By the Numbers

Volume 16, Number 4 The Newsletter of the SABR Statistical Analysis Committee November, 2006

Review

Academic Research: Errors and Official Scorers

Charlie Pavitt

The author describes a recent academic study investigating the change in error rates over time, and speculating on the role of the official scorer in the "home field advantage" for errors.

This is one of a series of reviews of sabermetric articles published in academic journals. It is part of a project of mine to collect and catalog sabermetric research, and I would appreciate learning of and receiving copies of any studies of which I am unaware. Please visit the Statistical Baseball Research Bibliography at www.udel.edu/communication/pavitt/biblioexplan.htm. Use it for your research, and let me know what is missing.

David E. Kalist and Stephen J. Spurr, Baseball Errors, Journal of Quantitative Analysis in Sports, Volume 2, Issue 4, Article 3

In its short existence, JQAS has shown a tendency to present articles that are long on method but short on interesting substance (case in point, another piece in Volume 2 Issue 4 relevant to the tired old topic of within-league parity). Kalist and Spurr's effort is a welcome change. It is, in short, an analysis of changes over the years in error rate, based on two data types: aggregated data for teams across seasons, and game-by-game data from Retrosheet starting with 1969.

Don't we already know that error rate has gone down substantially over the years? And aren't errors an unreliable measure of fielding given that they are defined by possibly biased scorers sitting in the press box? Sure, but the authors performed their study in a manner that allowed them to ask and answer a series of questions about factors that might impact on error rates, including the issue of bias. Herein lies the value of their effort.

Many of the findings based on the aggregate data (error rates per season) are no surprise. Error rates have declined over time, as we well know. First-year expansion teams made significantly more errors per season than the norm, as did the replacement players stocking the majors during World War II. Stolen bases per game, used as a proxy for team speed, were positively related with errors; others have previously noticed the speed/error association.

Interestingly, the National League has consistently "boasted" more errors than the American League; the authors are unsure why, but comparisons both before and after the appearance of the designated hitter in the junior circuit indicate that this is probably not the reason.

Things get more interesting when the authors move on to the Retrosheet data (error rates per game). All of the findings for the aggregated data are replicated; error rates have declined in an approximately linear fashion by about one-fourth since 1969. Error rates are lower in the warmer months than in September and, in particular, April; in night games (probably due to more consistent lighting conditions) rather than day, and on artificial turf (here, more even surface) as compared to grass.

And, last but most, errors are greater for home teams than for visitors. Although the authors claim that it is possible that home teams are more error free due to familiarity with the ballpark, other research attempting in vain to determine the reasons for the consistent 54% home field advantage in baseball (many of which I have reviewed here over the years) makes this explanation unlikely. The more probable culprit, both in my eyes and those of the authors, is biased umpiring (or, in this case, biased official

In this issue

scoring). When the home team is at bat, questionable calls go in favor of the batter to improve his batting average, decreasing visitor error rate; if any analogous bias exists when the home team is in the field, it would be to increase home team errors to protect home team pitchers' ERAs (or so the authors speculate). The general finding of referee bias is apparently not unique, as Kalist and Spurr cite similar findings in soccer, basketball, and figure skating. However, the bias has decreased since the free-agent era began; the authors propose the possibility that the greater amounts of money involved in player compensation may have made official scorers take their task more seriously.

Charlie Pavitt, chazzq@udel.edu ♦

Informal Peer Review

The following committee members have volunteered to be contacted by other members for informal peer review of articles.

Please contact any of our volunteers on an as-needed basis - that is, if you want someone to look over your manuscript in advance, these people are willing. Of course, I'll be doing a bit of that too, but, as much as I'd like to, I don't have time to contact every contributor with detailed comments on their work. (I will get back to you on more serious issues, like if I don't understand part of your method or results.)

If you'd like to be added to the list, send your name, e-mail address, and areas of expertise (don't worry if you don't have any - I certainly don't), and you'll see your name in print next issue.

Expertise in "Statistics" below means "real" statistics, as opposed to baseball statistics - confidence intervals, testing, sampling, and so on.

Member	E-mail	Expertise
Shelly Appleton	slappleton@sbcglobal.net	Statistics
Ben Baumer	bbaumer@nymets.com	Statistics
Jim Box	jim.box@duke.edu	Statistics
Keith Carlson	kcsqrd@charter.net	General
Dan Evans	devans@seattlemariners.com	General
Rob Fabrizzio	rfabrizzio@bigfoot.com	Statistics
Larry Grasso	l.grasso@juno.com	Statistics
Tom Hanrahan	Han60Man@aol.com	Statistics
John Heer	jheer@walterhav.com	Proofreading
Dan Heisman	danheisman@comcast.net	General
Bill Johnson	firebee02@hotmail.com	Statistics
Mark E. Johnson	maejohns@yahoo.com	General
David Kaplan	dkaplan@UDel.Edu	Statistics (regression)
Keith Karcher	karcherk@earthlink.net	Statistics
Chris Leach	chrisleach@yahoo.com	General
Chris Long	clong@padres.com	Statistics
John Matthew IV	john.matthew@rogers.com	Apostrophes
Nicholas Miceli	nsmiceli@yahoo.com	Statistics
John Stryker	john.stryker@gmail.com	General
Tom Thress	TomThress@aol.com	Statistics (regression)
Joel Tscherne	Joel@tscherne.org	General
Dick Unruh	runruhjr@iw.net	Proofreading
Steve Wang	scwang@fas.harvard.edu	Statistics

Few Saves, Little Glory – A New Way of Measuring Middle Relievers

Gary Gillette and Pete Palmer

Statistics specific to middle relievers currently number only one – the "hold" – which is unsatisfactory in several ways. Here, the authors come up with a new statistic, the "quality relief," in an effort to make the middle reliever's performance more measureable.

