum

It's conventional sabermetric wisdom that players improve up to the age of 27, then start a slow decline that weeds them out of the league sometime in their 30s. Can we quantify the effects of aging? Is there a way to try to figure out, for instance, exactly—or even approximately—how much we expect the average player to decline between the ages of 30 and 37?

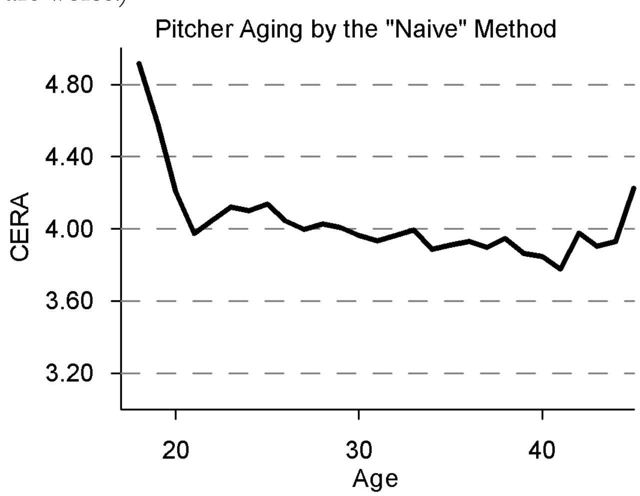
The question appears to be pretty straightforward, but it's harder than it looks. I won't have an answer in this article, at least not a specific numerical one. But we're going to have fun looking.

The Naive Method

One way of looking at the issue, which I'll call the "naive" method, is to just measure the skill of players at various ages. What's the average performance of a 21-year-old player, or a 26-year-old, or a 34-year-old?

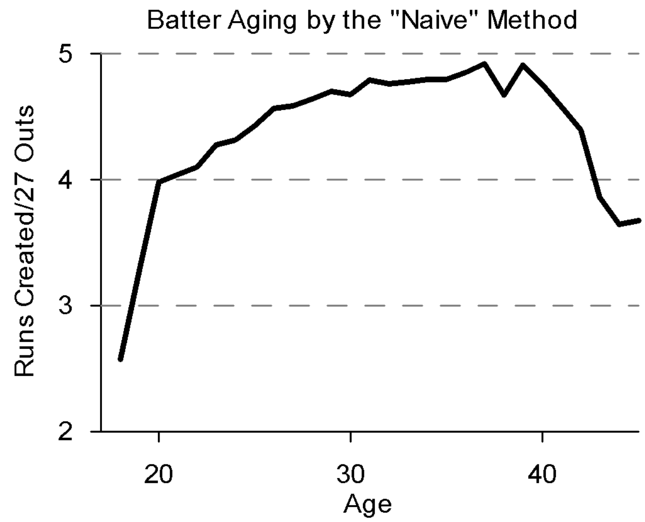
Let's start with pitching. I took all pitching player-seasons from 1947 to 2007 and calculated the pitcher's Component ERA. (Component ERA, or CERA, is a measure of how many earned runs per nine innings the pitcher "should have" given up based on the composite batting line of the batters who faced him.) I then calculated the overall CERA by pitcher age.

Here's what the result looked like. (Keep in mind that the graph is denominated in CERA, so higher numbers are worse.)



It's pretty flat after age 22, which suggests that you can't tell how good a pitcher is by just knowing his age.

For batters, I measured their proficiency in Runs Created per 27 outs (RC27), which is an estimate of how many runs a team would score if its batting line comprised nine copies of the same player. As it turns out, the equivalent curve for batting isn't quite what we expected. It shows a steady increase from ages 21 to 39. Here's the graph. (Higher is better, of course, for the batting curves.)



So just looking at these two curves, we might think that age doesn't matter a lot, at least between, say, 24 and 40. But that would be wrong.

On the graph, it looks like 35-year-olds do at least as well as 30-year-olds. But what's the graph is telling us is that those 35-year-olds in the major leagues are as good as 30-year-olds in the major leagues. But that's not what we want to measure. We want to know how much players decline between 30 and 35, and to do that, we need to compare the same players at both ages.

But the group of 35-year-olds in the above graph aren't the same players as the group of 30-year-olds. There are a lot more batters age 30 than batters age 35. Between 30 and 35, many batters decline so much that they're out of the league. So when you look at the 35-year-olds, you're seeing only the survivors, the players good enough to stick around into (baseball) old age.

