BASEBALL ANALYST

FEB. '87 JOURNAL OF SABERMETRICS VOL. 27

Well, you've probably been wondering a lot lately how many runs Tip O'Neill might have driven in in 1887, and the *Analyst* has taken two extra months to get here and tell you. Due to the *Abstract* crunch, there was no December, 1986 issue of this publication, so that we now owe you a whole issue as well as those six pages we owed you anyway. What we're going to do to catch up is do an issue in February--this one--an issue in March and an issue in April, with the March issue taking the place of the December one that was missed.

There are four articles in this issue:

Page 2 Tip O'Neill: Estimating his 1887 RBI
Dallas Adams

Page 4 A Look at 'Turf' Teams and 'Turf' Players: 1983-1986
Sandy Sillman

Page 8 The Implications of Leadership Research on Baseball
Charlie Pavitt

Page 12 Some Relief Pitching Statistics From 1986
Philip Meneely

I think I have enough material to scrape together the March issue, so I'll dispense with the usual grovelling and begging for articles here and just express my gratitude to Dallas, Sandy, Charlie and Philip for making their research, all of which I think is quite interesting, available to us. You'll be hearing from me again shortly.

Bill James

TIP O'NEILL: ESTIMATING HIS 1887 RBI'S

by Dallas Adams

This subject of this study was suggested by my friend Bill Rubinstein. He pointed out that the Macmillan Encyclopedia presents RBI data for the American Association (1882-1891) only for the final four years of its existence. Which means that there is no formal count of Tip O'Neill's RBI's for his wonderful 1887 season. Noting that the St. Louis Brown's, O'Neill's team, scored 1131 runs that year, Bill wondered whether O'Neill could possibly have reached 200 runs batted in; and asked if it was possible to estimate O'Neill's RBI's for that year.

My first thought was that O'Neill's statistics were remarkably similar to those compiled by Hugh Duffy for Boston (which scored 1222 runs) in 1894:

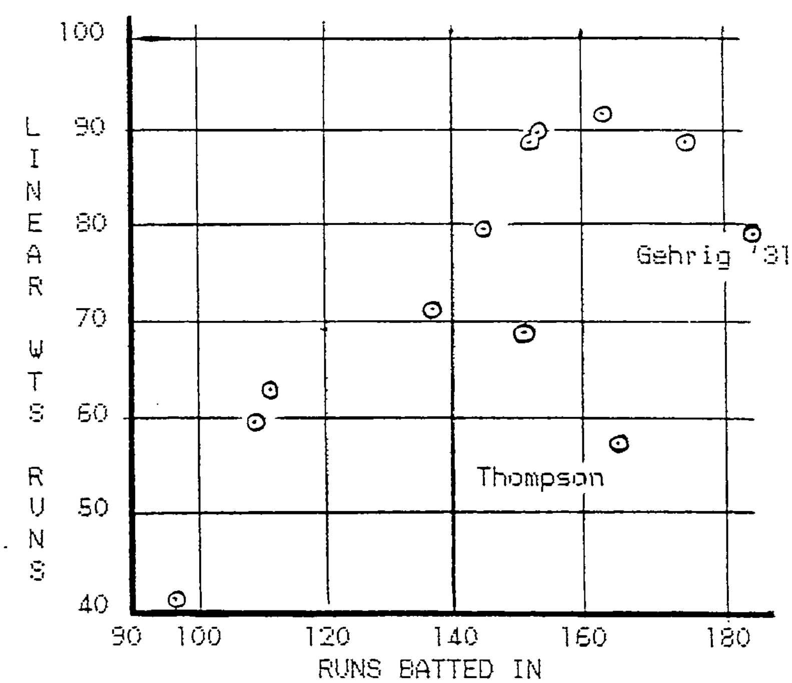
	YEAR	G	AB	R	H	2B	3B	HR	RBI	BB	K	SB	BA	SA	LNWT
O'Neill	1887	124	517	167	225	52	19	14	---	50	--	30	.435	.691	88.5
Duffy	1894	124	539	160	236	50	13	18	145	66	15	49	.438	.679	79.8

The "LNWT" data are Pete Palmer's Linear Weights Runs, as taken from The Hidden Game.

Using Bill James' Similarity Scores technique (1986 Baseball Abstract, pages 24-29) and omitting the RBI and strikeout comparisons, one finds that O'Neill and Duffy show a similarity of 959, which is extremely high. That figure would probably be reduced slightly, a few points at most, if O'Neill's RBI and strikeout data were known. In any case, the close similarity of the above two statistical lines is apparent. And that suggests that O'Neill probably had about the same number of RBI's as did Duffy; or perhaps a few more, given that he, O'Neill, also had the higher LNWT figure.

That latter point drew my attention to the possible relationship between Palmerian linear weights runs and runs batted in, especially for heavy-hitters on high-scoring teams. So, I compiled a list of players in the top-3 of league linear weights and who also played for a team scoring over 1000 runs.

PLAYER	YEAR	LNWT	RBI
Brouthers	1891	59.8	108
Duffy	1894	79.8	145
Kelley	1894	63.5	111
Thompson	1895	57.7	165
Ruth	1930	89.5	153
Gehrig	1930	88.4	174
Ruth	1931	91.5	163
Gehrig	1931	79.9	184
Ruth	1932	71.2	137
Gehrig	1932	68.8	151
Gehrig	1936	88.2	152
Williams	1950	40.9	97

