Defensive Regression Analysis:
A New, Objective and Reliable Pitching and Fielding Evaluation Method That Can Be Applied Throughout the History of Baseball

By Michael A. Humphreys

INTRODUCTION

> "As late as June 4, 2002, . . . there were still big questions about baseball crying out for answers; a baseball diamond was still a field of ignorance. . . . No one had established to the satisfaction of baseball intellectuals exactly which part of defense was pitching and which fielding, and no one could say exactly how important fielding was. No one had solved the problem of fielding statistics." Michael Lewis, *Moneyball*, p. 98 (W. W. Norton & Co. 2003).

Thanks to the efforts of many people, a solution to the problem of estimating the run-impact of fielding using traditional fielding and pitching statistics—and evaluating the run-impact of pitching independent of fielding—may be within reach. In this article, I would like to introduce a new system that provides runs-saved estimates for pitchers and fielders using traditional statistics publicly available throughout the history of baseball—in other words, without recourse to non-public "zone" data, which has only been compiled since the late 1980s.

The system is called Defensive Regression Analysis ("DRA"). (The acronym conveniently rhymes with ERA; fans unfamiliar with regression analysis can think of it as Defensive Runs Analysis. I'll provide a simple description of regression analysis in Part I below.) To the best of my knowledge, DRA is the first defensive system that

(i) deals with pitching and fielding simultaneously, i.e., on a fully integrated basis,
(ii) does not rely on any subjective weights or factors,
(iii) uses only publicly available statistics in existence throughout the history of baseball (e.g., no Caught Stealing or Sacrifice Hits Allowed), *and*
(iv) estimates, using simple "linear weights" equations (similar in form to Linear Weights equations used for evaluating hitters in the official baseball encyclopedia, *Total Baseball*), the runs saved (allowed) by pitchers and fielders (a) relative to the league average and (b) independently of each other.

A team's pitcher and fielder DRA ratings add up to the team DRA rating, i.e., an estimate of the number of runs the team *should* have allowed. The standard error of such estimate in the 1974-2001 study used to create DRA, 19.7 runs, is *less* than the standard error for

regression analysis models for team runs *scored*, as well as those of all other systems for evaluating offense described in the recent book co-authored by two former Chairs of the Sports Section of the American Statistical Association. *See* Albert and Bennett, *Curve Ball: Baseball, Statistics, and the Role of Chance in the Game* (Copernicus Books 2001) (hereinafter, "*Curve Ball*"). In other words, the "parts" under DRA "add up" as well or better than the "parts" under Linear Weights or Runs Created or any other system for batting and baserunning of which I am aware.

What is probably most exciting about DRA is that it provides individual fielder runs-saved estimates that match up well with runs-saved ratings derived from proprietary zone data, such as Mitchell Lichtman's Ultimate Zone Rating ("UZR") system (available at baseballprimer.com) and zone-based evaluations posted by Diamond Mind ("DM") on its website. (Disclaimer: I have neither sought nor received any endorsement or any other support from Diamond Mind. The DM evaluations that I cite are all publicly available on its webpage. My interpretations of DM evaluations are entirely my own.) DRA ratings have the same "scale" as UZR ratings, and an over 0.8 correlation with UZR ratings, when the latter are adjusted to incorporate DM evaluations.

Zone-type ratings, as you're probably aware if you're reading this, are based on highly detailed records of the actual number of balls hit into each of approximately 80 "zones" on the field. Zone data began to be collected in the 1980s because of the fundamental problem with traditional fielding statistics: they provide no direct record of fielding *opportunities*. Judging fielders based on *gross* plays made (e.g., the total number of fly balls caught by an outfielder) was therefore no more reliable (actually far less reliable) than judging batters by the gross number of their hits and walks. Zone ratings essentially compare the number of plays made by the fielder compared to what the league-average fielder would make, given the same number and pattern of balls hit into his zones. UZR converts the "plays made" numbers into runs saved, based on the change in expected runs allowed if a given play is made or not made; i.e., depending on whether the play made, on average, prevents a single or extra-base hit.

Provided in Part II below are position-by-position charts of DRA and UZR runs-saved ratings for all 82 players who played at least two seasons full-time (130 or more games) at one position during the three-year time period (1999-2001) for which I had access to both UZR and DRA data. I'm not aware that any other non-zone fielder rating system has ever been compared with zone ratings in as systematic a way.

Although you will have to be the ultimate judge, my sense is that the average rating per player over that time period under DRA essentially matches up with the UZR rating and/or DM evaluation close to 95% of the time. In other words:

> Given two or three years of publicly available data, DRA provides a reliable estimate of whether a fielder is basically average (+/- half a dozen runs per season), meaningfully above or below average (+/- a dozen runs a season), or exceptionally above or below average (+/- two dozen runs per season).

I am *not* promising exact matches. For example, DRA rates Nomar Garciaparra as a +5 (runs saved) shortstop per season in 1999-2000, whereas UZR rates him at –6. Given the imperfection in statistics, including UZR data, as well as the *practical* significance of runs-saved numbers of that relatively modest magnitude, I view that as an acceptable match, meaning that both systems identify Nomar as an essentially average-fielding shortstop. The *relevant* findings under DRA (and supported by UZR and DM) would include, for example, that Derek Jeter is costing the Yankees enough runs to justify moving him to third base, and that Pokey Reese was saving his teams so many runs at second base (at least in 1999-2000) that it made sense to let him play, even though he is weak hitter.

I think it fair to say that out of the 35 players with three years of data, there is only one clear and significant error, and out of the 47 players with only two years of data, there are only four clear errors. In addition, the errors are "conservative". By that I mean there are no false "positives" or "negatives" in the study; DRA might fail to identify a good fielder (in the study, it did not fail to identify any bad fielder), but it doesn't rate fielders in the study as meaningfully good or bad who aren't good or bad. Though DRA might be slightly more conservative than UZR, the runs saved (allowed) ratings under DRA nevertheless have approximately the same "scale" of impact as UZR ratings: the highest DRA ratings are as high as (and no higher than) the highest UZR ratings; the lowest DRA ratings are as low as (and no lower than) the lowest UZR ratings.

One simple way of quantifying the *overall* match between DRA and UZR/DM is to "regress" the UZR average ratings per player onto the corresponding DRA ratings. Regression analysis reveals that the average DRA rating of players in the study has, on average, almost precisely the same "scale" as the average UZR rating (the "coefficient" for DRA under the regression equation is nearly 1.0, actually, 0.95), and a correlation of 0.7. When, for reasons explained in Part II.A.4, UZR ratings are adjusted to account for DM commentary, admittedly in a somewhat subjective manner, the correlation improves still further, to slightly over 0.8. (All of the above results are provided in detail at the very end of Part II.)

Unlike zone systems, DRA can provide ratings for players throughout the history of baseball. Part III (and the Appendix to Part III posted as an Excel file) provides DRA ratings of all players who played at least 130 games at a single position for at least five seasons anytime between 1974 and 2001, the time period for which I had convenient access to the relevant data. I hope you will spend some time looking over these ratings, as they demonstrate even better than the three-year DRA-UZR-DM comparison the basic reliability of the system. In the majority of cases, players peak in their youth. The year-to-year ratings show remarkable consistency, including when players change teams. DRA ratings drop, sometimes sharply, after players are injured. There are almost no weird run-saving values—historically high assists and putout totals do not result in absurdly high ratings. Great as he was, and notwithstanding his historic assists totals, Ozzie Smith was *never* "saving" 50 runs a season, even at his peak—20 runs a year was more like it. What made Ozzie probably the greatest shortstop in history (and, therefore,

probably the greatest *fielder* in history) is that he consistently performed at or near that level for about ten years, before declining to an average level of performance.

Although I will provide a description of the principles and methodology of DRA, as well as the results of several diagnostic tests, the "linear weights" equations per position are not provided. (The format of the equations, however, is shown in Part I.A.) I am currently approaching several major league organizations regarding DRA, which can serve as a simple tool for double-checking zone ratings (which, due to their computational complexity, have something of a "black box" quality) and for evaluating minor league fielders (for whom zone data is unavailable). When I say that DRA is a simple tool, I mean that it is *far, far* simpler than any other pitching and fielding system (zone or non-zone) that has been described in print or a public Internet forum, either fully or in general terms. All of the equations can fit on one page, and although most elements of the equations are completely novel, they would make immediate, intuitive sense to any serious baseball fan after a brief explanation.

I can appreciate that it is somewhat difficult to place much faith in a system when not all of the details about it are available, although most fans seem to feel comfortable accepting UZR and DM ratings without having access to the data and all of the calculations under those systems. In the case of UZR, Mitchell Lichtman has done an excellent job of explaining the core "concept" behind zone ratings and the factors that must be considered in using zone data intelligently. Fans, therefore, feel comfortable "buying into" the system, even though the underlying data is proprietary. Tom Tippett at Diamond Mind is also very clear about the factors he considers in evaluating fielders although, again, his data is not publicly available. My hope here is that

(i) the description of the general principles and methodology of DRA will reassure you that the core concepts of the system are sound and that great care has been taken in making the system work,

(ii) the results under DRA, especially how the ratings in the 1999-2001 data set compare with UZR and DM, will instill confidence in the system and pique interest in learning more about it, and

(iii) the *im*perfection of the output will reassure the skeptics among you that the ratings were not just cooked up.