Ever since the current strategy of limiting closers to pitching primarily for one inning – and even then, only with a lead – was invented by Tony La Russa in the late 1980s, the save has become a much less interesting statistic. Moreover, the drastic limitation of the closer's role in the late 1980s accelerated and enhanced the concurrent trend of specialization in middle relief, giving the bullpen defined roles for two pitchers: the closer and a setup reliever.

The bloating of pitching staffs in the 1990s, plus the ever-increasing reliance on shuttling players from Triple-A and Double-A to the majors, inevitably led to further definition of relief roles. Soon many teams had two setup pitchers, one of whom was typically reserved for 8th inning duty and the other who was typically employed in the 7th inning. If possible, managers also preferred to have one righty and one lefty available for setup duty.

And it got worse – or better, if you were a marginal minor league pitcher who never would have carved out a big-league career if pitching staffs were kept to 10 roster slots. The old mandate of having at least one left-hander in the bullpen soon became a craving for dual southpaws. One lefty, of course, was generally a designated "hit man" whose role it was to face opposing teams' best left-handed hitters when the game was on the line. These situational lefties would often average far less than an inning per appearance, leading to the creation of the acronym LOOGY, for the cumbersome phrase "left-handed one-out guy."

Despite the greatly increased emphasis on – which is not to say importance of – relief pitching, only two new stats have entered the baseball vernacular as a way of measuring the effectiveness of relief pitchers. The first, the "Tough Save," is a meritorious stat that helps sort out the wheat from the chaff in an era when saves are so easy to come by. The second, the "Hold", is a much less worthy stat that attempts to measure the effectiveness of middle relievers.

When talking about the battalions of middle relievers, the hold is the most commonly cited statistic today aside from wins and ERA; the stat is about 15 years old now. Unfortunately, as a measurement of middle relief pitching, the hold has major problems. The definition of a hold is to enter the game in a save situation and leave with your team still ahead. This is what SportsTicker uses in its daily box scores. STATS, Inc., made a slight refinement some years after they started counting holds that required the pitcher get at least one out. However, the STATS figures are not as widely used as the Ticker numbers. (By the way, 1892 of 2011 Ticker holds qualified for STATS holds in 2004.)

The biggest problem with holds is not just how easy they are to get, but is inherent to the definition of the statistic. Holds measure middle relievers by using a metric designed for closers. If the stat starts with a save situation, its utility in evaluating middle relievers is severely compromised.

But there are also other problems. Directly related to the save situation part of the definition is the fact that a pitcher cannot get a hold if he enters the game with the score tied or his team behind. Pitchers are eligible to earn wins or losses in those situations, of course, but that leaves the majority of relief appearances without an evaluative stat. Another problem is that runners on base when a pitcher enters the game are not differentiated, nor are runners left on base by a middle reliever considered in any way.

So we decided to come up with something new and better. We started out by looking at the top pitching performances for 2004 in terms of Player Win Averages, invented by the Mills brothers in 1969. What they did was calculate the probability of winning the game based on score, inning, baserunners and outs, looking at the before and after for each event and how it raised or lowered the probability. The decimal place is eliminated, so 1000 player win points equal 1 win over average. Since each team starts at 500, gaining 1000 points in two games would be two wins, compared to the average of only one. If the win probability goes up from 50 percent to 70 percent, that is 200 points. The win probabilities were generated from a simulation, since in actual play, the sample size for most cells are too small to give a reliable number even if several seasons are surveyed.

Looking at all the pitching performances for 2004 with Player Win Averages of 200 points or better, we find 867 by starters, 339 by middle relievers, and 130 by closers. The starters had 615 wins, 18 losses, and 234 no-decisions. The closers had 68 wins and 62 saves. Of the middle relievers, however, there were only 77 holds, 57 wins, 3 losses, and 4 blown saves. The other 198 appearances got nothing.

A couple of examples show some of the problems. On April 5, 2004, Rafael Betancourt entered the eighth inning for the visiting Indians against the Twins with a 4-3 lead, one out and runners on second and third. He allowed one run to score to tie the game (getting a blown save), then pitched a scoreless 9th and 10th. The Indians lost 7-4 in the 11th.

Scott Stewart preceded Betancourt on the mound. He came in ahead 4-0 in the eighth, with the bases full and one out. He allowed a single and two doubles for 3 runs before departing and leaving the rest to Betancourt. Stewart got a hold. Betancourt had 302 player win points, Stewart minus 455. Of the 623 blown saves in 2004, 24 actually resulted in positive points for the pitcher.

The worst outing that earned a hold in 2004 belonged to Mike DeJean of the Orioles at Anaheim on May 21. He lost 531 points by entering at the start of the 8th inning with Baltimore ahead, 3-1. He left with a run in, runners on second and third, and none out. Darwin Cubillan relieved and allowed both runners to score, giving DeJean a hold and the loss. Cubillan lost 149 points for his efforts. Of the 2011 holds in 2004, 364 had negative player win points.