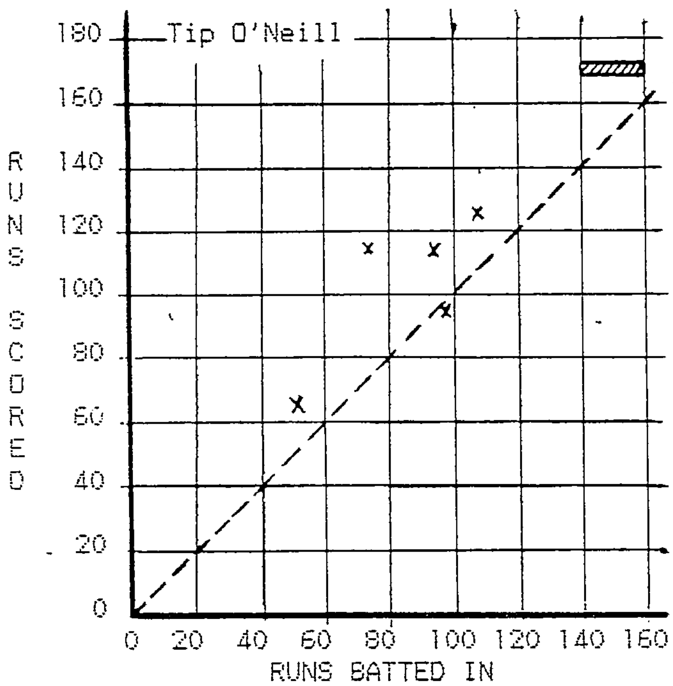


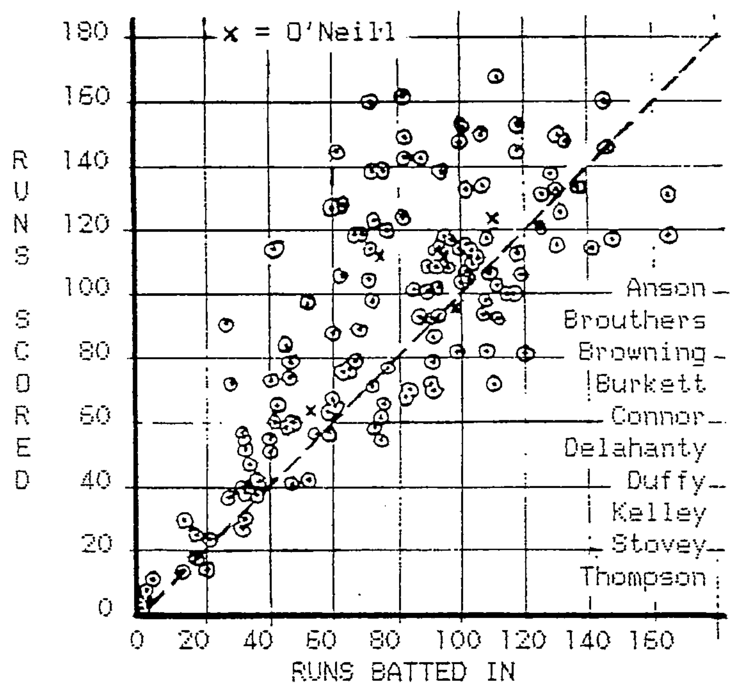
From the graph one can see that there is a general relationship between linear weights runs and runs batted in. Indeed, if the data points for Thompson and Gehrig (1931) are ignored, the correlation between the two parameters is quite strong.

Therefore, the graph suggests that O'Neill, with his linear weights value of 88.5, is most likely to have had about 160 RBI's.

Perhaps, though, O'Neill, like Gehrig and Thompson, had an unusually high (in comparison with his linear weights performance) frequency of driving in runs. No. It turns out that he was average (possibly even a trifle below average) in this regard. The graph, below left, shows O'Neill's runs scored versus runs batted in performance for the five seasons (1888-92) of his career for which we have RBI data. The diagonal line represents a 1-to-1 ratio between runs batted in and runs scored. His best RBI/R ratio occurred in 1888 when he had 98 RBI's and 96 runs scored; in the other four seasons he had more runs than RBI's. This implies that for O'Neill in 1887, a reasonable guess would be that he compiled between 140 and 160 RBI's -- the region indicated by the shaded box on the graph -- for his 167 runs scored.

The graph, below right, presents the same information for ten famous batsmen of O'Neill's era, plus O'Neill himself (his data points are denoted by "x"). As can be seen, some of O'Neill's contemporaries had seasons when they were adept at driving in more runs than they scored; mostly, however, they -- like O'Neill -- drove in fewer than they tallied themselves. Hence, O'Neill appears not to have had any special ability for driving in runs; he appears to have gotten about as many RBI's as he should have.





One final item: the runs scored by the 1887 Browns were fairly evenly distributed among the regular players. Six men had 100 or more, and Welch had 98. This suggests, weakly perhaps, that the runs batted in were also rather evenly distributed.

All of the above considerations lead me to feel that O'Neill probably had about 160 RBI's in 1887.

[io]% cat turf.doc

A LOOK AT 'TURF' TEAMS AND 'TURF' PLAYERS: 1983-1986

Sandy Sillman

Bill James,in the BASEBALL ABSTRACT,has had a running commentary on the question of which teams have played better on artificial turf. He has shown that some teams,such as the '83 Phillies, have a great record on turf even though they don't look at all like a turf team...while other 'classic' turf teams,such as the '85 Cardinals,do not in fact play any better on turf than on grass. It seems to me that we should be able to find out once and for all what types of teams really are turf teams,and what teams are grass teams. I have put together some information on the subject,and although I haven't found a 'once and for all' answer,the stuff I did find is worth passing on.

First off: Bill James (in the '86 BA) put together a collection of 'turf teams' -- teams with high batting averages,speed,and fielding -- in order to see how well these teams really did on turf. As he suggested,it should be possible to stand this study on its head -- to select a group of teams that really DID play well on turf and see what their team characteristics are. If you compare teams with the best turf record to the teams with better grass records,you should get a clear picture of "turf" and "grass" teams.

Here's what I did: I looked at team records for the years 1983 to 1986 and compared their record on grass with their record on turf. I did this for AWAY GAMES ONLY. (This avoids the home-field bias that you find,for example,in the record of a Houston team that plays half its games in the Astrodome.) By way of summary,I formed a 'turf index'* that expresses the difference (in games) between a team's turf record and grass record.