Should it be the case that DRA is of more interest to fans than major league organizations, I would welcome the opportunity to publish the DRA equations in a book that rates the greatest fielders throughout major league history.

Were I to reveal the DRA equations here, this article could be condensed to about ten pages, and anyone who reads Bill James, Pete Palmer, Dick Cramer or TangoTiger would "get it" and endorse it immediately. However, I'm trying to make the case for DRA indirectly, so I have to bring more subtle and lengthier arguments to bear. As I'm using statistical techniques and approaches never tried before, some amount of

background information about them is in order. Most importantly, I'm ultimately relying on the *output* of the system to make the *case* for the system, and discussing that output is a lengthy task.

I have, however, divided the article into four Parts, so you can focus on what is most interesting to you and skim the rest. Part I provides an overview of the principles and methodology of DRA, including many new insights about Defense Independent Pitching Statistics ("DIPS") (the relative impact of pitchers and fielders on whether batted balls fall in as hits). Part II compares 1999-2001 DRA results with UZR ratings and DM evaluations. Part III provides historical DRA results from 1974-2001. Part IV addresses various miscellaneous issues, including how DRA ratings can complement and improve zone ratings, the practical relevance of applying DRA to evaluate minor league fielders, DRA's role in stirring and settling various Hall of Fame debates, how and why DRA can adapt to changing pitching and fielding dynamics and record-keeping over the course of major league history, as well as how DRA ratings (combined with Linear Weights batting and baserunning ratings) could be converted into replacement-level Win Shares and Loss Shares.

* * *

ACKNOWLEDGEMENTS

Before we go any further, I'd like to thank several people.

Dick Cramer reviewed several earlier versions of DRA, and was the first person to appreciate the underlying logic of the method. Without his encouragement, I doubt I would have had the patience to keep working on DRA. In addition, Dick's research supports certain key assumptions under DRA. Finally, as a founder of STATS, Inc., Dick made possible the compilation of the data against which DRA could be tested. I'd like to thank Neal Traven at SABR for forwarding my first article about DRA to Dick.

Chris Dial, Mike Emeigh, Mitchell Lichtman ("MGL"), Charlie Saeger, and Tom ("TangoTiger") at baseballprimer.com have all provided key information and insights. Charlie developed years before anyone else many fundamental ideas for fielding evaluation. Throughout this article, I frequently use Charlie's terminology of "context-adjusted" fielding plays, although our systems are different. MGL has provided an invaluable service to the sabermetric community by publishing the results and basic methodology of his UZR system. Mike Emeigh's fielding articles helped me see a way of improving UZR, and MGL graciously accepted the suggested changes, which Chris and TangoTiger had previously suggested as well. Although there will be times when I will point out certain UZR ratings that might still be a little "off", UZR *is* fundamentally reliable, DRA isn't perfect either, and without the work and insights of Chris, Mike, MGL and Tango, there wouldn't *be* a system against which I could test DRA. Tango also provided team-level UZR data that helped me make the most recent and *very* significant

improvement in DRA. I'd also like to thank whoever is or was responsible for putting together the baseball-reference.com website, which is extremely well-designed.

I owe an enormous debt to all of the people who have contributed to Retrosheet, as well as to John Jarvis, both for his creative articles and for putting together team-level Retrosheet data in an easy-to-use format on his website, which was the primary source of data for DRA. As will be explained in Part IV.C, DRA would not have been possible without Retrosheet data, but can nevertheless be applied to seasons in baseball history for which we as yet do not have such data.

Pete Palmer is the class act of sabermetrics, and has been an invaluable help to me in understanding the quality and scope of statistics throughout baseball history.

Steve Pappaterra is a great and good friend who first introduced me to the work of Bill James twenty years ago.

Bill James wrote me a kind and encouraging e-mail when I sent to him an over-long and under-organized letter in which I tried to describe an early version of the DRA method. Although DRA differs significantly from Bill's defensive Win Shares system, I was inspired to work on this project after reading his *Win Shares* book. Moreover, I would never have been able to develop DRA had I not learned from Bill that ". . . fielding statistics make vastly more sense if you look at them from the top down than they do if you look at them from the bottom up. * * * To make sense of fielding statistics, sometimes you have to start with what the team accomplished, then ask how they accomplished that, and only then work toward the question of which player gets credit for that success." *Win Shares*, p. 11. DRA is founded upon the basic principle—introduced in *Win Shares*—that *everything has to add up*.

* * *

I. PRINCIPLES AND BASIC METHODOLOGY OF DRA

A. A "Bare-Bones" Summary

DRA is, as far as I know, the first system to use regression analysis to evaluate fielding and pitching, or at least the first defensive system to use regression analysis so comprehensively. For the sake of brevity, I'm going to provide a two-page summary of the DRA "process", with particular focus on one position (shortstop), and rely on some technical terminology that for some of the readers I'm trying to reach might be a bit daunting; I'll explain the technical terminology Part I.B immediately below. Under DRA, we

(i) Collect team-level defensive data of a sufficient sample size

(e.g., the runs allowed, strikeouts generated, walks allowed, shortstop assists, etc, per team; DRA was developed using team-level data for all teams in 1974-2001);

(ii) "adjust" each defensive variable for each team (strikeouts, shortstop assists, etc.) using simple arithmetic formulas, so that each such event is at least *arithmetically* independent of the others

(e.g., determine the number of (a) strikeouts generated by the team's pitchers, taking into account the number of batters facing the team's pitchers ("BFP"), *not innings pitched*, thus yielding "adjusted SO" ("**a**SO"), (b) assists by the team's shortstops, taking into account the number of batted balls in play allowed by the team's pitchers ("BIP"), *not innings played*, thus yielding "adjusted" A6 ("**a**A6"), and (c) the number of shortstop *errors*, taking into account the number of chances, or "**a**E6");

(iii) "regress" the *arithmetically* adjusted team-level fielding plays data (e.g., arithmetically "adjusted" assists at short, or "**a**A6") in the sample "onto" all the contextual variables derivable from publicly available data that we think *might* have an impact on the *distribution* of BIP, the opportunity to record assists on double play pivots, and the *positioning* of shortstops for purposes of fielding BIP

(e.g., regress aA6 "onto" pitcher variables, such as variables measuring BIP against left-handed pitching and a variable estimating the relative tendency of the team's pitchers to generate ground balls or fly balls, but also variables such as an estimate for runners on first, which impact positioning and double play opportunities);

(iv) eliminate the variables that do not have a statistically significant impact on the number of assists made at shortstop, and re-run the regression;

(v) treat the "residuals" of the final, stripped-down, regression analysis result as the "skill" plays made at shortstop, i.e., the *regression*-"adjusted" aA6, or "**a**aA6"

(e.g., if the regression result indicates that each marginal BIP against left-handed pitchers *increases* aA6 by 0.0150, and the team had 1000 "excess" BIP allowed by left-handed pitchers, we *decrease* the team's **a**A6 by 15 to reveal the shortstop plays *not* "given" to the shortstops by a high level of left-handed pitching, in order to provide a better estimate of fully context-adjusted shortstop assists, or **a**aA6);

(vi) repeat steps (ii) through (v) for plays-made data at all of the other fielding positions (including pitcher and catcher), thus yielding the "context-

adjusted" plays made at all of the positions (e.g., aaA1, aaA2, aaPO7, aaA7, etc.);

(vii) "regress" team-level *runs-allowed* data "onto" team-level context-adjusted pitching data (e.g., aSO, aBB, aHR, aWP, etc.) *and* the team-level context-adjusted plays made in the field determined in steps (ii) through (vi) (e.g., aaA6, aE6, aaPO7, aE7, etc.);

(viii) eliminate all fully context-adjusted variables that do not have a statistically significant impact on runs allowed (e.g, all errors, except at pitcher and right field) and re-run the regression; in order to

(ix) determine the average value, in *runs*, of each context-adjusted play made (e.g., the average runs-saved value of each aSO, aHR, aaA6, etc.).

A team's defensive rating at a position is derived from the *number* of context-adjusted plays made at such position (determined in part through regression analysis), e.g., aaA6, multiplied by the average *run-value* of each such play (determined through regression analysis).

The formulas for each position (including pitcher) thus resemble (and are no more complicated than) Pete Palmer's "Linear Weights" equation for batters found in *Total Baseball*. For example, shortstops ratings are based on their assists ("A6"):

> A6 (adjusted arithmetically for BIP, or "**a**A6") +/- A*Factor A +/- B*Factor B +/- C*Factor C, etc., *equals* **a**aA6
>
> (where the A, B and C "weights" are derived from regression analysis and contextual Factors A, B and C are derived from publicly available information).

The rating at shortstop equals the context-adjusted plays made, aaA6, multiplied by the regression-calculated weight in runs for each aaA6, adjusted so that the rating is +/- the league average runs-saved rating at shortstop.

In the case of pitchers, the formulas work for their individual stats, with one exception described in Part I.D.1. In the case of fielders, an individual's rating at such position is based upon the team rating at that position, pro-rated for his innings fielded, and adjusted up or down for his rate of plays made relative to the team rate of plays made at that position.