Interestingly, closers don't do that well when looking at PWA per game because they don't pitch enough innings. The top three relief appearances last year were all by middle relievers in extra inning games, with two of them in the same game. Joe Roa pitched a scoreless stretch from the 13th to the 17th in a 18-inning, 6-5 home loss to Oakland on August 8 for 640 points. Justin Duchscherer came in with the score tied, 2 outs and runners on first and second in the last of the 10th, pitching through the 14th with no runs allowed for Oakland, worth 636 points. Neither pitcher got a hold because the score was tied at the time. Duchscherer also pitched 5 scoreless innings (10th through 14th) on May 11 at Detroit and got the win and 640 points as the A's won 5-4 in the 15th. These ranked 7th, 8th and 9th for best performances of 2004.

The top save performance, at number 46 overall, was Jesus Colome of Tampa Bay at Arizona on June 20, worth 498 points. Colome went 1.2 innings, coming in with runners on first and third and one out in the 8th, and the Rays ahead 2-1 (which was also the final score). The worst performance of the year, at minus 965 points, was by Matt Mantei. On April 18, Arizona at San Diego, Mantei came in with 3-run lead in the last of the 9th and allowed 4 runs without getting an out. (Player-win wise, it would not matter how many outs he got as long as he allowed the 4 runs to lose the game.)

A problem with Player Win Points is that when entering in the last inning, the details of what happens do not matter, only whether the game was won or lost. A pitcher entering with a 3-run lead can strike out the side, or allow 2 runs and finish with the bases loaded, but both give the same result provided that he is not relieved. (Of course, the same effect occurs with a save.) Another problem is that one crucial at bat can make a big difference. Season totals have a standard deviation of around plus or minus 1500 points, yet an individual at bat can often be 200 points or more. There were 2000 of these at-bats in 2004, out of 188,000 at bats total.

Ultimately, Player Win points are too complicated for any relief stat that hopes to become commonly used. However, any system chosen for a new stat should correlate well with them. Taking a cue from the Quality Start stat, we propose an analogous statistic that we have called Quality Relief (QR). Here a pitcher must allow less than one run for every two innings pitched, including our allowances for inherited runners.

Inherited runner stats are widely quoted for relief pitchers, but they also have their own problems. Obviously, all inherited runners are not equal and there has been no accounting of runners on base when the pitcher leaves the game, even though those runners are inherited by the following pitcher. These factors tend to cancel out over a season but can make a big difference in a given game. Allowing few inherited runners to score is a plus for a reliever, and those who keep most inherited runners from scoring often see their sparkling stats quoted, but we are really interested in the whole package, not just one part.

The first step in allocating inherited runners among pitchers is to calculate the probability of each runner scoring. The pitcher leaving would be charged with the potential runs, while the new pitcher would be charged with the difference between the potential and actual runs scored. Thus a relief pitcher would get credit for preventing runs from scoring and also be charged with those who scored beyond what would have been expected as well as a portion of those he left on base if he was replaced.

By assigning half a run each for a runner left on 1st with no outs, on 2nd with one out, or on 3rd with two outs, we can use a simplified version of the actual scoring potentials to calculate potential runs from inherited runners. This allows us to more easily split the responsibility between pitchers and comes reasonably close to the actual averages.

Combining this split of inherited runners for the 21 possible base-out situations produces Table 1. As can be seen, if any runner scores from first or second after being inherited with two outs, the relief pitcher bears sole responsibility. Conversely, the starter bears sole responsibility for runners left on second or third with no outs.

While this is a simplification, it works pretty well and comes close to the actual averages. In fact, in 13 of the 21 possible inherited runner situations, it is the same as if the actual run-scoring potential were rounded to the nearest half-run. Over a full season, this approximation usually comes within 10 percent (or a couple of runs) of what the precise total would be for most relief pitchers.

Table 1 – Simplified Inherited Runner Scoring By Situation

	0 out	1 out	2 out
1st	0.5	0	0
2nd	1.0	0.5	0
1st and 2nd	1.5	0.5	0
3rd	1.0	1.0	0.5
1st and 3rd	1.5	1.0	0.5
2nd and 3rd	2.0	1.5	0.5
Loaded	2.5	1.5	0.5

Table 2 shows the data from 2006. The frequency column (freq) shows the number of times each situation (sit) occurred. The average column (avg) shows how many runners scored on the average from a given situation, while the approximate column (approx) shows the simplified version that can be used to more easily split the responsibility between pitchers.

For example, if a reliever came in with the bases loaded and one out, the previous pitcher would be charged with an approximate 1.5 runs (the actual figure being 1.34). The reliever could get a credit of half a run if he allowed only one run to score, but would be charged with half a run if he allowed two runs to score.

For 2004, looking at the difference between expected and actual inherited runners that scored shows Scott Eyre at the top of the list with 11 fewer runs allowed than expected. However, this is misleading, since Eyre usually pitched very briefly and simply left most of his runners for the next pitcher. He assumed 64 inherited runners, with 20 expected to score, though only 9 did. However, the key stat is that Eyre passed on 57 runners to the next pitcher, of which 11 scored. So he was really only average.