For example, Damaso Garcia created 3.7 runs per game at age 30. Mike Schmidt created 8.9.

But, at age 35, Garcia was long out of baseball. Schmidt, of course, was still going strong, but not quite as strong as when he was 30—his RC27 was only 7.3.

Garcia declined, from 3.7 to out of baseball. Schmidt declined, from 8.9 to 7.3. But if you take the average at each age and ignore the fact that Garcia didn't last, you get:

- Average of two players at age 30: 6.3
- Average of one player at age 35: 7.3

So you wind up with your conclusion completely backwards: this analysis seems to say that players improve when they age, whereas in reality, they all declined! The naive analysis severely underestimates any decline.

You can construct more elaborate examples than the simple two-player model I've done here. Or you can take a look at Bill James' debunking of this naive method back in the 1982 Baseball Abstract.

And you can even do a real-life example. Pick an arbitrary year and find all the 30-year-old hitters. Go five years later and look for the same players. You'll find that:

- Many of the players are out of the league by 35.
- Of the players still around, most of them will have experienced significant declines.

but ...

- If you compare the average 30-year-old to the average surviving 35-year-old, there won't be that huge a difference.

So the naive analysis doesn't work. Let's try something else.

Paired-Seasons Analysis

Suppose we find all the 23-year-olds in the league last year and note how they did. Then let's check how they did this year at age 24. Conventional wisdom is that 23-year-olds are improving at that young age, so we should find, on average, that we see better performance at 24 than at 23.

The big question though is, how much weight do you give to each player?

As a 23-year-old in 1979, Rance Mulliniks created 1.2 runs per game, but he had only only 65 at-bats. The next season, now with Kansas City, he created 3.8 runs per game, but again, in very little playing time: only 54 at-bats.

On the other hand, Steve Sax was already a regular in his early 20s. In 1983, his RC27 was 4.1 as a full-time player with 623 at-bats. In 1984, he dropped to 2.9 runs per game and lost a bit of playing time (but still had 569 at-bats).

Mullinks improved a lot in very few at-bats. Sax dropped significantly, but as a full-time hitter. We shouldn't weight them equally, should we?

Common sense says that since Sax had about 10 times as many plate appearances as Mulliniks, we should weight him 10 times as much.

But what happens if a player has significantly different numbers of at-bats in his two seasons? For instance, Jerry White had 278 at-bats in 1976 when he was 23. But he only had 21 at-bats in 1977. What do we do then?

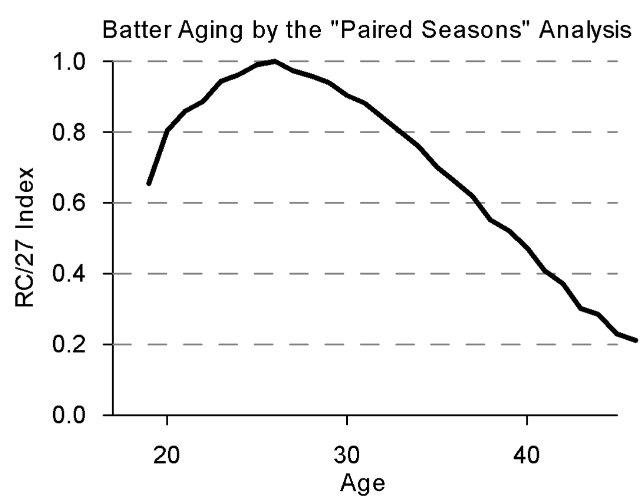
Typically, in studies like this, analysts use the smaller of the two at-bat figures of the two seasons and use that as the weight. So you wind up with these weights:

- Rance Mulliniks: +2.6 in 54 at-bats
- Steve Sax: -1.2 in 569 at-bats
- Jerry White: -2.0 in 21 at-bats

That seems like a reasonable way to do it. And if you go ahead and take the average of these three guys, with these weights, you get a difference of –0.91, and you'd conclude that hitters decline by almost a whole run between the ages of 23 and 24.