Here are the best turf teams over the last four years:

TEAM	ROAD GRASS	ROAD TURF	TURF INDEX*
Atlanta 83	17-23	25-18	+6
Philadelphia 83	16-26	24-15	+9
Pittsburgh 83	19-23	24-15	+6
Oakland 83	18-37	14-12	+6
$ Atlanta 84	18-23	26-16	+7
$ Philadelphia 85	14-28	20-19	+7
$ Cincinnati 85	22-25	20-13	+7
Stl Louis 86	15-22	17-12	+6 (thru Sept 10)
Montreal 86	16-21	18-13	+5 (thru Sept 10)
Oakland 86	12-33	12-10	+6 (thru Sept 10)

---

* Here is a sample 'turf index' calculation: Houston '85 had an overall road record of 39-42,which 'should' break down into 23-25 grass,16-17 turf. In fact the Houston record was 28-20 grass and 11-22 turf -- a difference of 10 games from the 'neutral' turf/grass split. So the Houston turf index is -10.

Here are the best grass (or worst turf) teams:

TEAM	ROAD GRASS	ROAD TURF	TURF INDEX
Cubs 83	15-18	13-35	-8
NY Mets 83	13-19	14-34	-6
San Diego 83	20-21	14-28	-6
$ Boston 84	37-20	8-16	-7
$ Philadelphia 84	25-17	17-22	-7
$ Houston 85	28-20	11-22	-10
$ NY Mets 85	22-11	25-23	-6
$ San Fran' 85	18-26	11-31	-6

You will notice that Philadelphia appears in successive years among BOTH the best turf teams and among the worst! This sudden shift in grass/turf performance from year to year is not unusual. The Mets,'Stros,Cardinals,Expos and Red sox all show similar tendencies. In fact,there are only two teams that show year-to-year consistency through this 4-year period: the Braves (good on turf) and the Giants (on grass). More on this later.

Now,most of the grass/turf records on the above list are not statistically significant,and could easily be flukes. The Oakland records are especially suspect: Oakland's turf opposition in '83 and '86 included two very weak teams. Still,if there is any convincing difference between a 'turf team' and a 'grass team', we ought to find it by comparing the teams on this list.

I would love to get my hands on team records for '83 and '86 so I could do a complete comparison of the performance of all teams on the list. Unfortunately I have detailed records only for the '84 and '85 teams (marked '$' in the table). Here are their composite records:

	OFFENSE							OPPONENTS' OFFENSE						
	runs	hits	2b	3b	HR	DP	SB	runs	hits	2b	3b	HR	DP	SB
GRASS	690	1440	243	41	140	133	110	670	1390	255	33	120	135	130
TURF	660	1350	240	35	120	120	140	665	1365	220	42	122	128	158

A virtual washout!True,the GRASS teams hit more home runs than did the TURF teams...but the amount is not very significant. (The ST.LOUis '86 team will strengthen the turf home run effect,but then again,the St LOU 85 team almost made the GRASS list.) Except for home runs,there is really nothing that distinguishes GRASS teams from TURF teams. The turf teams did steal more bases,but they also had a worse base-stealing percentage than did the GRASS teams; and their defence against steals was worse.

There is only one possible difference between GRASS and TURF teams in this table; and that lies in the opponents' doubles. Both GRASS and TURF teams hit an identical number of doubles, but the GRASS teams gave the opposition more doubles. An indicator of weak outfield defense,maybe? Aside from this, if you can find a pattern to the performance of these 'GRASS teams' as opposed to the 'TURF teams', you're doing better than I.

Incidentally,the cumulative four-year records (83-86) on road grass androad turf show very little persistant patterns.The best 4-year grass and turf habits are:

Atlanta	65-87 grass	85-78 turf	(index +16)
San Fran	68-85 grass	58-104 turf	(index -14)

The N.L. teams that play on turf at home show a very slight advantage in their road turf records. This advantage amounts to only 26 games out of a thousand played,though. A table with the 'turf index' for all teams is included at the end of this.

TURF PLAYERS: Next,I want to turn my attention to the question of which type of PLAYERS do better on grass or on turf. Here there are some real surprises. Players,like teams,have their grass/turf records change greatly from year to year: Keith Hernandez leapfrogged from a .230 batting average on turf in 1984 to .350 in 1985. There is one player who consistently hits much worse on turf than on grass: Kirk Gibson. He regularly has lost over 100 points on his batting average when playing on turf. I haven't any idea why.

But turf performance is not really about batting average. Using the 'park adjustment factors' for offensive performance in different parks (1985 Baseball Abstract) you can find out just what effect turf has on batting average. The 10 turf parks have a combined park adjustment factor for batting average of 1.013; this means that a .250 hitter on grass transported to Artificial Turf Hitters' Heaven,miraculously blossoms into...a .2503 hitter. Turf parks don't really increase batting averages. They DO increase extra-base hits -- doubles and triples increase by 7%. Theoretically the extra doubles could work to the advantage of a club that lacks power,making inside-the-park runs easier to get.

But what TYPE of hitter gets the extra base hits? Is it the speed artist who gets the most benefit out of artificial turf,or is it the slow-footed slugger? After all,even Dave Kingman has to run the bases sometimes.

I selected a group of speedsters -- classic turf players -- and a group of sluggers to look at in depth for grass/turf performance. In order to avoid home-park biases,I selected only players who play on artificial turf at home,and whenever possible I selected speedsters and sluggers from the same ball teams. The players I looked at,all for 1985,were: Tommy Herr,Vince Coleman, Willie McGee,Juan Samuel,Lonnie Smith,and Tim Raines (speedsters), and Jack Clark, Dave Parker,Mike Schmitt,George Brett and George Bell (sluggers). I then summed up the total number of doubles and triples hit by these people on grass and on turf,and calculated a 'partial slugging percentage' -- the contribution to slugging percentage made by doubles and triples only,found by taking:

2 x (# doubles) + 3 x (# triples)

and dividing by times at bat. This gives a measure of 'inside-the-park' slugging.

The results were exactly the opposite of what you would expect:

	PARTIAL SLUGGING PERCENTAGES GRASS	TURF
SPEEDSTERS	.140	.155
SLUGGERS	.116	.155

Turf was actually MORE helpful to the sluggers than to the speedsters. True,the speedsters hit more doubles and triples than the sluggers did overall...but the change from grass to turf helped the sluggers more.