Table 2 – Actual Inherited Runner Scoring By Situation

sit	freq	runs	avg	approx
0 out				
1	11405	4423	0.39	0.50
2	3639	2205	0.61	1.00
12	2886	2876	1.00	1.50
3	503	421	0.84	1.00
13	1031	1276	1.24	1.50
23	670	951	1.42	2.00
123	781	1398	1.79	2.50
1 out				
1	13055	3509	0.27	0.00
2	6035	2515	0.42	0.50
12	5031	3377	0.67	0.50
3	2025	1303	0.64	1.00
13	2262	2006	0.89	1.00
23	1797	1976	1.10	1.50
123	1981	2657	1.34	1.50
2 outs				
1	13269	1772	0.13	0.00
2	7640	1819	0.24	0.00
12	6320	2283	0.36	0.00
3	3105	816	0.26	0.50
13	3016	1227	0.41	0.50
23	1952	938	0.48	0.50
123	2297	1679	0.73	0.50

Eyre did the same thing in 2005, leading the league in inherited runner prevention, with 78 runners, 23 of those expected to score, and only 12 actually scoring – for a surplus of 11. But he passed on 41 runners, of which 11 scored, getting back to average. The 2006 leader was a true gem. B.J. Ryan inherited 29 runners, expected 9 to score but allowed only 1 of them – and he left no runners for anyone else.

Since a Quality Start is unrelated to the score of the game and unrelated to whether the pitcher wins or loses, we propose that the middle reliever be measured the same way. We also examined further refinements where the score of the game was taken into account and a "setup" stat was awarded if the score was close: no more than 3 runs ahead or the tying run on deck (as with a save), and no more than 3 runs behind. This covered 83 percent of the total relief appearances in 2004; the number of Quality Reliefs in those games was 7770, or 70 percent, so the stats that resulted from those refinements were very similar to the overall QR stats. One stat is usually better than two, though, and since the pitcher has no control over when the manager uses him, we decided to simply count Quality Reliefs (analogous to Quality Starts), with no provision for the score of the game.

Using this new QR method, there were 9335 Quality Reliefs out of 13,418 relief appearances in 2004, or 69 percent. Of the top 2023 performances (100 Player Win points or better), all but 13 earned QRs. Most of those that did not involved allowing 1 run in less than 2 innings. For example, the 726th rank relief was on April 15, Arizona at Florida, where Matt Mantei came in with a 2 run lead, 2 outs and runners on 1st and 2nd. He was responsible for both inherited runners based on the approximate method. Mantei allowed a double that cashed one run, which was charged to him via the inherited rule, then finished up the 9th, going 1.1 innings and allowing only that run.

Of the 350 worst performances in 2004 (minus 400 points or worse), only 4 got QRs due to allowing 2 or more unearned runs. In the case of an error, the Player Win Average method charges the batter and the pitcher as if the batter made an out, but it does not eliminate any further responsibility after the third out would have been made. Since we are using earned runs only, it is possible with a three run lead to allow 4 unearned runs in the 9th after an error and effectively get off the hook, as Rocky Biddle did on June 1, although he certainly did not do his job. (Biddle did get the loss, however.) The relief pitcher has some responsibility for overcoming errors by his defense, but for now, we have not included unearned runs in our methodology.

The leaders for 2004 are shown below, based on percentage of Quality Reliefs (QR) with a 40-game minimum. CLS are close games and CQR are Quality Reliefs in those games; as can be seen, the figures in close games are pretty much the same as the overall ones.

2004	tm	g	ip	qr	pct	close	cqr	gf	sv	pwa
Joe Nathan	MIN	73	72	66	.904	72	65	63	44	4569
Mariano Rivera	NYA	74	79	65	.878	74	65	69	53	3834
Billy Wagner	PHI	46	48	40	.870	45	40	39	21	2053
Francisco Rodriguez	ANA	69	84	59	.855	68	59	29	12	2311
Trevor Hoffman	SDN	55	55	47	.855	55	47	51	41	1913
Brad Lidge	HOU	80	95	68	.850	79	67	44	29	4750
Scott Linebrink	SDN	73	84	62	.849	71	60	7	0	1907
Armando Benitez	FLO	64	70	54	.844	64	54	59	47	3739
Steve Kline	SLN	67	50	56	.836	60	50	22	3	218
Shingo Takatsu	CHA	59	62	49	.831	54	44	45	19	1770
Mike Gonzalez	PIT	47	43	39	.830	42	34	12	1	217
B.J. Ryan	BAL	76	87	63	.829	74	61	19	3	2632
Eric Gagne	LAN	70	82	58	.829	69	57	59	45	4647
Kiko Calero	SLN	41	45	34	.829	37	31	4	2	1223
Ryan Madson	PHI	52	77	43	.827	51	43	14	1	1407
Tom Gordon	NYA	80	90	66	.825	80	66	15	4	3474
Danny Kolb	MIL	64	57	52	.813	60	48	48	39	757
Keith Foulke	BOS	72	83	58	.806	72	58	61	32	1507
Todd Jones	CIN	51	57	41	.804	49	40	10	1	1365
Chad Cordero	MON	69	83	55	.797	65	51	40	14	2617
Jason Isringhausen	SLN	74	75	59	.797	72	58	66	47	2644
John Smoltz	ATL	73	82	58	.795	73	58	61	44	4245
Mike Remlinger	CHN	48	37	38	.792	46	37	6	2	329
LaTroy Hawkins	CHN	77	82	61	.792	74	59	50	25	467
Francisco Cordero	TEX	67	72	53	.791	66	52	63	49	3235

Of the top 25 pitchers in QR Percentage last year, 15 were closers and 10 were middle relievers. That seems reasonable if we're trying to create a measurement that doesn't ignore middle relievers – especially when one considers that closers are usually the best pitcher in a team's bullpen.