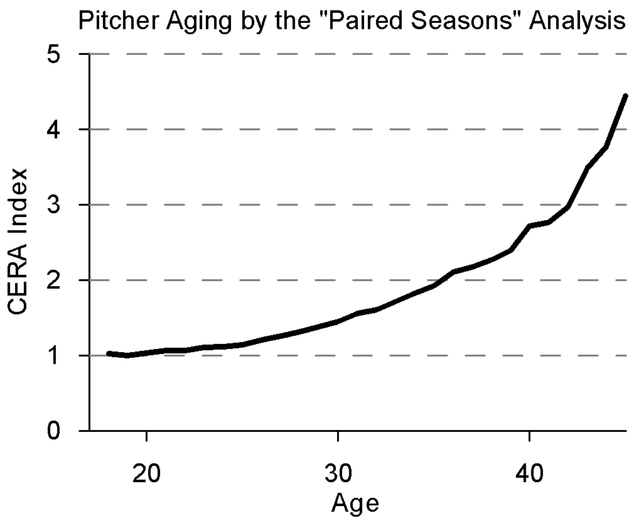
Of course, this is only three arbitrarily selected players. I reran this study, but using all hitters from 1947 to 2007 (I also weighted seasons by outs instead of at bats). And it turns out that the results confirm the conventional wisdom: from age 23 to 24, hitters improve from 4.54 to 4.68. That's about a 1.7 percent increase in performance, as measured by RC27.

We can repeat this analysis for every pair of ages. Then we'd be able to compare any two different ages just by multiplying the percentages together. I did that, and here's what I got:



The graph shows performance "relative to peak;" it turns out the peak is at 26, a bit short of what conventional wisdom (and various studies, including that Bill James chapter from 1982) holds.

Here's the same study repeated for pitching:



It's a whole other story for pitching. Instead of seeing young pitchers improve, as we did for young hitters, we see the peak is at age 20 and that pitchers decline immediately and continuously from there.

Other than that, the results don't look too unreasonable. Sometimes they look a bit exaggerated—a pitcher with an ERA of 3.30 at age 27 projects to over 5.00 by age 34—but you have to keep in mind that we included players who are well out of the league by then. It could be that half the 27-year-olds are still around at 34 and have an overall ERA of 4.00. But the other half are out of baseball, and if you made them come back and pitch, they'd be over 6.00. Half at 4.00 and half at 6.00 does indeed work out to 5.00. So the results aren't necessarily unreasonable.

Paired Seasons Analysis: The Flaw

But this analysis suffers from a problem: the same kind of selective sampling we saw in the naive method.

There's a fair bit of luck in individual player statistics. A pitcher with a talent level that would typically produce a 4.00 ERA could easily get lucky and have a good year with an ERA of 3.00. Or he might get unlucky and wind up at 5.00.

Those occurrences tend to even out—there'll be about as many 3.00s as 5.00s in the long run. And the luck usually won't repeat itself next year; both the lucky pitchers and unlucky pitchers will wind up back at their normal levels in subsequent seasons.

However, it's likely that the pitcher with the 5.00 ERA won't get as much playing time next year as the pitcher who went 3.00. So while both pitchers may wind up at 4.00, they'll do it in different numbers of innings. Suppose the lucky pitcher gets 200 innings of work, but the unlucky one gets only 100 innings. The two lines then look like:

- Pitcher L: 3.00 ERA in 200 innings last year; 4.00 ERA in 200 IP this year
- Pitcher U: 5.00 ERA in 200 innings last year; 4.00 ERA in 100 IP this year

Remember that for each pitcher, we weight by the lower of their two innings pitched. Which means:

- Pitcher L: declines 1.00 ERA in 200 innings
- Pitcher U: improves 1.00 ERA in 100 innings

If you average those out, giving pitcher L twice the weight of pitcher U, you get an overall decline of 0.33 ERA.

See what's happened? By our assumptions, there was actually zero decline over the two seasons. But because the pitcher more likely to decline was given more innings, we wind up with a selective sampling problem, and then we observe an overall decline where we shouldn't have.

The problem is even more acute when some players retire. If pitcher U was released after his unlucky season, we'd be left only with pitcher L. He'd decline by a whole run, there would be nobody to offset him, and we'd wind up estimating a 1.00 decline—much worse than 0.33 and much farther from the real-life value of zero.

This method, then, will always cause us to overestimate declines, and underestimate improvements. The graphs are almost certainly overly pessimistic, for both batters and pitchers, because of this selective sampling issue. Players probably don't decline anywhere near as quickly as what the curves seem to tell us.

Correcting the Selective Sampling Problem

So is there a way to fix this problem? Maybe. The problem was caused, mostly, by players getting lucky or unlucky in the first of the two years. Perhaps if we correct for that, we might be able to get better results.