Now,this speedster/slugger difference includes just a few people over one season. The results may not hold up when a larger group is included. But I will be very surprised if it could be shown that turf benefits speedsters MORE than sluggers. Turf is not a factor like short fences which benefits only a select group of people who can swing for those fences. Turf benefits every hitter,regardless of speed or style.

CONCLUSION? I wouldn't want to argue that slow,power-hitting teams are the ones that perform BEST on turf; although the speedster/slugger effect may explain why some slow teams (like Philadelphia in '83) excel on turf. Two turf factors cancel each other out here: if the power hitters benefit most from artificial turf; the teams without power benefit from the overall increase in inside-the-park offense. Also I have not considered at all the effect of turf on pitchers -- a likely big effect. But the bottom line for hitters looks like a washout...as do the team records.

Personally I'm not convinced that there are ANY real park effects -- not even an effect due to park size. Often when people cite park effects,it serves a useful purpose -- it counteracts the 'devil theory of park effects' described by Bill James (1985 Abstract). That is: a baseball man might say: 'the Red Sox need to have a good hitting team because Fenway favors hitting teams.' I would reach the same conclusion by a different route: 'the REd Sox had better have a good hitting record,because Fenway inflates their batting averages and if they hit poorly even in Fenway,then their hitting is REALLY bad.'

The Implications of Leadership Research on Baseball:
1. Past Research
Charles Pavitt

I'd like to offer my two cents on an issue that has been coming up repeatedly in the _Abstract_ over the past couple of years - the match between manager and team. Bill James made a number of claims in the 1985 _Abstract_ (pp.213-215) that I wish to comment on in this and a subsequent essay. One is that "managers _do_ make a difference" (p. 215); more specifically, that sometimes it is in a team's best interests to replace one manager with another. It turns out that there has been some research published in the social scientific journals on the impact of changing managers on team performance. This work was performed to test some theories about the impact of management change on organizational success, with baseball used as an example because different teams' performances are clearly quantifiable (via won-loss percentage) and comparable to one another (via league standings). They also provide a study of the manner in which the quality of research can be improved by constructive criticism.

It all started with an article by Grusky (1963) proposing two possible theories for the relationship between managerial change and team performance. The "common-sense" theory implies one-way influence from manager to performance; poor managing means poor performance, and replacing the poor manager with another at least provides the potential for improvement. The "vicious cycle" theory implies two-way influence between manager and performance; frequent firings cause team disruption which in and of itself leads to poor performance and more firings. In his analysis, Grusky found a correlation of -.4 between the number of managerial changes made by teams and the teams' average league standing between 1921 and 1941. A second analysis for 1951 to 1958 revealed a correlation of -.6. In other words, Grusky verified the not-particularly-newsworthy fact that poor teams change managers more than good teams. Grusky saw this finding as support for the "vicious cycle" theory.

Gamson and Scotch (1964) replied with the claim that Grusky had ignored a third theoretical possibility. The "scapegoating" theory implies _no_ influence; managers have little effect on performance, which is chiefly the result of organizational strength as a whole. Managers are fired in order to provide someone other than themselves for the owner or general manager to blame for their team's failures. Gamson and Scotch claimed that the negative correlation between team performance and number of managerial changes _accross_ seasons verified by Grusky could support either the "vicious cycle" or "scapegoating" theories. _Within_ seasons, however, firings should lead to better performance if the "common sense" theory is right (assuming that poor managers are generally replaced by better ones), worse performance if the "vicious cycle" theory is correct, and no change in performance if the "scapegoating" theory is correct. Gamson and Scotch found that, of 22 midseason managerial changes between 1954 and 1961, 13 teams proceeded to do better and 9 to do worse during the rest of the season. Despite the plurality in favor of the "common-sense" theory, they claimed that there was no reliable difference between 13 and 9 (without running statistical tests to be sure) and that the "scapegoating" theory was supported.

Grusky's rebuttal (1964) was quite interesting. He claimed that the "vicious cycle" theory further implies that managerial replacements from within the organization will be less disruptive and, as a consequence, more successful than replacements from outside of the organization. The "scapegoating" theory makes no comparable claim. Grusky found 23 midseason managerial changes from 1954 to 1961, with 14 improving and 9 worsening performance (and claiming that Gamson and Scotch must have missed one of the former). The team improved in seven of nine cases (78%) where replacements were from within the organization against 7 of 14 (50%) from the outside, supporting the vicious cycle. Grusky went on to claim that within-season comparisons were unfair anyway, as a manager can get fired with his team at 2-14 (as with the 1959 Indians) so that the subsequent manager is bound to do better. A fairer test would compare the record of the previous year for the departed manager with the record of the replacement. A reanalysis (with only 19 firings) resulted in the following:

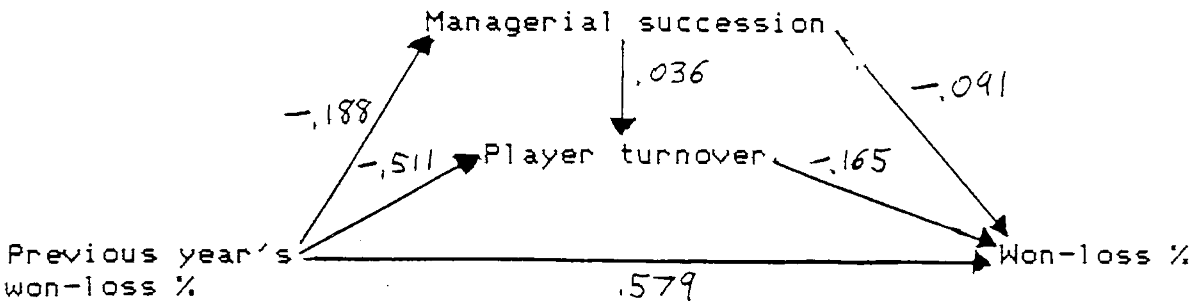
	New Manager's Performance		
Type of Change	Better	Same	Worse
Within organization	7	1	0
Outside organization	2	1	8

Within-organizational changes led to improvement in most cases, while outside-organizational changes led to worse performance, in strong support of the "vicious cycle" theory.