Here are the leaders for 2005 and 2006:

2005	tm	g	ip	qr	pct	close	cqr	gf	sv	pwa
Mariano Rivera	NYA	71	78	64	.901	70	63	67	43	2605
Chad Cordero	WAS	74	74	64	.865	73	64	62	47	2694
Joe Nathan	MIN	69	70	59	.855	68	58	58	43	1881
Arthur Rhodes	CLE	47	43	40	.851	46	39	8	0	577
Kyle Farnsworth	DET	46	43	39	.848	44	37	16	6	1264
Al Reyes	SLN	65	63	55	.846	57	47	18	3	1965
Jason Isringhausen	SLN	63	59	53	.841	59	51	52	39	1558
Billy Wagner	PHI	75	78	63	.840	72	60	70	38	2419
Todd Jones	FLO	68	73	57	.838	66	55	55	40	2042
Roberto Hernandez	NYN	67	70	56	.836	64	53	20	4	840
Juan Rincon	MIN	75	77	62	.827	73	60	18	0	1566
B.J. Ryan	BAL	69	70	57	.826	67	55	61	36	1087
Derrick Turnbow	MIL	69	67	57	.826	66	54	62	39	2864
Scott Eyre	SFN	86	68	71	.826	82	68	15	0	2672
Cliff Politte	CHA	68	67	56	.824	64	52	14	1	3027
Bob Howry	CLE	79	73	65	.823	74	60	24	3	1292
Trevor Hoffman	SDN	60	58	49	.817	58	47	54	43	1734
Bob Wickman	CLE	64	62	52	.813	63	51	55	45	2300
Kiko Calero	OAK	58	56	47	.810	54	43	15	1	522
Luis Ayala	WAS	68	71	55	.809	63	50	18	1	756
Brian Fuentes	COL	78	74	63	.808	71	56	55	31	3396
Jim Mecir	FLO	52	43	42	.808	49	39	13	0	-457
Todd Williams	BAL	72	76	58	.806	58	45	12	1	563
Huston Street	OAK	67	78	54	.806	57	45	47	23	3178
Kevin Gryboski	ATL	31	21	25	.806	23	19	7	0	-249
Mike Gallo	HOU	36	20	29	.806	34	28	5	0	-241

2006	tm	g	ip	qr	pct	close	cqr	gf	sv	pwa
Jonathan Papelbon	BOS	59	68	53	.898	57	51	49	35	3989
B.J. Ryan	TOR	65	72	58	.892	63	56	57	38	3813
Joe Nathan	MIN	64	68	57	.891	62	56	61	36	4386
Cla Meredith	SDN	45	51	39	.867	44	38	11	0	2592
Trevor Hoffman	SDN	65	63	56	.862	62	54	50	46	3070
Wes Littleton	TEX	33	36	28	.848	25	20	6	1	874
Takashi Saito	LAN	72	78	61	.847	68	58	48	24	3183
Billy Wagner	NYN	70	72	59	.843	69	58	59	40	2738
Akinori Otsuka	TEX	63	60	53	.841	58	48	48	32	307
Chad Gaudin	OAK	55	64	46	.836	40	33	13	2	911
Dennys Reyes	MIN	66	51	55	.833	57	47	8	0	1128
Rafael Soriano	SEA	53	60	44	.830	49	41	14	2	1821
Fran. Rodriguez	LAA	69	73	57	.826	68	56	58	47	4130
Mariano Rivera	NYA	63	75	52	.825	63	52	59	34	2272
Joel Zumaya	DET	62	83	51	.823	61	50	12	1	2803
Joe Kennedy	OAK	39	35	32	.821	38	31	8	1	893
Duaner Sanchez	NYN	49	55	40	.816	47	39	15	0	1805
Pat Neshek	MIN	32	37	26	.813	24	20	3	0	797
Gary Majewski	WAS	46	55	37	.804	36	29	14	0	-1232
Randy Choate	ARI	30	16	24	.800	22	18	3	0	-53
David Aardsma	CHN	45	53	36	.800	28	22	9	0	674
J. Duchscherer	OAK	53	56	42	.792	51	42	17	9	2214
J.J. Putz	SEA	72	78	57	.792	69	56	57	36	3244
Brad Thompson	SLN	43	57	34	.791	25	21	16	0	-128
Mike Myers	NYA	62	31	49	.790	56	46	6	0	310
Joe Beimel	LAN	62	70	49	.790	46	37	10	2	1521

For 2005, 14 of the top 25 slots in QR Percentage are taken by middle relievers. Again, that seems reasonable. Note: three of 2004's best middle relievers – Francisco Rodriguez, Todd Jones, and B.J. Ryan – were closing games in 2005.

An obvious comparison for QR percentage is to save percentage, which is very widely quoted for closers. Save percentage stats (saves divided by saves plus blown saves) have a problem in that middle relievers have many opportunities to get a blown save, but very few to earn saves because the closer usually replaces them if they do their job right.