All of the ratings add up to the DRA estimate of runs allowed by the team. As mentioned in the Introduction, the standard error of such estimate in the 1974-2001 sample is *less* than that for any system for rating team *offense* of which I am aware.

B. A Little Bit of Theory

A (relatively) brief discussion of the basic theory behind regression analysis and how regression analysis has been applied to evaluate run *scoring* may help explain why regression analysis has never been tried before to evaluate run *prevention*.

Multi-variable linear regression analysis ("regression analysis") is a statistical tool for determining the marginal impact each variable in a multi-variable model has on the ultimate outcome being modeled. As explained in the book *Curve Ball*, regression analysis can be used to estimate the impact of each type of batting and baserunning event on the total number of runs a team scores.

For example, if you provide a computer with a sufficiently large sample of rows of historical annual team-level data consisting of at-bats minus hits, walks, singles, doubles, triples, home runs, stolen bases, etc. (collectively, the variables), as well as the actual number of runs each team scored that season (the ultimate outcome being modeled), and (politely) "ask" the computer to "regress" team runs scored "onto" the variables, the computer can perform a regression analysis that will estimate the marginal increase or decrease in team runs scored that is associated with (loosely speaking, statistically correlated with) each variable. Regressions usually show that for each additional home run hit, a team will, on average, score an additional 1.5 runs, assuming all other variables are held constant. For each *two* additional home runs, a team will score, on average, an additional 3.0 runs, and so forth.

As explained in *Curve Ball*, the run-values per offensive event have been verified empirically by "change-in-state" models developed by George Lindsay and Pete Palmer. "Change-in-state" models analyze the changes in expected runs scored before and after each offensive event (e.g., a home run) in large numbers of actual baseball game situations, based upon the change each offensive event (e.g., a home run) has on the runs scored, the number of outs, and the position of baserunners after each event. *See Curve Ball*, pp. 178-205.

For regression analysis to "work" (i) each variable must have a reasonably *straight-line* relationship to the ultimate outcome and (ii) each variable must be reasonably *independent* of the *other* variables. The first assumption enables us to say not only that if a team hits an additional 2 home runs it should score an additional 3 runs, but also that if a team hits *20* more home runs than average it should be expected to score, all other variables being equal, *30* more runs. The second assumption is necessary for the technique to reveal the independent marginal impact of each variable—if the variables are *correlated* with *each other*, the computer can't calculates the marginal, *independent* impact of the variable, because the computer can't "hold" all the other variables constant while it's "calculating" the marginal impact of the variable under consideration—they're "moving" with (or against) the relevant variable.

Although the process of run-scoring is *not* linear, an individual player's contribution to the number of runs his team scores—and the marginal impact of each element of a team's offense—*is* approximately and reasonably linear over the range of

typical major league run-scoring scenarios over the course of a season. That is why Pete Palmer's Linear Weights equations work so well for batters. The latest version of Bill James' Runs Created is essentially linear. *See Curve Ball*, pp. 230-41. In addition, the impact in terms of team *wins* of a given player's *run* creation is reasonably linear. "Within the range where the teams are clustered, a linear representation of value works perfectly well—exactly as Pete Palmer has always insisted that it did." *Win Shares*, p. 108.

Similarly, the process of *allowing* runs is approximately and reasonably linear. TangoTiger has done interesting work on how extremely good individual pitchers should have a disproportionate, non-linear impact on the number of runs their teams allow. I agree this is true to some extent. Pitchers (unlike batters) "create" their own run "context", and their skills "interact" with each other and with fielders. If The Big Unit strikes out twice as many batters as a typical pitcher, the baserunners he "takes away" significantly reduce the expected impact of whatever singles his fielders might "allow". I performed a DRA analysis of the most dominant starting pitcher in history, inning-per-inning: Pedro Martinez. The estimated runs saved under DRA by Pedro and his fielders (who, over the course of his career, were very slightly better than average) was indeed about 5% less than the number of runs Pedro actually "saved", as measured by runs allowed by his teams while he was pitching. Even assuming there is some non-linearity for extreme pitchers, the effect is fairly modest, and *fielding* has a linear impact, as *team-level* pitching staff quality is much less likely to have the effect Tango has described. (For those of you who are familiar with regression analysis, the "residuals" generated under the various regressions used in DRA did not reveal any non-linearity.)

There is, however, one fundamental difference between offensive and defensive statistics: offensive statistics are generally not significantly correlated, whereas defensive statistics are *highly*, and *by definition*, cross-correlated.

Just because a team *draws* a lot of walks does *not* necessarily "prevent" or "cause" it to *hit* a lot of singles, or even a lot of homeruns. Each event is thus reasonably independent. Therefore, regression analysis of offensive statistics has long provided reasonable run-value estimates for various discrete offensive events. Doubles and homeruns show some minor cross-correlation (teams that hit more doubles tend also to hit more homeruns), which probably explains why the weights for doubles and home runs are slightly different under regression analysis and Lindsay-Palmer "change-in-state" models. *See Curve Ball*, p. 203 (home run weight of 1.4 under Palmer methods). But, in general, a simple regression analysis of batting and baserunning data works. When you plug the regression-generated run-weight values to the actual offensive data per team, you obtain an estimate of the number of runs the team *should* have scored. The "error" in such estimate, that is, the "residual" not explained by such procedure, can be quite low (in a 1954-1999 regression analysis, the standard error was 0.142 runs per game, or 23 runs per 162-game season). *See Curve Ball*, pp 178-84.

In contrast, each strikeout by a team's pitchers literally eliminates an opportunity for the team's fielders to make a play. Thus shortstop assists and outfielder putouts are

by definition *negatively* correlated with strikeouts. In addition, "raw" fielding play data at each of the positions has extremely strong *positive* cross-correlations. For example, unadjusted shortstop assists are strongly correlated with second and third base assists, as they are both affected by "ground ball" pitching, so if you ran a simple regression of team runs allowed on all defensive plays, the computer wouldn't be able to "tell" the marginal impact of the shortstop plays, because in most cases in which the shortstop plays go up, second and third base plays go up as well. This problem probably explains why no one has ever used regression analysis before to evaluate fielding. You can't just "regress" runs allowed "onto" traditional team defensive statistics, and come up with anything useful.

DRA "untangles" the cross-correlations among all the defensive events measured by traditional pitching and fielding statistics (e.g., strikeouts, infielder assists, outfielder putouts, etc.), *both* arithmetically and through regression analysis, so that each such "event" is context-adjusted and *independent* and, therefore, *can* be reliably associated (through regression analysis, and with statistical significance) to *runs*, the "money of baseball, the common denominator of everything that occur[s] on the field." *Moneyball*, p. 131.

We first adjust each defensive variable *arithmetically*, so that each is no longer (negatively) correlated *by definition* with any other variable. If you just took strikeouts per inning pitched and shortstop assists per inning played, each would be negatively cross-correlated by definition, as each strikeout takes away an opportunity to record a shortstop assist, given the *fixed* number of outs in a game/season. Therefore, we *put* each event into the relevant "context" or "denominator" of opportunities, which *differs* by position: BFP for SO; BIP for A6. (In contrast, *offensive* events all *share* the same "context" or "denominator" of opportunities—outs.) At this point we now have arithmetically "adjusted" variables: e.g. **a**SO and **a**A6.

To eliminate the *remaining* cross-correlations among the variables, we use regression analysis and the fact that baseball defense has a clear direction of causality: pitchers cause more or fewer shortstop chances; shortstops don't "cause" pitchers to allow more ground balls. We can therefore regress pitcher variables "onto" defensive events to reveal the *number* of plays made by fielders at each position *not* "explained" by pitcher variables. That number is the "residual" of the regression analysis: the unexplained part that reflects, on average, fielder skill. To calculate that "skill" residual per team, we just "back out" the effect of the pitching variables for each such team. If the regression result for the 1974-2001 sample indicates that each marginal BIP against left-handed pitchers on average *increases* aA6 by 0.0150, and a team has 1000 extra BIP against left-handed pitching, we *decrease* the team's **a**A6 *by 15* to reveal the shortstop plays *not* "given" to the shortstops by a high level of left-handed pitching, to provide a better estimate of fully context-adjusted shortstop assists, or **a**aA6, that are a better estimate of skill plays made.

Now we can run a "global" regression of runs allowed data "onto" all of the now statistically independent variables: aSO, aBB, aaA6, etc., etc., to reveal the statistically significant run-impact of each marginal event.

When you apply these run weights to the number of context-adjusted plays made by a team at a particular position, you effectively arrive at an estimate of the likely change the players at such position brought about in the number of runs their team should have allowed, given the statistically significant relationships among all of the publicly available statistics of the team and between those statistics and the actual runs allowed by the team. You can then "individuate" the ratings per-player at each position in the manner described above in the "Bare-Bones Summary".

C. Some Preliminary Results to Indicate We're On the Right Track

Although the basic "idea" behind DRA, as summarized above, is fairly simple (though, again, entirely new), the devil, as always, is in the details. It has taken me more time than I care to admit figuring out the simplest and most effective ways in which to (i) eliminate the cross-correlations between and among pitching and fielding events and (ii) improve the model's fit to (a) what we all "know" about various events and (b) UZR ratings and DM evaluations.