Thus ended this particular debate; but the issue resurfaced fifteen years later, with Allen, Panian, and Letz (1979) using a technique called "path analysis" to test a model of managerial succession. To perform path analysis, one first proposes a model of the causal relationship between a number of different variables, including both direct and indirect relationships. All of these relationships must be one-way, or a more sophisticated method must be used. Allen et al. hypothesized that three variables influence a team's won-loss percentage; amount of player turnover directly, frequency of managerial succession both directly and indirectly through its effect on player turnover, and the previous season's won-loss percentage, both directly and indirectly through its effect on both frequency of managerial succession and player turnover. Second, one calculates the correlations among the variables. Based on data from 1920 to 1973, Allen et al. found the following, with player turnover based on the 8 position players and 5 busiest pitchers as measured by innings pitched:

	won-loss percentage	managerial succession	player turnover
previous year's won-loss %	.681	-.188	-.517
won-loss %		-.221	-.476
managerial succession			.132

Third, one calculates partial correlations between the variables in the model. The resulting path model looked like this:



Partial correlations show the direct relationship between the variables in the model. The difference between, for example, the .681 correlation between previous and current year won-loss and the .579 partial correlation between the same variables is that the .681 also includes the indirect effects of the former on the latter via the paths through managerial succession and player turnover. Anyway, all of these partial correlations except for the effect of managerial succesion on player turnover were statistically significant; but given the huge sample size (934 team-seasons), many others are small. Most relevant to our interests is the very small, if negative, effect for frequency of managerial succession on won-loss percentage. Also of interest is the relatively large effect of previous won-loss percentage on player turnover.

Ignored in all of this were the theoretical arguments of Grusky and Gamson-Scotch from fifteen years earlier. Turning their attention in that direction, Allen et al. calculated the expected deviations from .500 of teams given different "types" of managerial change. The results were:

No managerial change	.004
Inside organization, between seasons	.017
Inside organization, during season	-.023
Outside organization, between seasons	.006
Outside organization, during season	-.036
Multiple successions	-.037

I would interpret these findings as supportive of the disruption implications of the "vicious cycle" theory. Teams are affected negatively by changes during seasons, positively by changes between seasons. Replacements not from the organization are less successful than replacements from inside the organization. Change in and of itself is slightly disruptive. Grusky appears correct.

Beyond these findings, there are two implications of this work that I wish to point out. First, it is extremely important for the advancement of baseball research as a field of study that we continue to provide constructive criticisms of one another's work. Without Gamson and Scotch, Grusky would probably have been satisfied with his early, trite work and not motivated to perform his later, more interesting analyses. Second, we need to continue searching for baseball research studies published in journals of the various sciences. Thorn and Palmer made a great start on this in The Hidden Game of Baseball, but as my review shows, there is more. For example, there is a fairly large literature on the economics of baseball, including a book (Baseball Economics and Public Policy by Jesse William Markham). I have unfortunately been unsuccessful in my attempts to persuade some economists I know to write a review of this stuff for all of us thus far.

References

Allen, M. P., Panian, S.K., & Letz, R. E. (1979). Managerial succession and organizational performance: A recalcitrant problem revisited. Administrative Science Quarterly, 24, 167-180.

Gamson, W. A., & Scotch, N. A. (1964). Scapegoating in baseball. American Journal of Sociology, 70, 69-72.

Grusky, O. (1963). Managerial succession and organizational effectiveness. American Journal of Sociology, 69, 21-31.

Grusky, O. (1964). Reply to Gamson and Scotch. American Journal of Sociology, 70, 72-76.

A good book on path analysis, mostly written in English rather than mathematics, is Causal Modeling by Herbert B. Asher, published by Sage.

SOME RELIEF PITCHING STATISTICS FROM 1986

Philip Meneely
3836 NE 90th St
Seattle, WA 98115

The statistics used to evaluate relief pitchers date to a time when the relief pitcher was not a significant part of a team's staff. In 1986, however, the bullpen pitched about 30% of the innings in each league--30.4% in the AL and 31.2% in the NL, ranging from 25% for Boston to 36% for San Diego. Clearly, there is a need for some new stats, or for some revisions of the old ones. I have been working on this for the past few years, and while I have not come to any one grand stat that tells you how much better Dave Righetti was than Ron Davis, or how many games more the Dodgers would have won if they had the Mets bullpen, I do have some ideas that I present for your criticisms and suggestions.

There are two basic statistics used to evaluate relief pitchers: ERA and saves. I'll consider some ways to improve each one, or to replace with a better stat.

ERA/TRA: We know that the job for any pitcher is to prevent runs from scoring, so the principal stat has to be the total run average (TRA=9*runs allowed/innings pitched). This is preferable to the ERA simply because it makes no difference to the outcome of the game if a decisive run was earned or unearned. It also has the advantage of bypassing official scorers' opinions and the rules about inherited runners aboard on an error. The TRA for each major league team is given in Table 1. Part A gives the NL, part B the AL. The table is broken down by starters (ST) and relievers (REL), and the TRA has been adjusted for the home park (using Bill James most recent park adjustment figures). I also include some other stats I'll discuss below, and the league averages for comparisons.

I'd like to draw several comparisons from the table. First, we can compare each team's bullpen or starters to the league average, or to any other team. The teams with the best rotations: Houston and the Mets; and Boston and California in the AL. We can keep track in the future to see if the best rotations are usually the pennant winners. Boston's rotation is remarkable when adjusted for Fenway--about 3.95 runs per game. We can also look at the worst rotations: the Padres, Cubs, Braves, Phillies and Reds; and the Mariners and Rangers, with the Brewers close behind. The teams with the best bullpens: the Giants and Cardinals; and the Royals and White Sox. A bit of surprise here; I would not have included the Giants and White Sox when thinking of good bullpens. The worst bullpens are more predictable: the Cubs (how good can their bullpen be if they have to trade two of them to get Ron Davis?) and Dodgers; and the Twins and Indians were far worse than any other AL team.