In 2004, the top 29 pitchers by saves (14 or more) had 919 saves and 145 blown saves, an 86 percent success rate. A total of 127 middle relievers (40 or more games and less than 14 saves) had 156 saves and 338 blown saves, for only 32 percent. The closers averaged 0.27 inherited runners per appearance, while the middle relievers averaged more than twice that (0.60).

Ace relievers used to be called "firemen," meaning they came in with runners on base and the game on the line after the starter had faltered and they were expected to put out the fire. Now they hardly enter under those circumstances and nowadays have relatively few chances to blow a save by allowing an inherited runner to score because they have so few inherited runners. Thus the number of blown saves is more meaningful for a middle reliever, and the save percentage stat should be used only for closers. The record of consecutive saves without a blown save is somewhat hollow, since coming into a game with the score tied and losing is not considered a failure, as happened to Eric Gagne twice during his long streak.

As every baseball fan knows, the high-profile closers get all the glory. But a large portion of the relief burden falls on their teammates in the bullpen, and the Quality Relief stat gives us a better picture of how setup pitchers and other middle relievers are performing when they don't earn many saves.

Notes

- Relief pitchers have a built-in advantage in earned run average. The reason is that they are more apt to face a batter with one or two outs than a starting pitcher. If you had three pitchers in the game, one who pitched until there was one out each inning, another until 2 outs, and the third to finish up, and they all pitched equally well, the first pitcher would have an ERA more than double that of the third pitcher. Most runs are scored with 2 outs, but most runners who score get on base with none out, as they have more opportunities to get around the bases. The average batter with none out will score about 16 percent of the time, compared to 12 percent for one out and 7 percent for two outs. Relievers also get a small benefit for getting credit for an inherited runner being put out, such as a double play or caught stealing, with no liability if the runner scores instead. Relievers in general have an ERA about 10 percent lower than starters (4.17 to 4.62 in 2004). About 20 percent of the difference is from the advantage just mentioned, while the remaining 80 percent is from pitching better. Relievers had an OPS of .841 compared to .875 for starters: 4% lower. (Run scoring is proportional to twice the difference is OPS.)

- Of the 4856 games started in 2004, 2299 were Quality Starts. Of the top 867 by Player Win Averages (200 or more points), all but three made it; in each case the starter allowed 4 earned runs. If a pitcher has a big lead, allowing a few runs does not count heavily against him. Using potential runs instead of actual runs, the number decreased to 2279, meaning more often than not, the relief corps helped preserve the starter's record.

- There have been many articles on the faults of the current relief strategy. Bill Felber had an excellent piece in the fourth edition of *Total Baseball* and Dave Smith gave a fine presentation at the SABR convention in 2004, both showing how the current strategy has not increased the probability of winning the game in the late innings. Jim Poserina's article in the 2004 edition of the *Baseball Research Journal* also addressed the problem well. Going back as far as 1990, Pete Palmer had an article in *The Show* magazine showing how the current strategy was ineffective.

This research was originally presented at the 2005 SABR convention in Toronto. In a different form, it also appeared on espn.com in August, 2005.

Gary Gillette, Baseballist@EarthLink.net
Pete Palmer, PetePalmr@aol.com ♦

Numbers Behaving Weirdly – the At-Bat Conundrum

Abbott Katz

It is a statistical truth that a player's ability cannot be accurately estimated from a small number of at-bats. However, what happens if you take every player with a small number of at-bats, and thus create a large sample? Here, the author finds significant differences between players with few at bats and players with many.

Belaboring the obvious – in lead paragraphs, no less - doesn't quite meet industry standards for reader enchantment, but journalistic duty impels me just the same to report the following truth, important if annoyingly evident: namely, that there is next to nothing to learn from tiny samples - of voters, of experimental subjects, of at-bats, of anything. Imagine a hitter with a dozen or so plate appearances and try to say something meaningful about his batting average; do so, and you're likely to have your lineage besmirched, or your passport impounded, or something similarly dire.

Push the matter to its reductio ad absurdum. There's apparently nothing – nothing – to be said about a hitter whose season comprises exactly one solitudinous at-bat. There is, by the lights of statistical convention, nothing whatever to talk about here, apart from that gruesome picture pasted to my passport (I've since grown a beard).

But what if you could harvest all the one-at-bat seasons windblown across decades worth of seasons and mash them together into one aggregate BA? The at-bat total now would be imposingly large - but even so, would the accumulated figure enable us to say anything about the numbers?

Inspired by that very question, your intrepid correspondent went about doing the math. There were, between the years 1959-2005, 1791[1] one-at-bat seasons; their aggregate batting average (all hits/all 1-bats) comes to… .109.

I then computed averages for all at-bat increments over that period - that is, an Aggregate Batting Average (ABA) for all batting seasons comprising 2 at-bats, 3 at-bats, etc…all the way up to Willie Wilson's Everest-like 705 in 1980. The results exhibit an almost eerie linearity, describing a slow, nearly ineluctable ascent through the cohorts. Thus, for example, the ABA for 17 at-bats is .171; for 30 at-bats, .165, for 50, .180; it surmounts the .200 threshold at the 77 AB mark, and so on. And the averages continue to slope upwards; at 300 at-bats the ABA figures to .249; at 400, .266, and at 600, .289. For reasons of space I've synopsized these results in bundles of 25 at-bat cohorts (I will be happy to email all 705 rows worth of at-bats upon request, and/or any other data conveyed here):

The data are, if nothing else, extraordinarily orderly, almost mysteriously so. The correlation between number of ABs and BA - .881.