In general, a player's talent is closer to average than his performance indicates. A pitcher with an ERA of 2.00 is probably not that good; a pitcher with an ERA of 6.00 is probably not that bad. Same for hitters—a batter who hits .340 is probably performing a bit over his head. And a guy who hits .220 is probably just having a bad year.

(To see why this is true, consider a guy who goes 3-for-4. He's certainly not a .750 hitter; he's probably a .290 hitter or something who just had a good day. The same is true, to a lesser extent, for a week, month or season. That means that batters who hit .340 for the season are probably, on average, just .310 hitters who had a good year.)

Suppose that we have three players this year, each of whom gets 500 at-bats. One of them hits .300, one of them hits .270, and the third hits .240.

- Player A: .300 in 500 AB
- Player B: .270 in 500 AB
- Player C: .240 in 500 AB

In reality, player A is a .285 hitter who got lucky. Player B is a .270 hitter who did about what was expected. Player C is a .255 hitter who got unlucky. (And again, let's assume no age-related changes.)

Because of their performance, they won't all get 500 at-bats again next year. Player A will, because he hit .300. But player B might get only 400 at-bats, and Player C might get released.

So next year, only players A and B survive and hit exactly according to their abilities:

- Player A hits .285, for a decline of .015 in 500 at-bats.
- Player B hits .270, for a decline of .000 in 400 at-bats.
- Player C is gone.

Overall, we find a decline of .0083. This contradicts our omniscient knowledge that the real change is zero.

That is, the paired-seasons method is overly pessimistic; it will show a steeper age-related decline than actually exists.

But suppose that instead of measuring the actual decline, we measured the decline based on expected talent. Then, instead of saying that player A declined from .300 to .285, we say that he stayed level, from a talent level of .285 to an actual performance of .285. If we do that, we get exactly the right answer.

Of course, that leads us to another difficult problem: without being God, how do we know player A's actual talent? We don't. But even a decent guess at his talent will give us a better estimate. Let's use a technique called regression to the mean.

We know that a player who hits .300 is probably less than a .300 hitter. Suppose we bring him down to .290 (instead of .285). And suppose we don't regress the .270 hitter at all, because he's right at the league average. Then,

- Player A declines .005 in 500 at-bats
- Player B declines .000 in 400 at-bats

And we get a decline of .0027 points—still not the exact figure of .0000, but much closer than the unregressed estimate of .0083.

If we regressed perfectly and brought player A all the way down to .285 where he belongs, we'd get exactly the correct answer of .000.

So if you're able to take the players in Year 1 and regress them to the mean by the exact amount so that the regressed estimate matches their talent, you'll eliminate the effects of selective sampling and get the correct results.

Regressing

What I'll do now is try to find the appropriate average level of regressing to the mean and apply that level to all the players. But again, that requires an answer to a hard question: how much regression is appropriate? I don't know. For now, the best I can do is to try a couple of different amounts and see what comes out.

For a first attempt, let's try this: we'll regress enough so that a pitcher with 200 innings pitched winds up 10 percent closer to the mean. To do that, we add 22.2 innings of average pitching to his record. (That gives him 222.2 innings; 10 percent of that is the 22.2 we added.)

Tom Tango has pointed out that if you've chosen the correct number of innings to add for one pitcher, it's the same number for all pitchers. So we'll add 22.2 innings of average ERA to every pitcher in the sample. Those with 200 innings will regress 10 percent, and those with only 22 innings will regress 50 percent (which is appropriate, since a pitcher with a 2.00 ERA in 22 innings is much more likely to be average than a pitcher with a 2.00 ERA and 200 innings).

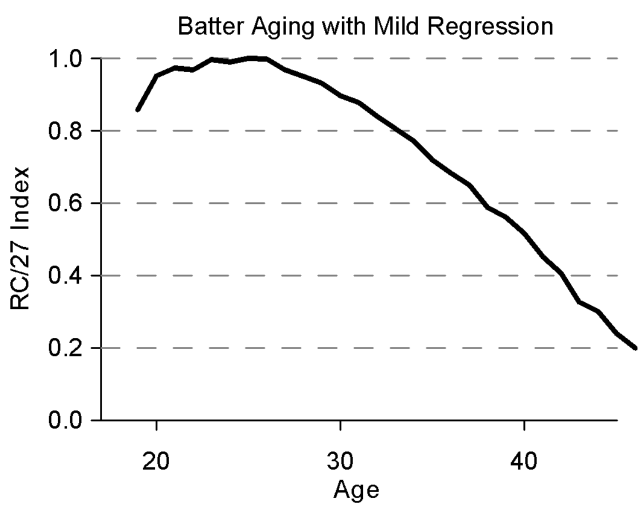
If we do that, and rerun the matched pairs calculations, we get an aging curve that's less steep. Here it is for pitching:

Now, instead of 40-year-old pitchers having three times the ERA of the young guys, it's only about two-and-a-half times.