Another useful approach is to compare a team's starters to their bullpen. Which teams were hurt (or helped) most by their bullpens? The column labeled STAFF DIFFER shows the difference between the starters' TRA and the relievers' TRA. The average bullpen in each league is .3 runs per game better than the starters. We'll look first at teams helped most by their bullpen--those with the largest STAFF DIFFER. In the AL, the teams with the best bullpens show up, the Royals and the White Sox, but more interesting were the Brewers and Rangers. Neither had a good rotation, but each had a good bullpen. Of course, these were the two teams that committed themselves to young pitchers; the relievers in each case outperformed the starters. In the NL, the most interesting case is Atlanta, which actually had a fairly good bullpen, but bad starters. In the other two NL teams with large staff differences, the Padres and Giants, the difference is due to the weakness of the rotation (Padres) or the strength of the bullpen (Giants). On the other side, we can see in both leagues that the teams hurt most by their bullpens (the largest negative STAFF DIFFER) were the teams with the worst bullpens in each league--Minnesota, Cleveland, and Los Angeles--and the ones with the best starters--Boston, Houston, and the Mets. In fact, Boston's bullpen was quite bad, which became more evident in the Series.

But do the managers use their bullpens appropriately? We can ask that by comparing the %RIP column (percent relief innings pitched) with the staff differential--teams with the highest staff differential should have the highest %RIP. Likewise, the teams with the lowest staff differential (the largest negative number) should have the lowest %RIP. And, by and large, they do. With one exception, the Braves. Hello, Chuck Tanner. The Royals and Tigers were a little slow on the hook too, but at least their starters were better than average. The only manager we can question is Tanner, who had a good bullpen, and didn't use it.

The remaining four columns in Table 1 show the on-base percentage and batting average allowed by the starters and relievers for each team. (A brief note on how these were calculated. BA is hits/at bats. We don't know at bats against a pitcher, but we do know IP (which is a third of the outs recorded) and hits allowed. That is close to, but not identical to at bats. In fact, in each league, in both 1985 and 1986, the league ratio of at bats to 3*IP+hits is .96. I got this by using the league offensive stats. So, the denominator of BA for pitchers is .96*(3*IP+hits). A similar estimate was used for on-base percentage. Again, in both years and in both leagues, the ratio was .96. This doesn't mean that it's the same for all teams, so these are estimates. But they're close.)

Back to Table 1. The job of a pitcher is to prevent runs from scoring, which is what the TRA shows. We also know that the components of the James' run created stat for offensive players include times on base and total bases in the numerator. These are similar to OB% and SLG%, of course. And since any offensive stat can be used as a pitching stat, we should be able to come with numbers comparable to TRA by using OB% and SLG%. If we had the data, which we don't. It is difficult to get total bases allowed by a pitcher, and

I don't have it. So this is part of what we want, the OB%. I threw in BA as a bonus, because I still think in terms of a batting average of .280 or .320 and not an on-base percentage of .340 or .380.

Why do we want to bother with this if TRA does the job? Well, TRA misses on one thing and that is the runners the relievers allow to score that are not charged to them. You are familiar with the case--the starter walks the first batter in the inning, the next guy gets an infield hit, and the manager pulls the starter. The reliever's first pitch is hit 457 feet, the game is lost, and the starter is charged with the worst of it. Not fair, you say. And I agree. But it is not obvious how to change that. We could use OB% and SLG% to help estimate these runs, but it would be nice to use the runs actually scored in the game. My approach this past year was to charge fractional runs--every run was divided into fourths, each fourth reflecting one base. Thus, runners allowed to score were charged to the pitcher who allowed them to advance. In our example, the starter would be charged with .75 runs instead of 2, and the reliever with 2.25 instead of 1. You may object on aesthetic grounds, there is not such thing as .75 runs. But you didn't object when I wrote about a TRA of 3.95, so it isn't consistent for you to object now. You may object on other grounds, that it requires too much bookkeeping or that each base is not equally important in determining a run. The second may be valid, although my response is the time-honored reply to the Vince Colemans and Miguel Dilones and John (the Jet) Jeters who keep coming up--you can't get to second base without crossing over first.

I kept track of fractional runs for the Twins and Mariners pitchers in 1986. I won't show you the data, I don't want to look at data on two bad pitching staffs any more than you do. The interesting point was that fractional runs had little effect on either the starters or relievers, about .2 on the TRA. The Twins bullpen was charged with 24.5 fractional runs, the Mariners with 26.75. The biggest effect was on Matt Young, who was charged with 9.25 additional runs for people he allowed to score. That's quite a few for a reliever; it increased his TRA by more than a full run. No starter was really affected much, the most being 10 runs subtracted from Mike Morgan (which improved his TRA by .4) and 6.75 subtracted from Frank Viola (which improved his TRA by only .25). I don't know if the relatively small effect seen is due to the nature of these staffs or managers; maybe they preferred to stick with the starter rather than go to a questionable bullpen. Maybe a team with a real stopper would be more interesting. I also don't know how big of an effect was expected--maybe Young's additional 9.25 runs is equivalent to 61 homers, but I doubt it. If anyone has ideas or data, I'm interested.

The reason I kept fractional runs was that I was looking for a more complete way to measure relief performance, a way to include the runners inherited by a reliever. I think these stats work toward a solution, particularly if we could supplement them with a

SLG%. Other partial "solutions" are always being put forth. These include strikeouts, percent of first batters retired, winning percentage and so on. These are not solutions, because none of these things has been shown to affect the basic job a reliever has to do--get people out without allowing runs to score. And that brings up the most famous non-solution of all, the save.

SAVES. The other basic statistic used to evaluate relief pitchers is the save. It has some fundamental flaws. First, it is completely arbitrary; the definition has changed at least twice in the past twenty-five years. Second, not all pitchers have the same number of opportunities. This may be because they play for a poor team or because the manager doesn't use them when the game is close. Third, all saves have equal weight, whether the reliever held a two run lead for a third of an inning or a one run lead for three innings. Of these, the easiest to remedy is the second, and the hardest the third.