And one mustn't think the Aggregate BAs are slapdash constructions, that for example lots of guys with 17 ABs hit .529 and lots of others hit .117, and that the resulting .171 merely grafts a makeshift mean atop a chaotic dispersion. 82% of the 17 at-batters hit less than .250, or, given the 17-at-bat denominator, had 4 or fewer hits. At 40 AB the sub-.250 figure was 82%; at 100 AB (virtually no pitchers here) the line still held at 71%. At 8 at-bats the less-than-.250 measure reads 66% - meaning that 313 of the precisely 475 8 at-batsmen had one or no hits.

Moreover, if you substitute parameters and correlate Plate Appearances (simply here, ABs+BBs) instead of ABs with batting average, on the grounds that power hitters with relatively low averages tend to walk a lot and thus artificially depress their at-bat totals to match their low averages, our new correlation reads: .877.[2]

Thus the data here speak to us with a kind of relentless monotony: batters with few at-bats do significantly less well than batters with many, the evidence obtaining again and again across over 6,000,000 at-bats.

But another demurral lurks behind the data. A great many of the low at-bat accumulators are doubtless pitchers, and because they're "only" pitchers, one needs to confront the temptation to toss them from the sample, as if they don't, or shouldn't, count. It seems to me this recommendation would contort the data. Pitchers are baseball players who happen, in the main, to be bad hitters. Must they be excluded as

[1] Data source: Sean Lahman's collection of seasonal statistics, downloadable from http://baseball1.com/statistics/. It's free, but they accept donations. The analyses were performed with Excel's Pivot Tables.

[2] I thank Phil Birnbaum for this and several other insurgent perspectives on the data cited here.

a result? And in any case, if one confines the data to post-1972 American League seasons, during which the DH contracted pitcher at-bats to the vanishing point, the correlation that emerges between ABs and BA: .855.

But readers may fail to share my thrall with these numbers, branding it a rather cheap statistical thrill indeed. They can lob back a perfectly sensible and unexciting reason to account for the data. *Of course* hitters with few at-bats have markedly lower batting averages, they'll rebut - precisely because they're bad hitters, just the sorts of people who don't get to play terribly much. Albert Pujols gets 590 at-bats; bad hitters get 17 – right?

Sounds good. But what entitles the spoilsports to say with any measure of confidence that these cameo players are bad hitters? *They only have 17 at-bats.*

And if one grants that each of these miniscule seasons tell us nothing by themselves, but rather fits just one more tile into a grand statistical pattern - so that the aggregate .171 of 17-at-bat hitters comprises one mighty, superseding gestalt instead - then we'll have to allow in turn that, even though these players are overwhelmingly lousy, we're incapable of describing any particular player in those terms. Is that what we want to say?

Nor can we treat these data in the same way as we might regard Albert Pujols' 590 turns at the plate, if were we to dice these into 35 or so bundles of 17 at-bats. In this case, a 3-for-17 Pujols "slump" could be joined to his other 34 bundles, as they all emanate from the same talent source (although even this formulation makes me nervous). The data here instead total 277 different 17-at-bat seasons, springing from different bats and different times.

That points us to another problem. If the hitters described here are in fact bad ones, I've treated myself to that finding because so many of them in the low at-bat domain have performed poorly, and because the strength of the resulting distribution is so striking and persuasive. That is, I seem to know that Jim Pisoni (he was actually on the Yankees) is a bad hitter because so many of his 17-at-bat colleagues were similarly bad, thus corroborating the pattern. But that inference makes no sense. Why should Pisoni's 3-for-17 in 1959 have anything to do with Todd Van Poppel's 3-for-17 in 2004? Can we say the two were beholden to the same ubiquitous statistical, or even social forces, 45 years apart?

And what if Pisoni's performance were the exception instead? That is, suppose the 17-at-bat aggregate BA came to .250; do we now suddenly turn around and write off Pisoni's .177 as nothing but a routine case of random fluctuation?

Now let us consider one last question about the data, a rather significant one. Couldn't one assert that the low at-bat, low-average guys are "really" better hitters than the data suggest, and that these unfortunates simply suffered their bad-luck slump early on, thus scaring their job-insecure managers into remanding them to the bench? Isn't it possible that giving these players additional, significant ABs would have righted their averages to their "actual" talent levels?

For example, to simplify the numbers: Maybe our 17-at-batters were really, in sum, .230 hitters, and not deserving of the collective .171 stigma the record inflicts upon them. That sounds most possible, indeed; perhaps these fellows had simply careened into the dark side of the standard deviation, having been waylaid by a mere, "unlucky" 17 at-bats.