We can regress even more. Here's another pitching graph, adding 84 average innings (about a 30 percent regression for a pitcher with 200 innings):

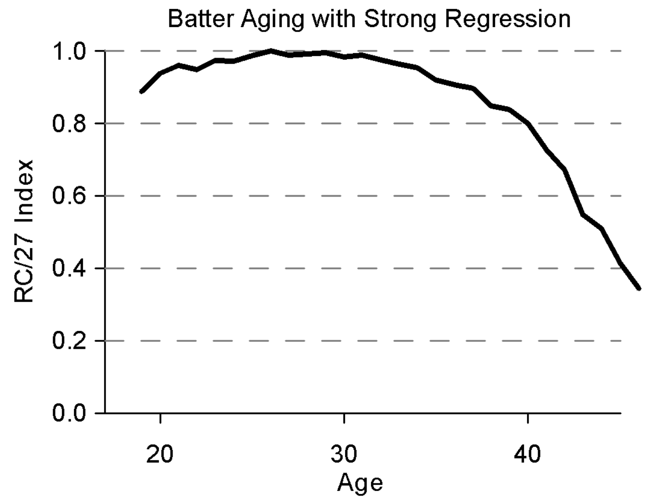
It's really flat now, perhaps too flat. A pitcher at 4.50 at age 30 now rises to only 5.40 by age 40. Intuitively, it seems like maybe we regressed too much.

For hitters, let's start by adding 66 league-average outs, which moves a player with 400 outs about 14 percent closer to the mean:



The curve flattens a bit. Without regression, there was an obvious increase from 19 to 26. Here, once the hitter reaches age 20, it's pretty flat to age 28.

We can regress even more—I've added 257 league-average outs to each batter (regressing 39 percent for a batter with 400 outs) to the following graph:



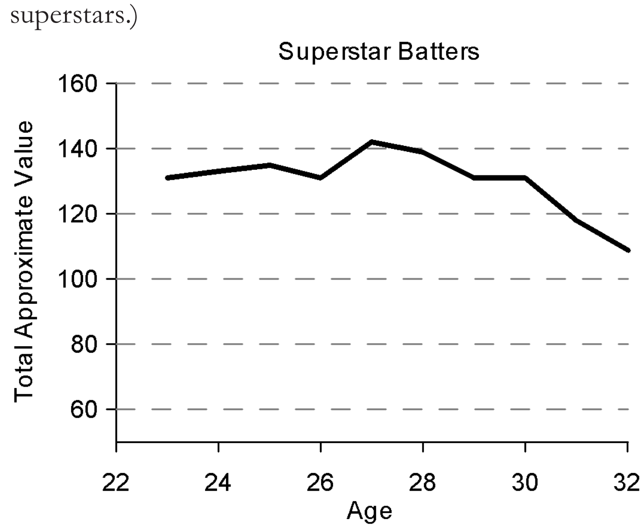
Again, this is probably a bit too flat.

So which curve is right? The honest answer is, I don't know. It depends on what the correct level of regression is. That's a question for further research.

However, there is one piece of objective evidence we can use right now to help us decide. In his 1982 study, Bill James gave us the total of "approximate values" for hitters at every age. The grand totals aren't all that useful for our purpose here, because "approximate value" includes playing time, and we want to look at rate statistics, not playing-time-dependent statistics.

We're lucky, however, in that Bill broke down the hitters by talent level. One of his categories is "superstar." We can probably assume that superstar hitters should all be full-time players regardless of the ups and downs of their performance, which means that we don't have to worry about varying playing time.

Here are those superstars by age. (I've included only ages 23-32, out of fear that outside that range, playing time variations might become significant even for superstars.)



At these ages, the curve seems pretty flat, nothing like the graceful rise we saw for young players in the non-regressed case. Actually, I think this curve is closest to the "14 percent" curve for hitters, so I'd bet that curve is reasonably accurate, at least for superstar-caliber hitters.