Under the latest DRA model, with one minor exception having no impact on fielder ratings, no context-adjusted pitching or fielding event has a cross-correlation with any other context-adjusted pitching or fielding event greater than 0.2. Before such transformations, many such events had cross-correlations in excess of 0.6. As a general rule of thumb, correlations below 0.3 are generally viewed as non-significant.

The resulting run values per context-adjusted event are very much those you would expect. Each marginal walk given up in the 1974-2001 data was associated with giving up an additional 0.34 runs. The Palmer weight for a walk *drawn* is 0.33. Each homerun was associated with an additional 1.44 runs. The Palmer weight is 1.40 runs. Strikeout values varied according to run environment—approximately .31 runs saved in the high-offense 1990s; closer to .26 runs saved in the low-offense 1970s. Pete has similarly found that out values vary between .30 and .25 in the same way, and MGL has found that outs in the form of strikeouts are just slightly more valuable to a defense than other outs; i.e., outs in the form of strikeouts in the high-offense era have been worth about .31 runs saved rather than .30 runs saved. (DRA fielding ratings take into account—simply and objectively—the changing value of outs over time.) The run-value of a wild pitch/passed ball/balk ("WP") was .37 runs allowed, slightly more than the Palmer weight for a stolen base, but that makes sense, because more than one runner can advance on a WP. The value of a stolen base allowed was not stable—under the current model it appears to be half of the UZR value per stolen base, and well below the Palmer weight. But I'm living with it, and the catcher ratings reflect the lower value. In the catcher ratings section in Part II.B.8 I'll suggest some reasons to believe that the weight might be correct.

One of the cardinal principles under DRA is that *only* statistically significant statistics and relationships are used in the model. Non-significant factors are dropped, thus simplifying the model. Errors—except at pitcher and right field—have no statistically significant relationship with runs allowed, *after* taking into account plays made. This makes sense, to me at least. On *average*, the errors made at, say, shortstop, are no more damaging than the failure to make a play. A bobbled ball that stays in the infield is less damaging than a ball that finds its way to the outfield. In contrast, an error on a throw to first is sometimes worse than not making the play at all (unless the play is successfully backed up by the catcher or second baseman, the batter can often advance to second). On average, the mix of such errors is essentially run-neutral, after taking into account the fact that the play was not made successfully. (I believe the UZR weight for allowing a batter to reach base on an error rather than a base hit is less than 0.05 runs more damaging.) Pitchers, of course, can't position themselves to improve their range, and have limited mobility to make a play after delivering a pitch, so errors become significant. In right field, I suspect the reason errors are statistically significant is that throwing errors on throws from right field to third are the likeliest to result in a run scoring on the play.

This finding has a significant impact on how *pitchers* should be evaluated. The traditional method of rating pitchers is *Earned* Runs Average ("ERA"). ERA is a flawed method for rating pitchers, not only because of the "DIPS" issues discussed in Part I.D immediately below, but because ERA focuses on the *wrong* fielder factor, errors, for purposes of rating pitchers "independently" of fielders. A pitcher on a poor fielding team might have an unfairly high "Earned" Run Average because the team made relatively few errors but had poor range, while a pitcher on a strong fielding team might have an unfairly low "Earned" Run Average because the team made more errors but had good range. It is true that errors slightly correlate with poor fielding, but they have no statistically significant impact (except at pitcher and right field) once you take into account context-adjusted plays made.

But the problem is even worse than that. Error rates are *much* higher on ground balls than fly balls. So ground ball pitchers, who help *cause* errors, *benefit* by doing so, as the higher number of errors results in a lower "Earned" Run Average. The opposite applies for fly ball pitchers. Bill James has noted that

> ". . . [T]he Robin Roberts family of pitchers . . . tend to have lower component ERAs than actual ERAs. The component ERA is a formula which looks at a pitcher's hits allowed, walks allowed, and home runs allowed, and says, "this is the ERA that these should add up to." It'a good formula, but it's not a *perfect* formula, and pitchers such as Roberts, Newcombe, Jenkins and Hunter tend to have higher actual ERAs than component ERAs. Robin Roberts never led his league in ERA, but led his league in component ERA twice."

Roberts was an extreme fly ball pitcher, as were Jenkins and Hunter (I don't know about Newcombe). Roberts hurt his ERA by giving his fielders easier chances. DRA would redress this problem in rating pitchers.

The DRA-UZR-DM ratings comparison is provided and discussed in Part II.

D. Assumptions Supported by Other Studies

I stated in the Introduction that DRA has no subjective weights or factors. That's still true—each context-adjustment and run weight is the result of an objective determinant of opportunities to make plays (e.g., the number of batters facing the team's pitchers or the number of BIP allowed by the team's pitchers) or a statistically significant regression result, and *only* statistically significant variables are present in the model. However, I did make exactly four assumptions, each of which is supported by empirical data outside of the model. The assumptions are based in part on new insights regarding DIPS. Readers who already feel "burdened by too much information" should feel free to skip ahead to the ratings; I'm only providing this background information in case the full methodology is ever made public, and somebody would otherwise feel that they had been misled regarding some of the bolder assertions made here.

1. Responsibility for Estimated Infield Fly Outs

DRA, partly by default and partly by design, assigns responsibility for estimated infield fly outs to a team's pitchers.

Infield fly outs are assigned to a team's pitchers by default because—under the principle that *everything* has to add up—*somebody* has to take responsibility for them, and I know of no reliable method for assigning credit for them to fielders.

Put simply, we don't know with any degree of precision how many infield fly outs are actually caught by each fielder. I know of no reliable method for estimating the number of infield fly outs that are caught by middle infielders, due to the large number of putouts from runners caught stealing and force outs/double plays at second. I know of no method for separating out estimated unassisted ground out plays at first from estimated unassisted putouts at first. Probably most putouts at third are fly outs, but Bill James has found no relationship between third base putouts and fielder skill. At third, as well as other infield positions, it is almost always the case that a fly ball caught by one infielder could have been caught by another, so that even if we knew the actual number of fly balls caught by an infielder, it would largely reflect his tendency to "hog" chances, and not his contribution to *team* fielding success. For all of these reasons, I believe that UZR and DM do not include infield fly outs in infielder ratings.

The number of infield fly outs caught by a *team*, however, *can* be easily estimated (infield putouts—including non-strikeout putouts at catcher—minus *team* assists) and *does* contribute to team success, as ball-hogging nets out. Sure, there are plenty of first base unassisted ground outs and a few unassisted ground ball putouts at second base, but the *variation* from team-to-team overwhelmingly reflects variation in fly outs allowed by pitchers (as well as foul territory), as I'll discuss in the first baseman ratings section of Part III. And yes, sometimes infield fly outs overlap with outfield (particularly

centerfield) fly outs, but the *team* benefits whether or not the play is made by centerfielder or a middle infielder.

In determining where the credit really belongs for estimated infield fly outs, it is important to remember that infield fly outs represent a peculiar kind of pitcher-generated BIP—with the rare exception of "at 'em" line drives, all such BIP outs represent *weakly* hit balls that zone data reveals are caught by *somebody* on the team something like 95% of the time.

In light of these facts, it makes sense to assign credit for infield fly outs to a team's pitchers by *design*. Infield fly outs represent, in a fun-house mirror sort of way, the mirror opposite of home runs. Home runs are batted balls that by definition fielders *can't* reach, and so we assign "blame" for them to pitchers. Infield fly balls are batted balls that in the overwhelming number of cases fielders *always* reach. So shouldn't we assign "credit" for them to pitchers?

The problem, of course, is that traditional data doesn't tell us the number of infield fly outs allowed by individual pitchers. Right now, DRA leaves them as a team-level statistic (which is effectively how they're treated under UZR, I believe). As Voros and others have shown, *completely* ignoring BIP outcomes gives us pretty good value estimates for individual pitchers. Until more "Retrosheet" data (discussed in Part IV) is available that tracks infield fly outs per pitcher, I think that it might be reasonable, however, to allocate the team credit (blame) for infield fly outs among the team's pitchers, initially pro-rata, then adjusted up or down based upon each pitcher's non-homerun hits allowed ("Hits Allowed") relative to the other pitchers on his team. This is because per-pitcher Hits Allowed differences are probably, on average, largely explained by the relative tendency to give up infield fly balls. A great deal of data, reported at baseballprimer.com and elsewhere, shows that *total* fly balls are converted into outs at a significantly higher rate than total ground balls. Since infield fly outs are almost *always* caught, the out-conversion rates for *outfield* fly balls and infield ground balls might be very similar, on average, so that most of the difference in BIP out-conversion rates by pitchers is probably accounted for by infield fly balls. I readily admit that (i) differences in the relative abilities of an outfield and an infield could impact the relative Hits Allowed of a fly ball or ground ball pitcher on a given team and (ii) centerfielders such as Andruw Jones can play havoc with infield fly out numbers. For both of these reasons, I would report pitcher ratings showing the "stable" and "reliable" SO/BB/HR factors separately from infield fly out estimates.