The way to adjust for pitchers not have the same number of opportunities is to use a save percentage. Table 2 presents one for the bullpens of the National and American League. How do we define a save opportunity? My definition is simple--either the bullpen gets the save, or it blows the lead. The first column is saves, the second is squanders. A squander is simply losing the lead, or, in the event of a tie, losing the tie. The save percentage (SV%) is then 100* SV/(SV+SQ). We see immediately that the best bullpens by this measure were not Montreal and the Yankees, which led their league in saves, but Houston and Boston. Both Montreal and the Yankees benefitted by the larger number of opportunities. In fact, Montreal was below the league average in SV%. So this is a simple way to adjust for the differences in opportunities. I haven't applied to to individuals here, but clearly it could be. If saves are to be used at all.

And that leads us to another point. Saves are flawed because they are arbitrarily defined. We can accept the definition for what it is, a definition and not a statistic, and work towards a more realistic view of the same thing. For this, I kept track of how many times a bullpen lost or held the lead, the SQ and HL columns of Table 2. The advantage here is that the situation is completely unequivocal--a team either is ahead, tied or behind. A tie, for my purposes went both ways; if a team was ahead, the tie counted as a lost lead. The statistic on %HL is then 100* HL/(SQ+HL). The best bullpens here are clearly the Mets and Boston, and the worst are LA and Minnesota. It makes a certain amount of sense that the best teams would lead the league in their ability to hold leads, although Milwaukee and San Francisco were also very good at this. Of course, saves alone would not have seen this, and save percentage would have missed the Giants.

Suppose we combine Tables 1 and 2, as some grand measure of a bullpen. Two discrepancies leap out, Boston and Pittsburgh. Boston in Table 1 has a bad bullpen, bad in TRA, OB%, and BA. But, according to Table 2, they held the lead well. Pittsburgh is the opposite. In Table 1, they were better than average in all three categories; the BA against their bullpen was .228. In Table 2, they are below average, only LA was worse. These discrepancies bother me. The HL% in Table 2 is not perfectly correlated with TRA in Table 1, but there is a relationship. And it makes sense that the two things might go together. There are several ways to explain the discrepancy. One way is that different pitchers contributed to the the two stats. Every bullpen has pitchers they use when behind; maybe for Boston those pitchers were especially bad, and for Pittsburgh, they were actually quite good. (The converse also holds; that the pitchers used to protect a lead were the best pitchers for Boston, and the worst for Pittsburgh.) As a result, the TRA would be affected in games when the %HL is not. There may be some of that here, but neither team had a real well-defined role for everyone, at least until Schiraldi showed up for Boston, and he wasn't used that much. A bigger problem is the one mentioned earlier in connection with saves. All held-leads and squanders are given the same value. It is possible that Boston was using its bullpen when ahead by several runs, so that the runs allowed did not affect the outcome of the game. Maybe Pittsburgh used its bullpen more when the game was tied or they were only a run ahead; in that situation, every run allowed would affect the outcome of the game. I checked it out, with the results shown in Table 3.

What I have done is to break down the entrance of the bullpen by game situation, and then compiled the TRA for each bullpen in each situation. For example, Boston's bullpen entered the game four or more runs behind 17 times; when they did this, their TRA was 3.55. I included two control teams for each league. These are the teams that, from Table 1, most closely matched Boston and Pittsburgh for HL% and staff differential. The NL controls are Montreal and Cincinnati; the AL controls are Milwaukee and Cleveland. The reason I matched for HL% is obvious; that is what seems out of line for the TRA the bullpen compiled. The reason I matched for staff differential rather than TRA was that I wanted to look at different responses to similiar situations. For example, the team closest to Boston in bullpen TRA is the Yankees. But they are clearly a poor control because Pinella was replacing a bad starter (on average) and McNamera was replacing a good starter. Thus, the prediction in comparing the Yankees with Boston is that the starters got pulled in close situations in New York but not in Boston. That's true, by the way, but it doesn't make for a good comparison. Even though Cleveland's TRAs are not particularly similar to Boston's, they were nearly identical in that the bullpen was .7 of a run worse than the starters.

With all of that, let's look at the data. Remember our problem: Did Boston pitch well when ahead and poorly when behind, thereby increasing their HL% and TRA, or did they get more opportunities when the runs didn't matter much? For Pittsburgh, we ask the opposite side of the same question. And in both cases, the answer is clear. Pittsburgh pitched very well when behind, and very poorly when ahead, particularly when one or two runs ahead. (They did pitch well when the game was tied.) Compare that to Montreal, who had the same HL%. The Expos, with a worse TRA, pitched better when ahead than the Pirates did. Cincinnati was pretty consistent under all situations. So what doesn't show up in the Pirates individual pitchers shows up here--they pitched poorly when ahead, for whatever reason. They did not get an unusual number of close games, which would also have accounted for the discrepancy.

For Boston, the answer is also pretty clear. First, they got a lot of chances when the game wasn't close, 15 times with a four or more run lead. And of those 15, ten were two innings or less, so they were often called in to mop up the last two innings of blowouts. They were not called on when the game was close, only 13 times with the score tied, compared to Cleveland's 21. And they pitched a little better when ahead than when behind. (I offer no explanation for the -2 results.) What is the most interesting thing in the AL data is the result with Milwaukee. They were exceptional at holding a lead. We can give credit to Plesac and Clear, since they were the two most responsible for those remarkable stats.

In summary, we can see that saves can be better used when expressed as a SV%, and that the same concept can be better served by the %HL. I've not shown these stats for individual pitchers, but that is clearly an appropriate application. HL% will serve even better than SV% when applied to individuals. The biggest problem is with weighting SQ and HL. I've devised various systems, but I don't like any of them enough to pass on. A good system of weighting these things has to include three components--IP, runs allowed, and game situation. Ideally, it will combine them in a simple, logical way. The fractional run adjustment to TRA is probably not new, but I think it could add significantly to our understanding. Maybe we could eventually use it for assigning wins and losses, too. I'm surprised by the small effects I saw. Again, if anyone has data on other teams, or ideas and criticisms, I would like to hear.