But if that scenario applies, then it follows that for every .171 hitter at 17 ABs we'd find a "lucky" .290 hitter at the same 17-at-bats, who would have thus presumably "earned" more at-bats and disappeared into the higher at-bat strata. That is, we should have had as many lucky guys populating the upside of the deviation as those inhabiting the low end after 17-at bats. But if this serendipitous crew were given serious at-bats, its members could not be counted on to prevail at .290 – *they'd fall back to .230.*

Table 1 – BA for various numbers of AB in a season

AB cohort	Sum of ABs in cohort	avg
1-25	101689	.142
26-50	143295	.181
51-75	197598	.186
76-100	175815	.210
101-125	147317	.230
126-150	152186	.240
151-175	159071	.244
176-200	170006	.249
201-225	175425	.248
226-250	181354	.249
251-275	194463	.253
276-300	186698	.253
301-325	215767	.259
326-350	237263	.260
351-375	231864	.261
376-400	259617	.263
401-425	273722	.264
426-450	271858	.265
451-475	319303	.268
476-500	349952	.271
501-525	377741	.273
526-550	409654	.278
551-575	443686	.279
576-600	404178	.284
601-625	316547	.286
626-650	173659	.288
651-675	75213	.295
676-700	25330	.303
701-725	2110	.324

Yet the data don't confirm this surmise. There were about 19,000 1-to-75-at-bat seasons played out during the 1959 to 2005 span. Their aggregate batting average: .178. The aggregate 75-plus AB average during that time, also comprising close to 19,000 individual seasons: .264. Were the above analysis in point, we'd see a huge centripetal tug toward .230. We don't.

But the other explanatory alternative – that 17 at bats, considered in isolation and distanced from any wider context or pattern, should allow us to say something meaningful about a hitter – is heresy.

Now time to step back. What I am *not* proposing, that last paragraph notwithstanding, is a repeal of the laws of probability. To quote Gertrude Stein, a coin flip is a coin flip is a coin flip. But the larger question - to which we don't have an unassailable answer - is the extent to which "luck" insinuates itself into baseball games in the first place (observe Bill James' unease about this point at www.sabr.org/cmsfiles/underestimating.pdf). In spite of the sabermetric fondness for the luck parameter, we need to acknowledge that luck as we tend to understand it is a variable, and impacts various sports variously.

For example, the British columnist Robert Crampton asserts, in a different sporting connection:

> At other games (tennis, badminton, pool, etc.) the inferior player always has what they call in boxing a puncher's chance, as when Henry Cooper floored Ali in the fourth in Wembley in 1963. But squash isn't like that. Squash is like chess: the inferior player gets crushed, every single time.[3]

No room for standard deviations here. Chess is less luck-ridden than poker - far less. Is poker less luck-ridden than baseball? We don't yet know where to notch baseball on the continuum.

Moreover, is there really any shortage of human endeavors in which small samples *are* taken rather seriously? A college student typically graduates after having taken about 40 courses. Is that number too small to justify a confident conclusion about his/her GPA? More to the point, a National Football League season comprises 16 regular-season games. 16 games is less than 17 at-bats; and are we prepared as a result to declare that virtually no NFL season conveys any qualitative meaning? Now think about soccer. Games in the British Premiership league - the entity that umbrellas the Manchester Uniteds and the Chelseas, etc. – average about 2.14 goals between any two competing teams. With scoring activity that infinitesimal one might assume "luck" could grab hold of lots of games, infiltrating a fluke goal here and there, and thus serving to squeeze lots of teams towards the .500 mean. But check the standings; disparities in the Premiership are gaping. How much luck here?

Of course I'm winding down with more questions than answers. But perhaps we need a bit of a re-think about small numbers and small samples. Maybe sometimes they matter a bit more than we tend to assume - even in baseball. And one more question, if you will: can I have my passport back?

Abbott Katz, akatz@hotmail.com ♦

[3] Robert Crampton, *The Times [UK] Magazine*, January 14, 2007, p. 74.

Get Your Own Copy

If you're not a member of the Statistical Analysis Committee, you're probably reading a friend's copy of this issue of BTN, or perhaps you paid for a copy through the SABR office.

If that's the case, you might want to consider joining the Committee, which will get you an automatic subscription to BTN. There are no extra charges (besides the regular SABR membership fee) or obligations – just an interest in the statistical analysis of baseball.

The easiest way to join the committee is to visit http://members.sabr.org, click on "my SABR," then "committees and regionals," then "add new" committee. Add the Statistical Analysis Committee, and you're done. You will be informed when new issues are available for downloading from the internet.

If you would like more information, send an e-mail (preferably with your snail mail address for our records) to Neal Traven, at beisbol@alumni.pitt.edu. If you don't have internet access, we will send you BTN by mail; write to Neal at 4317 Dayton Ave. N. #201, Seattle, WA, 98103-7154.

Submissions

Phil Birnbaum, Editor

Submissions to *By the Numbers* are, of course, encouraged. Articles should be concise (though not necessarily short), and pertain to statistical analysis of baseball. Letters to the Editor, original research, opinions, summaries of existing research, criticism, and reviews of other work are all welcome.

Articles should be submitted in electronic form, either by e-mail or on CD. I can read most word processor formats. If you send charts, please send them in word processor form rather than in spreadsheet. Unless you specify otherwise, I may send your work to others for comment (i.e., informal peer review).

If your submission discusses a previous BTN article, the author of that article may be asked to reply briefly in the same issue in which your letter or article appears.

I usually edit for spelling and grammar. If you can (and I understand it isn't always possible), try to format your article roughly the same way BTN does.

I will acknowledge all articles upon receipt, and will try, within a reasonable time, to let you know if your submission is accepted.

Send submissions to:
Phil Birnbaum
88 Westpointe Cres., Nepean, ON, Canada, K2G 5Y8
birnbaum@sympatico.ca