To recap, we can't really measure, using traditional statistics, how many fly balls an infielder has caught (still less how many he caught that nobody else on his team could have caught) or how many infielder-catchable fly balls a pitcher has generated, but it makes sense to credit a team's pitchers (as a group) rather than infielders (as a group) for the impact of estimated infielder fly outs on the number of runs a team allows. Leaving infield fly outs as a team-level pitching statistic should generally not significantly distort pitcher ratings, and in time we may determine a logically and empirically defensible method for allocating infielder fly outs among a team's pitchers.

2. Responsibility for BIP Other Than Estimated Infield Fly Outs

Under DRA, fielders are assigned complete responsibility for all ball-in-play ("BIP") outcomes, other than estimated infield fly outs. In other words, the shortstop gets complete credit for each context-adjusted assist. Stated one more way, after accounting for how pitchers influence whether a ball is hit on the ground or in the air, and whether the ball is hit to the left or right side of the field, etc., etc., we attribute all credit or blame for plays made or not made (excluding infield fly outs) to the fielders. This approach implicates the still-raging debate about Voros McCracken's Defense Independent Pitching Statistics ("DIPS").

In a nutshell, Voros is generally credited with first observing that most pitchers have very little impact on whether BIP (at-bats not resulting in a strikeout, walk, hit batsman, or homerun) are converted into outs by their fielders or fall in as hits. Numerous analysts, including Voros, Bill James, MGL, TangoTiger, Tom Tippett of DM, and Dick Cramer have found, by a variety of different methods, that individual pitchers, with the rarest of exceptions, have not historically demonstrated an ability to cause the number of BIP they allow that are converted into outs over the course of a full season of pitching to differ by more than a handful of hits compared with a league-average pitcher. The entire controversy seems to be about whether Voros said that pitchers have "no" impact (he didn't, or if he ever did, he quickly corrected himself) and the extent to which those "handful" of hits per season per "exceptional" pitcher are worth thinking about. Some recent computer *simulations* by Arvin Hsu and Erik Allen at TangoTiger's Primate Studies page at baseballprimer.com suggest that full-time pitchers might have a larger impact: a pitcher who is one standard deviation better at BIP out conversion might save about seven hits over the course of a season, and the two-standard-deviation outlier might save 14.

My belief, based on the prior discussion and some more analysis provided below, is that (i) pitchers generally control, through their well-known tendency to generate ground balls or fly balls, the level of *infield* fly outs (the most *extreme* form of fly ball), which accounts for a significant portion of total variability in BIP out-conversion, team-by-team (consistent with the Hsu/Allen simulation results), and (ii) pitchers that generate more infield fly outs obviously generated more outfield fly balls than infield ground balls (and vice-versa), *but* (iii) out-conversion of infield ground balls and *outfield* fly balls is overwhelmingly controlled by fielders.

There is a good deal of indirect evidence to support ascribing responsibility for BIP outcomes (other than infield fly outs) to fielders.

First, *making* this assumption yields DRA ratings which match up well with UZR and DM ratings, which *do* take into account virtually *all* of the ways in which pitching staffs affect BIP. UZR factors in ball placement, speed, whether the ball was hit off a left- or right-handed pitcher, and whether the ball was a grounder, fly ball or line drive. I'm not aware of all of the factors DM considers in its "zone" model, but they are

probably similar to the factors included in UZR. If zone systems take every pitcher-controlled variable regarding BIP into account in rating fielders, and zone ratings basically match up with DRA ratings, which assume no pitcher impact, doesn't that strongly suggest that fielders control such BIP out-conversion rates, on average?

Second, zone data shows that individual fielders have a much greater measured impact on non-infield fly out BIP than pitchers. The number of BIP *allowed* by a starting *pitcher* per season probably exceeds the number of BIP *reaching* the "zones" of any particular full-time *fielder* per season, even at the most important fielding positions, such as shortstop. In other words, individual pitchers have more "opportunities" to affect BIP outcomes than individual fielders. Yet only a small number of the best pitchers in the history of the game have been demonstrated to have had a reliable effect on more than a handful of BIP outcomes per season, while zone data shows that many fielders every year have measured impacts of five-to-eight times that amount. DM reports that "in a typical season, the top fielders at each position make 25-30 more plays than the average. Exceptional fielders have posted marks as high as 40-60 net plays, but those are fairly uncommon. Recent examples include Darin Erstad in 2002, Scott Rolen just about every year, and Andruw Jones in his better seasons. The worst fielders tend to be in the minus 25-40 range." And though extremely high fielder ratings tend not to last for more than a few years, they have more year-to-year consistency than pitcher BIP outcomes, which brings me to the next point.

Third, the "persistency" of individual fielder ratings is *at least eight times* greater than individual pitcher ratings on BIP. Before proceeding with the UZR/DRA/DM comparison, I conducted a test to confirm that full-time fielders' UZR ratings in a given year actually predicted—to a statistically significant extent—their ratings in the following year. The absence of such "persistency" (an idea of Dick Cramer's) would indicate that fielding performance was random, not a skill. The percentage of fielders' plays made in any given year generally explained well more than 20% of the variation in their plays made in the following year. (In other words, when I regressed full-time Year 2 UZR runs for all full-time fielders onto full-time Year 1 UZR runs for the same fielders, the "r-squared" was generally greater than 20%, often *much* greater. The same held true for DRA.) Using a similar method, Dick Cramer recently conducted a study that showed that for full-time pitchers, the number of their Hits Allowed per BIP relative to the league average rate in any given year "explains" *at most* 2.6% of the variation in the number of their Hits Allowed in the following year. *See* Dick Cramer, "Preventing Base Hits", *The Baseball Research Journal*, Vol. 31, p. 89 (SABR 2003).

My hunch as to why some analysts do not allocate all BIP outcomes to fielders is that measures of *team*-level fielding impact on Hits Allowed show not that much more year-to-year persistency than individual pitcher Hits Allowed data. Dick's study suggested there might be about a 6.4% team-level fielding persistency, independent of park effects. *Id*. There is, however, a good reason why team fielding quality doesn't persist so much: a team's *fielders* don't "persist" so much. In collecting data for the 1999-2001 UZR/DRA comparison, I discovered yet again how hard it is for players to keep their jobs from year-to-year. On average, only two, maybe three players played

full-time at the same position for the same team for two successive seasons. Due to fielder turnover, the most revealing comparison, for determining the relative impact of pitchers and fielders on BIP, is *individual pitchers who pitch full-time* and *individual fielders who field full-time*. Thanks to Dick's study, we have a reliable measure of the *average* persistency of the effect of individual full-time pitchers on Hits Allowed. Thanks to MGL's UZR data, we have a reliable measure of the *average* persistency of the effect of individual full-time fielders on Hits Allowed. Fielders beat pitchers by at least eight to one, *even though responsibility for infield fly outs is effectively "ceded" to pitchers* under UZR and Dick's persistency tests.

Eventually, UZR and similar systems will provide a definitive answer regarding the relative impact of pitchers and fielders on Hits Allowed. How? By enabling us to perform an exact comparison of the average scale and persistency of *pitcher* "zone" ratings ("PZR") with the average scale and persistency of *fielder* UZR ratings. PZR can track for each pitcher the characteristics of all non-home run batted balls allowed, e.g., the actual rate of infield fly balls, ground balls, outfield fly balls, line drives, etc. given up by the pitcher, as well as the distribution of such BIP on the field, and assign run-expectancy values on such BIP *independent* of the effects of the pitcher's park and fielders. In other words, just as UZR rates fielders "independently" of pitchers and parks, PZR can rate pitcher BIP outcomes independently of fielders and parks.

What I am highly confident we will find is that (i) the variation between pitchers in average PZR ratings (the average "scale" of impact, or how much difference there is between pitchers on average) is almost completely explained by the variation in infield fly balls generated, (ii) the scale/variation in PZR ratings *excluding infield fly balls* will be *very* much less than that for UZR fielder ratings, and (iii) the year-to-year persistency of PZR ratings excluding infield fly balls will be *very* much less than that of UZR ratings.

To summarize, we'll determine that BIP outcomes are explained by

(i) the number of pitcher-generated, nearly-always-infielder-catchable fly balls (i.e., short, high flies that are "always" caught),

(ii) pitcher influence over how hard ground balls ("dribbler" or scorching one-hopper?) and outfield fly balls (high fly or "rope"?) are hit,

(iii) fielder ability,

(iv) fielder *turnover*,

(v) park effects (particularly on infield pop-ups and long outfield drives), and

(vi) luck.

I'm fairly certain that (iii) and (iv) will be found to be *at least* five times more important than (ii), and thereby justify the simplifying assumption under DRA that fielders are fully

responsible for their context-adjusted plays made, i.e., all BIP outcomes other than estimated infield fly outs. In the meantime, the fact that DRA ratings match up well with UZR ratings reassures me that this is the correct approach.

One intriguing finding under DRA: the number of runs saved/allowed per team on infield fly outs had a standard deviation in the 1974-2001 data set of 27, whereas the standard deviation in runs saved/allowed for all other BIP was 39. The ratio of such standard deviations (approximately 3:4) is very close to the ratio that Hsu and Allen have found in their simulation models between pitcher and fielder impact on BIP.