TABLE 1. TEAM TRA, OB% AND BA

1A. National League Pitching Staffs

TEAM	%RIP	ST TRA	REL TRA	STAFF DIFFER	ST OB%	REL OB%	ST BA	REL BA
CUBS	33.2	4.593	4.451	.152	.3310	.3508	.2679	.2768
MONT	33.5	4.350	3.845	.510	.3168	.3138	.2437	.2392
METS	26.2	3.569	3.613	-.043	.2931	.3238	.2301	.2441
PHILA	32.4	4.522	4.073	.453	.3303	.3256	.2594	.2628
PITTS	32.2	4.362	3.783	.602	.3281	.3157	.2606	.2285
ST L	27.8	3.824	3.426	.402	.3048	.3127	.2454	.2408
ATL	30.2	4.432	3.682	.810	.3374	.3212	.2665	.2451
CIN	32.4	4.562	3.915	.653	.3197	.3261	.2632	.2448
HOUS	30.0	3.695	4.049	-.327	.2859	.3131	.2209	.2275
LA	27.6	4.237	4.468	-.226	.3067	.3458	.2501	.2648
SD	36.3	5.097	4.310	.737	.3340	.3207	.2613	.2372
SF	32.4	4.156	3.474	.660	.3109	.3062	.2375	.2174
AVE	31.2	4.277	3.936	.341	.3164	.3231	.2505	.2443

1B. American League Pitching Staffs

TEAM	%RIP	ST TRA	REL TRA	STAFF DIFFER	ST OB%	REL OB%	ST BA	REL BA
BALT	31.4	5.022	4.496	.526	.3263	.3274	.2617	.2550
BOST	24.6	3.953	4.688	-.735	.3100	.3620	.2550	.2864
CLEVE	29.5	4.913	5.632	-.719	.3386	.3644	.2677	.2779
DET	28.0	4.581	4.118	.463	.3210	.3247	.2465	.2533
MILW	31.3	5.134	4.184	.950	.3282	.3235	.2705	.2488
YANK	34.9	5.100	4.558	.542	.3298	.3209	.2662	.2484
TOR	32.3	4.289	4.064	.225	.3231	.3130	.2629	.2428
CALIF	28.0	4.198	4.305	-.107	.3026	.3194	.2397	.2560
WSOX	32.9	4.614	3.717	.897	.3212	.3175	.2476	.2452
KC	30.4	4.551	3.411	1.140	.3192	.3135	.2605	.2389
MINN	27.3	4.822	5.905	-1.08	.3288	.3690	.2700	.2942
OAK	30.1	5.034	4.984	.050	.3252	.3416	.2431	.2469
SEA	32.1	5.277	4.474	.803	.3547	.3371	.2808	.2694
TEX	33.1	5.208	4.074	1.134	.3405	.3329	.2451	.2450
AVE	30.4	4.756	4.454	.302	.3264	.3328	.2585	.2571

Definitions: %RIP-percent of team innings pitched by relievers
ST-starters. REL-relievers. TRA=9*RUNS/IP. STAFF DIFFER=ST TRA- REL TRA.
OB% -on-base percentage. BA-batting average.

TABLE 2. PERCENT OF SAVES; PERCENT OF LEADS HELD

A. National League

TEAM	SV	SQ	HL	% SV	%HL
CUBS	41	43	59	48.8	57.8
MONT	50	55	64	47.6	53.8
METS	46	35	79	56.8	69.3
PHILA	39	40	67	49.4	62.6
PITTS	30	43	50	41.1	53.8
ST L	46	33	64	58.2	66.0
ATL	39	41	54	48.8	56.8
CIN	43	37	70	53.8	65.4
HOUS	49	34	75	59.0	68.8
LA	26	46	44	36.1	48.9
SD	34	45	58	43.0	56.3
SF	35	32	67	52.2	67.7
LEAGUE	478	484	751	49.7	60.8

B. American League

TEAM	SV	SQ	HL	% SV	%HL
BALT	38	28	58	57.6	67.4
BOS	41	19	63	68.3	76.8
CLEVE	33	42	56	44.0	57.1
DET	38	31	59	55.1	65.6
MILW	32	23	55	58.2	70.5
YANK	58	41	84	58.6	67.2
TOR	44	48	75	47.8	61.0
CALIF	40	30	65	57.1	68.4
WSOX	38	39	59	49.4	60.2
KC	33	32	60	50.8	65.2
MINN	24	41	47	36.9	53.4
OAK	37	45	57	45.1	55.9
SEA	26	40	46	39.4	53.5
TEX	40	37	69	52.0	65.1
LEAGUE	522	496	853	51.3	63.2

Definitions: SV-saves. SQ-squanders. HL-held lead. %SV=SV/(SV+SQ). %HL=HL/(HL+SQ)

TABLE 3. TRA BY GAME SITUATION

	RUN DIFFERENCE								
	−4	−3	−2	−1	+0	+1	+2	+3	+4
PITTS	3.02	4.44	3.44	4.43	3.30	5.64	6.93	2.50	4.74
	(25)	(18)	(18)	(15)	(22)	(14)	(15)	(6)	(11)
CINCINN	3.74	3.81	4.08	4.17	4.36	3.34	4.45	4.76	2.94
	(30)	(12)	(15)	(17)	(17)	(19)	(11)	(18)	(8)
MONT	5.00	3.88	2.00	4.53	3.28	3.39	3.72	2.60	6.55
	(28)	(15)	(15)	(15)	(22)	(22)	(11)	(10)	(8)
BOSTON	3.55	3.67	5.29	10.15	4.05	3.21	3.79	6.23	4.91
	(17)	(14)	(18)	(11)	(13)	(15)	(12)	(9)	(15)
CLEVE	7.08	6.75	3.45	3.75	5.60	6.55	4.33	4.22	5.14
	(25)	(16)	(8)	(11)	(21)	(14)	(17)	(8)	(9)
MILW	5.44	2.57	3.56	4.34	5.89	1.60	2.70	1.08	.89
	(33)	(14)	(11)	(15)	(18)	(14)	(13)	(3)	(11)

The top line is the total run average when the bullpen entered the game under those circumstances. The number in parentheses is the number of times this occurred. The -4 and +4 columns include 4 and greater.