But comparing curves by eye is pretty weak, and Bill James is measuring something different than we are, so I'd say any conclusions here should remain tentative for now.

A Puzzle: Why Don't Pitchers Appear to Improve?

In the hitting curves, we see that hitters improve up to approximately age 27, then start to decline. But the pitching curves show that pitchers seem to decline almost immediately! They start out great at age 20, then get worse and worse and worse until they're out of the league.

At first, it seems this must be wrong. There are a lot more 25-year-old pitchers in the major leagues than 21-year-old pitchers. But if younger pitchers are better, it should be the other way around. Indeed, if there are more MLB pitchers at 25 than at 21, you'd expect to see the same kind of curve we saw in the first graph.

So something must be going on. What could it be? A few possibilities:

Injuries

It could be that healthy pitchers don't decline as much as the graph suggests. The overall population is a combination of healthy young pitchers who might stay in the league and improve, and unhealthy young pitchers, who decline badly and may no longer stay in baseball. That effect might be stronger for pitchers than hitters; it does seem like batters don't have anywhere near as many career-ending injuries as pitchers do.

(However, if that's the case, then it's still true that the average pitcher declines. He declines because of injury, but that's still something a general manager or rotisserie player has to keep in mind when projecting his future.)

To check, I looked for "collapses," which I defined as a young pitcher going from above average to below replacement level (league ERA + 1) the next season. There were indeed more of those pitcher collapses than hitter collapses. So maybe this is part of the answer.

Maturity

In a blog conversation, someone suggested that perhaps pitchers have to learn to master a few different pitches before making the major leagues. That happens suddenly, in their early 20s. At that point, the pitcher would get called up but not necessarily progress any further.

In that case, you'd have a combination of two things:

- Most young pitchers don't yet have the mastery to pitch in the major leagues.
- Young pitchers' arms start to decline at a very young age.

That would explain why young pitchers get worse and worse, and also why there aren't many 21-year-olds despite the fact that their arms are stronger.

This hypothesis explains the data, but I'm not sure how plausible it is.

Statistical Illusion

Remember how the selective-sampling issue made declines look worse than they actually were by giving lucky players (who are set for a decline) more innings than unlucky players (who are set for an improvement)? That effect could be larger for pitchers than for hitters.

Suppose you have two hitters with .315 talent, but one hits .300 and the other hits .330. They'll both get full-time jobs with equal numbers of plate appearances, and so forth.

Next year, both return to .315. One declines, the other improves, but because they have the same playing time, the effects cancel out to zero.

Now suppose you have two star pitchers, both expected to have an ERA of 2.75. But, as the season plays out, one winds up with an ERA of 2.50, the other with an ERA of 3.00. They probably didn't get equal playing time. They might both be full-time starters, and each might have gotten 30 starts. But the 3.00 pitcher will have pitched fewer innings than the 2.50 guy, simply because he'll have had more bad games, where he'll have been taken out earlier. (For the sake of this example, let's suppose that you lose 25 innings for every 0.25 increase in ERA.)

Next year, when both pitchers revert to 2.75, the situation looks like this:

- Pitcher A: Last year, 2.50 in 250 innings. Next year, 2.75 in 225 innings.
- Pitcher B: Last year, 3.00 in 200 innings. Next year, 2.75 in 225 innings.

Again remembering that we use the lesser of the two innings, we get:

- Pitcher A: decline of .25 in 250 innings
- Pitcher B: improvement of .25 in 200 innings

Because of the differences in playing time, these average out to a decline of .027—not zero, as it did for the hitters.

So the playing time effect is worse for pitchers than for hitters. Is the extra effect enough to make the young pitchers appear to decline when they're actually improving? I'm not sure, but it's something to look at anyway.

Conclusions

If you're a general manager thinking of signing a 33-year-old slugger, and you want to know the future expectations for players of his age—well, other than being pretty sure that there's going to be a decline, there's not much we can tell you. What looks like a simple problem really isn't.

How can we get a better answer to the question? The most direct way would be to force teams to give their players the exact same amount of playing time from year to year, at least until they turn 40. Since that's not likely to happen, we'll have to hope someone comes up with better, more ingenious ways to measure the effects of aging.

Many of the ideas in this paper grew out of discussions with Tom Tango and Mitchel Lichtman, on our respective blogs. Thanks to Tom and Mitchel and the commenters.