3. Out Values of BIP Outs

In the course of refining the methodology for determining the number of context-adjusted plays made at various positions, I was presented with two approaches, each to some degree equally compelling from a theoretical perspective. One would yield significantly different out values (as distinguished from "hit-prevented" values) for infield v. outfield plays. I chose the other approach. Aside from the common sense behind such approach, it yielded a better UZR match. Again, I'm only belaboring the point in case years from now somebody wants to say that I was being misleading in saying that DRA has no subjective weights or factors.

4. Wild Pitches—Passed Balls—Balks

DRA assigns responsibility for passed balls ("PB") catchers, but also responsibility for wild pitches ("WP") and balks ("BK"); provided, however, that an adjustment (admittedly *ad hoc*) is made for knuckleball pitching.

Why should *catchers* be responsible for WP and BK?

Because such an approach yields ratings that best match up with what I consider to be the best catcher rating system, albeit one that requires play-by-play Retrosheet data.

TangoTiger has recently put together, using Retrosheet data, a brilliant analysis of the rates at which WP, PB, and BK (as well as SB, CS, Picked Off) occur for each pitcher/catcher pairing for catchers who played a significant amount of time in the 1970s and 1980s. The data permits a calculation of the rate at which catchers "allow" WP or PB per pitcher, compared with all other catchers who caught for that pitcher anytime in that pitcher's career. The data also permit a calculation of the rate at which pitchers "allow" WP or PB per catcher, compared with all other pitchers who pitched to that catcher. Although the data show that WP and PB are clearly more controlled by individual pitchers than catchers, catchers *do* have a non-trivial impact on WP and PB. When I "credited" catchers with WP and PB (and BK), the catcher DRA ratings generally matched up better with Tango's ratings.

Here's another way of thinking about the issue. Individual pitchers absolutely control WP more than catchers; indeed, they control PB more than catchers, on a "rate"

basis. However, even dominant starting pitchers pitch only one-fourth as many innings per-season as a full-time catcher catches. Furthermore, *team*-level performance is almost certainly more controlled by the catcher (and his small number of backups) than the pitchers, as the "mix" of pitchers should, on average, randomize away their individual impact. For purposes of making career assessments of catchers, this effect is even more powerful.

Balks are included in the catcher rating because I lumped them together with WP and PB in the regression analyses (for fairly obvious reasons), and it just isn't worth the trouble to take them out. Tango has concluded they're basically random occurrences.

II. 1999-2001 DRA-UZR-DM COMPARISON

In this Part I'll provide and review the results of the DRA-UZR-DM analysis for 1999-2001, the time period for which I had data for both DRA and UZR.

To "test" DRA, I wanted to compare how it performed against the state-of-the-art zone systems: UZR and DM. Thanks to MGL and TangoTiger, I was able to obtain exact UZR runs saved (allowed) ratings for each player, position and team. DM does not provide explicit runs saved (allowed) ratings for players, but does usually provide a description of a player's fielding ability—at least if it is notably good or bad—on its website, either under its annual "Gold Glove" review or its annual team reviews.

In subpart A, I'll describe how the test was put together; in subpart B, I'll provide and discuss the results, position-by-position; in subpart C, I'll summarize the results.

A. Whys and Wherefores of the DRA-UZR-DM Comparison

1. Test Population

The test I decided upon was a comparison of the average of the full-time ratings for every player who played at least two seasons full-time (130 or more games, or at least 80% of a full season) for one team per season (including Team A in Year 1 and Team B in Year 2) during the three-year time period (1999-2001) for which I had both UZR and DRA ratings. The total population of players satisfying these conditions was 82, plus 4 catchers. As I did not have access to UZR catcher ratings for 1999-2001, I will provide career DRA ratings for the two "extremes" in catcher ability, as observed by fans and quantified under UZR in the past: Ivan Rodriguez and Mike Piazza.

2. Why Two or Three Full-time Seasons?

As we all know, traditional data is very incomplete. What I wanted to find out was whether, given a sufficiently large sample of data, most of the contextual factors that are beyond the reach of traditional statistics would "average out", so that an intelligently designed system based on old-fashioned stats could—given a sufficient sample size—

approach the accuracy of "perfect" systems based upon new-fangled data. (Even UZR ratings for less-than-full-time seasons are subject to high variance.)

On the other hand, I didn't want the test to rely upon *too* many seasons of data. The most likely practical application of DRA may be in the evaluation of minor league players, for whom zone data is unavailable. If a system based upon readily available statistics required four or five seasons of data to be reliable, it wouldn't be of any use for evaluating such players, because it would take too long to find out if any given fielder was any good.

I don't know how many "minor-league" games the average major-leaguer these days has played, but I would be surprised if there are many players who have played fewer than 250 games in Division I college ball and/or organized minor-league ball. And the ones who are quickly promoted to The Show are usually so good that the fine points of their fielding ability are probably not especially important.

3. Assigning the Team DRA Rating to the Individual Player

For purposes of the DRA-UZR-DM analysis, I will equate an individual full-time player's DRA rating with the DRA rating for his *team* at the position he played. There are empirical, theoretical and practical reasons for doing this.

First, even if I had exact innings fielded data and used it to "individuate" the ratings, DRA ratings wouldn't necessarily be better for full-time players. In some data samples, I used the 2000-2001 baseball-reference.com innings fielded data to see if I could get a more accurate match with UZR. The results were inconclusive—sometimes slightly better, sometimes slightly worse. There are very good explanations why this happened. The 130-game requirement is a cut-off point—but the *average* number of games played per full-time season in the test is probably closer to 146, or approximately 90%. If a player plays that many games, there remains a significant risk that the contextual factors in the *small* sample of games the player *didn't* play in won't be representative, just based on *randomness*–so if you try to "individuate" the rating of the *part*-time fielder by pro-rating the team rating and adjusted it up or down based on the part-time players' plays made per inning fielded relative to the team average, the part-time players' rating (and, therefore, the full-time player rating, which is just the team rating minus the part-time players' rating) will be off. Think about it another way—how much impact could an exceptionally good or bad back-up fielder have on a team's rating in 16 games? Or even 32? The *maximum* in remotely normal cases should be +/- a handful of runs per season.

Second, the UZR team/position/player ratings demonstrate that, at least for two- or three-year average ratings for players who played 130+ games per season, there is no meaningful difference between the team and the individual player *UZR* ratings. None. (In Part II.C, I will include the team UZR rating as well as the individual UZR ratings to show this.)

Third, by demonstrating in the 1999-2001 study that the impact of part-time fielders always randomizes away within two or three years, I hope to reassure you that the career ratings covering *five* or *more* seasons (provided in the historical review in Part III) are basically reliable, even though they *also* assign the team rating to the individual who played at least 130 games for the team at the position.

If I ever do a comprehensive Win Shares-type project with the help of a computer programmer, I would use innings fielded data and the Bill James method for estimating innings where we lack the data, and the resulting estimated DRA per-player ratings would probably be reasonably good for part-time players (with one possible exception described in Part IV.C) and, to the extent they were off, the smaller number of games would not yield any substantial ratings distortions on a career basis. I just don't have the time or resources right now. For purposes of evaluating the truly significant fielders throughout history, the current method is, I believe, more than adequate, with the rare exception of a Gil McDougald or Jim Gilliam, who played outstanding defense at a number of positions part-time each season.

4. Role for DM Information

UZR is almost certainly the most accurate, most reliable fielding evaluation system that provides actual runs-saved estimates for today's fielders that any fan can look up. MGL's work on UZR has dramatically improved not only our appreciation of who the good and bad fielders are; it has also provided all sorts of new insights on how various contextual factors can impact fielding.

That being said, there have been times in the past when UZR needed to be fine-tuned, and we shall see that a few of the latest ratings seem to be a little off. That's why I decided to consult DM to resolve differences between UZR and DRA.

Before making the initial comparison between UZR and DRA, I ran a test to make sure that UZR was reliable. By running regressions of paired successive full-time player seasons, I could perform a "persistency" test at each position similar to the test Dick Cramer performed for Hits Allowed by pitchers (see Part I.D.2 above). Studies I had done revealed that year-to-year persistency under UZR was most reliable for full-time season ratings. (Yet another reason I have limited the study to full-time fielders: regressions that included UZR "rate" stats (runs per 162 UZR games) for less-than-full-time seasons were less reliable, for the simple reason that they included player seasons with smaller "samples" of performance.)

A similar persistency test I had run on UZR last year revealed that it wasn't yielding persistent ratings for full-time players, thus suggesting it wasn't capturing fielder skill. Mike Emeigh's fielding articles, along with results I was obtaining under DRA, gave me some ideas for fine-tuning UZR, and as mentioned in the Acknowledgments section, MGL graciously incorporated the suggestions. Recently updated UZR ratings were *strongly* persistent in the 1999-2001 sample, except at third base, right field and, possibly, shortstop. On the one hand, that might only be because the number of paired

seasons was low at right field and third. (At short, the ratings were persistent if I excluded the two players who played two *non*-successive seasons full-time seasons: 1999 and 2001, but not 2000.) On the other hand, as we shall see, there seem to be a few UZR ratings that are probably incorrect. UZR appears to generate slightly more extreme ratings than DRA or DM.

I'm *not*—repeat, *not*—"knocking" UZR; I *can't*, because I'm *relying* on it to give us the best evidence that DRA works. It's just that UZR requires a huge amount of data and calculation, much of which cannot be easily checked, and little of which is publicly available, and a minor glitch or two *might* remain in the system.

There are all sorts of subtle factors to consider when using zone data, which I won't discuss here. (See Part IV.A.) Suffice it to say that with high-quality zone data, we have "perfect" and "complete" information, but it's sometimes difficult to know precisely what to do with it. It is probably the case that whatever differences there are between UZR and DM evaluations probably arise from how all the variables tracked in "zone" data (ball speed, left- or right-handed pitching, park effects, zones "shared" by more than one fielder, plays made in unusual zones, sub-optimal positioning due to the presence of baserunners, etc., etc.) are factored in.

As MGL himself has said:

> "Keep in mind that UZR is far from a "perfect" metric - IOW, it does not reflect a player's defensive contribution to his team perfectly reliably and accurately. This goes without saying and is true for all metrics. However, for various reasons, I (and other people) am confident that it is a very good metric and is probably the best defensive metric out there, by far and away (DM and maybe some others have a similar metric, but they do not share their exact methodology, so I am excepting them from this statement—I am mainly referring to things like ZR, RF, etc.). That doesn't mean that it doesn't make lots of small mistakes, a few big mistakes, and 1 or 2 gigantic mistakes. All metrics will do that by virtue of random fluctuation alone. Can we identify those mistakes? Probably not. Can we assume that what it suggests about a player's defensive value is true ands accurate? Yes! That's the whole point of a good (not perfect) metric! To augment or even supplant (that's the dirty word in [the] scouting/sabermetric debate) what we think we know about a player. To borrow (and paraphrase) an axiom from Bill James (I think), if we are not shocked by some percentage of a certain metric, the metric has no use! That should be intuitively obvious! If we look at a metric and go through it and say "Nah that doesn't look right (A. Jones in 2002, Torii Hunter, J.T. Snow), I'll reject that one (or even, worse, I'll reject the whole metric)," or "Yeah that's what I thought, I'll accept that one," what would be the point of the metric? You would [/] need only go with whatever you thought in the first place. A rule of thumb is that if a metric is good, you should be surprised at maybe 10 or 20% of the results and very satisfied with the rest. I think that UZR falls nicely into that framework, although I've never gone through any of the files and calculated the percentage of

"surprises". Now, that 10 or 20% will vary depending upon the sample size of the ratings you are looking at of course. The smaller the sample size, the more there should be surprises. Also, even if it is a good metric, that doesn't mean that for all of those surprises the metric is correct and your presumption was not. But that's another topic altogether (using Bayesian probability), which I won't get into right now." MGL, e-mail comment to his article, *Super Linear Weights, 2000-2002*, baseballprimer.com, June 4, 2003.

MGL brings up an extremely important point: if a metric doesn't confound our expectations at least some of the time, it's useless. There will be a few such surprises as we survey the DRA-UZR-DM information for 1999-2001, as well as some of the career DRA ratings for 1974-2001. To be clear, however, I *won't* "second guess" UZR based on subjective preconceptions. Frankly, I *have* no subjective impressions about most of the fielders in the 1999-2001 survey. I'm just consulting one well thought-out, empirically supported, publicly available set of defensive evaluations, the content of which I must accept on faith (DM), in order to evaluate surprising results under *another* well thought-out, empirically supported, publicly available set of defensive ratings, the content of which I must accept on faith (UZR).

As I'll discuss in Part IV, DRA could have a valuable role to play in helping analysts identify and diagnose the "mistakes" that even MGL acknowledges above are possible in zone systems, including exceptionally well-designed systems such as his.

5. Coding Errors

I have done my utmost to double-check all reported results, especially the results reported by others (MGL and DM). As a great deal of the information analyzed here had to be selected by hand and input by hand, there are no doubt some coding errors. I welcome any corrections any reader has to offer.

OK. Enough explanations and caveats. Let's go through the results.

B. 1999-2001 DRA-UZR-DM Ratings, Position-By-Position

I will go through the positions in descending order of skill/importance, what Bill James long ago described as the Defensive Spectrum: shortstop, second base, center field, third base, right field, left field and first base. I will end with a review of the career DRA ratings (through 2001) for I-Rod and Piazza, the best and worst fielding full-time catchers over the past decade or so, as I lack the most up-to-date UZR ratings at catcher.

UZR infielder ratings include "DP" ratings; UZR outfielder ratings do *not* include "Arm" ratings. In Part III, in the context of the discussion of historical outfielder ratings from 1974-2001, I will discuss DRA arm ratings in the outfield. All numerical ratings are denominated in terms of runs saved or allowed relative to a league-average fielder; e.g., +25 means 25 runs "saved"; –12 means 12 runs "allowed".

The "Notes" column addresses examples where DRA and UZR seem to be reaching meaningfully different results. The "test" I decided to apply for consulting DM was as follows. If both the DRA and UZR ratings were in the +/- 6 runs "band" around zero (i.e., both systems rated the player as "average"), I left them alone. If either rating was outside that band *and* if the two ratings differed by *more* than 10 runs, I consulted DM.

The following "code" of comments applies: "dm=dra" means that DM information strongly supports DRA; "dm~dra" means that DM information is mixed, but on balance, appears to favor DRA over UZR; "?" indicates it is unclear whether DM supports DRA or UZR; "dm~uzr" means that DM information is mixed, but on balance, appears to favor UZR; "dm=uzr" means that DM information strongly supports UZR. The one reference to "dial=dra" refers to an instance in which Chris Dial's zone rating matches better with DRA. The one reference to "park" refers to a (Fenway) park effect.

DM commentary comes from three separate sources: *team* essays for 1999 and 2000, which contain capsule summaries of individual player performance, and the "Gold Glove" essay for 2001. DM's webpage does not provide team comments for 2001 or Gold Glove essays for 1999 and 2000. In general, this mix of essays generally does not provide commentary for average or below-average fielders in 2001. The team comments for 1999 and 2000 more than make up for the lack of "Gold Glove" essays for those years.

1. Shortstop

	'99		'00		'01		Avg		Notes
Player	UZR	DRA	UZR	DRA	UZR	DRA	UZR	DRA	
Rich Aurilia	-2	2	-3	0	17	0	4	1	
Royce Clayton	1	4	-7	-16	20	8	5	-1	
Derek Jeter	-17	-21	-10	-29	-19	-26	-15	-25	
Edgar Renteria	1	-10	11	12	2	18	5	7	
Miguel Tejada	13	5	5	-4	0	0	6	0	
Omar Vizquel	10	-8	1	-18	10	-4	7	-10	dm=dra
Deivi Cruz	3	11	-7	3			-2	7	
Nomar Garciaparra	-7	1	-4	9			-6	5	
Chris Guzman	-18	-8	-21	-12			-20	-10	
Neifi Perez	3	-2	18	18			11	8	
Rey Sanchez	31	21	20	22			26	22	
Alex S. Gonzalez			-9	-12	18	15	5	2	
Alex Rodriguez			18	9	8	-6	13	2	?
Alex Gonzalez	-17	-13			-3	-4	-10	-9	
Rey Ordonez	39	10			-6	-2	17	4	dm=dra

DRA and UZR basically agree that Aurilia, Clayton, Renteria, Tejada, Deivi Cruz, Nomar, and Alex S. Gonzalez were basically average over the '99-'01 period.

DRA and UZR agree that Rey Sanchez was outstanding and that Neifi Perez was pretty good, at least in 2000. DRA and UZR basically agree that Jeter, Guzman and Alex Gonzalez were clearly below average, with the DRA ratings for Guzman being less extreme and the UZR rating for Jeter being less extreme. We could quibble about a few single season ratings—UZR shows Clayton as a viable Gold Glove candidate in 2001; DM's 2001 Gold Glove Review ("DM GG") does not mention Clayton. On the other hand, DRA shows Renteria as a viable Gold Glove candidate in 2001, and DM GG doesn't mention him either.

The significant differences are over Vizquel, Ordonez, and, possibly, A-Rod. DM GG had this to say about Omar, my nomination for the most over-rated fielder in history:

> "[Vizquel] was one of three Cleveland infielders to be rewarded with Gold Gloves [in 2000]. But that infield was below the league average in turning ground balls into outs. And according to the STATS Major League Handbook, they were fourth *worst* in the league in converting double plays when grounders were hit in double-play situations.
>
> "The bottom line is that somebody isn't making nearly as many plays as people think . . . .
>
> "[In 2001], Cleveland's infield was 13th in the league in the percentage of ground balls turned into outs. And they were only a hair above the league average in double-play percentage.
>
> "You could argue that the infield looks bad because the corner guys -- Jim Thome at first, Travis Fryman and Russ Branyan at third -- don't cover much ground, and you'd be correct. Problem is, there's absolutely no evidence that their middle infielders are doing more than their share, either . . . .
>
> "Suffice it to say that Vizquel's range wasn't all that good this year."

UZR rates Vizquel above average; DRA rates him below average.

Rey Ordonez has a historically high UZR rating for 1999: +39. DM does not seem to suggest that Ordonez was having a historically outstanding season at short. "Error totals aren't usually a good indication of fielding prowess, but the four errors charged against Ordonez were impressive nonetheless." DM says nothing about his range. DRA rates Ordonez's 1999 season at +10.

Regarding A-Rod, DM seems to take a middle position between the moderately high rating he has under UZR and the barely above average rating he has under DRA. In 2000, DM's team comment for Seattle describes A-Rod's fielding in a manner that supports DRA: "While A-Rod lacks the great range of some other AL shortstops, he does rate above-average and has very good hands." UZR rates A-Rod's 2000 season at +18; DRA rates it +9. DM has nothing to say about A-Rod in 2001 in its Gold Glove report.

UZR rates him slightly above average (+8); DRA rates him as slightly below average (-6).

All in all, DRA appears to have "worked" in evaluating full-time shortstops during the 1999-2001 period.

2. Second Base

	'99		'00		'01		Avg		Notes
Player	UZR	DRA	UZR	DRA	UZR	DRA	UZR	DRA	
Roberto Alomar	5	3	-5	-5	-11	-4	-4	-2	
Ray Durham	-10	-2	10	11	22	12	7	7	
Jeff Kent	3	-8	-7	-6	17	23	4	3	
Edgar Alfonzo	5	2	4	8			5	5	
Jay Bell	0	-12	-9	-1			-5	-7	
Warren Morris	-7	0	-9	3			-8	2	
Pokey Reese	31	38	21	24			26	31	
Luis Castillo			6	-15	8	3	7	-6	dm~uzr
Mark Grudzielanek			-7	0	7	-2	0	-1	
Adam Kennedy			10	0	15	15	13	8	
Eric Young			-1	3	-5	-3	-3	0	
Damian Easley	-7	5			-11	1	-9	3	dm=dra

DRA and UZR basically agree at second base. In its 2000 Florida Marlins comment, DM classifies Luis Castillo among young players with "great speed and defense", so it's probably the case that UZR has measured his fielding better than DRA, though DM does not elaborate at all regarding Castillo's defense, and does not mention Castillo at all in its 2001 Gold Glove review. If DRA has failed to recognize his talent, it's not a talent of significant magnitude. DRA, UZR and DM all agree that Pokey Reese was outstanding and that Adam Kennedy was very good, particularly in 2001.

DM says in 1999 of Damian Easley, "In the field, he played well at second and even was pressed into emergency duty at shortstop, where he acquitted himself admirably." UZR rates him –7; DRA +5. In 2001, DM does not mention Damian in its Gold Glove review. UZR rates him –11; DRA +1.

3. Center Field

	'99		'00		'01		Avg		Notes
Player	UZR	DRA	UZR	DRA	UZR	DRA	UZR	DRA	
Mike Cameron	29	9	6	5	25	19	20	11	
Steve Finley	16	7	-19	-16	1	-2	-1	-4	
Doug Glanville	-1	7	-31	2	-16	15	-16	8	dm=dra
Andruw Jones	53	43	19	21	30	29	34	31	
Bernie Williams	-23	-2	-5	-10	0	-7	-9	-6	
Ken Griffey, Jr.	-19	-8	10	7			-5	-1	
Marquis Grissom	-17	-11	-26	-3			-22	-7	dm=dra
Ruben Rivera	21	13	37	3			29	8	dm~dra

Jose Cruz, Jr.			-28	-16	-31	-18	-30	-17	dm=dra
Jim Edmonds			4	0	-21	-8	-9	-4	
Kenny Lofton			27	20	20	-2	24	9	dm~uzr
Carlos Beltran	-8	-17			4	13	-2	-2	

There are some differences here, but mostly in terms of magnitude, not direction, and it would appear that DM commentary is more consistent with DRA ratings. As I mentioned before, the UZR and DRA ratings shown are just for catching fly balls, not for throwing out baserunners. (I'll discuss outfielder arm ratings in Part III.)

I think it is pretty clear, both by reference to DM and from indirect evidence, that the DRA rating for Glanville is probably more accurate than the UZR rating. According to the DM 2000 report for the Phillies:

> "Glanville's speed is still his best asset. He is a very good base stealer and has been a strong performer in center field, though 2000 was far from his best year defensively. Glanville made just four errors in 2000 and threw out nine runners."

UZR rates Glanville at –1 in 1999 and –31 in 2000; DRA rates Glanville at +7 in 1999 and +2 in 2000. Glanville stole over 90 bases during the three-year period and had a phenomenal 85% success rate. True, good base-stealers are not always good outfielders, as Lou Brock and Tim Raines have demonstrated. But it seems very unlikely that a major league centerfielder with two years of full-time experience as a "strong" center fielder, as well as outstanding speed and baserunning judgment, could have a –31 rating.

In the case of Grissom, Rivera, Cruz and Lofton, DRA and UZR agree in terms of direction; but I think it fair to say that DM generally agrees more with DRA regarding magnitude.

For 1999-2000, DM indicates that Grissom's fielding is clearly in decline, but DM gives no indication that it is *very* bad: "[In 1999,] Grissom's range in center field [wa]sn't what it used to be, so he d[id]n't make up for his weak bat with defense." UZR rates him at –17; DRA rates him at –11. DM indicates that "the 2000 season was yet another step down the ladder for Grissom", but gives no indication regarding his *fielding*. UZR rates him –26 (*notably* terrible); DRA rates him at –3. Grissom's two-year average UZR rating is one of the worst at *any* position during the time period; DM seems to say that Grissom was below average, but not *that* bad. His two-year average DRA rating is –7.

For 1999, DM has this to say about Ruben Rivera: "He showed . . . a strong arm and above-average range in center field." UZR rates Rivera at +21; DRA rates him +13. For 2000, DM ratchets up the praise for Ruben's fielding: "He . . . plays a superb center field . . . ." UZR rates Ruben at +37; DRA rates +3. The DM description would put him somewhere between those ratings. Over the course of the two-year period, according to DM, Ruben was good/superb, not excellent/outstanding/Gold Glove material. The two-year average rating under UZR is far above a Gold Glove standard (+29); the DRA rating is solid/above average (+9).

Jose Cruz, Jr.'s range is considered by DM in 2000 to be not "quite good enough for center field"; but DM gives no indication in 2000 or 2001 that Cruz is the next Greg Luzinski. His UZR ratings are -28 and -31; his DRA ratings are –16 and –18.

Kenny Lofton attracts no comment for his defense in the DM commentary on the 2000 season, but he is acknowledged as the second-best AL outfielder in 2001, so perhaps it is fair to say that the UZR rating is slightly more consistent with DM than the DRA rating.

Although UZR and DRA ratings for Andruw Jones are in almost perfect agreement, I think both systems overstate his value. Infield fly outs for the Braves have been freakishly low during the seasons in which Andruw's putouts have been freakishly high, thus suggesting that he was taking a lot of short fly ball chances that would have been caught by middle infielders anyway. I have serious doubts as to whether Andruw has really been contributing *that* much to Atlanta's success in the field.

4. Third Base

	'99		'00		'01		Avg		Notes
Player	**UZR**	**DRA**	**UZR**	**DRA**	**UZR**	**DRA**	**UZR**	**DRA**	
Jeff Cirillo	10	-4	3	-3	35	4	16	-1	?
Troy Glaus	-4	1	7	20	-21	-10	-6	4	
Chipper Jones	-10	-26	-1	-3	-7	-16	-6	-15	dial=dra
Joe Randa	0	9	1	3	0	0	0	4	
Robin Ventura	24	19	0	2	15	2	13	8	
Scott Brosius	18	4	11	-4			15	0	dm=uzr
Eric Chavez			-3	-4	26	10	12	3	dm=uzr
Corey Koskie			3	-8	19	8	11	0	dm=uzr
Mike Lowell			-15	0	4	4	-6	2	
Phil Nevin			-3	-4	-4	5	-4	1	

If DM is any guide, neither DRA nor UZR covered itself in glory at third. As I mentioned above, UZR ratings for successive seasons were not persistent to a statistically significant extent. Neither were the DRA ratings. In general, DM matches up better with UZR than DRA, and it's probably fair to say that third base is the one position that DRA has difficulty evaluating in the 1999-2001 study, though the historical ratings in Part III seem to be correct. For reasons I'll keep under my hat for now, careful application of DRA methods to (non-zone) Retrosheet data that goes beyond traditional statistics would probably enable DRA to handle third significantly better.

The only third baseman of the 1999-2001 period that anyone will remember for their fielding ten years from now are probably Robin Ventura, Scott Brosius, Scott Rolen, and Eric Chavez. The average ratings under UZR and DRA for Ventura are basically consistent. One of the undeniable errors under DRA is its failure to give Scott Brosius a clearly above-average rating. I'm not impressed *per se* by his having won a "real" Gold Glove in 1999, but DM clearly recognizes his value, and UZR got it right. Scott Rolen is not shown in the chart, because he played fewer than 130 games at third in 1999 and

2000. DM praises Rolen in 2001 as "amazing"; DRA rates him at +16 and UZR at +32, so UZR probably "got" him better, but DRA wasn't far off. The Chavez UZR average rating is closer than DRA to what DM had in mind (pun intended) in characterizing him as providing good defense in 2000 and as being a "strong candidate" for a DM "Gold Glove" in 2001.

Chipper Jones' ratings didn't quite "trigger" the DM test, but I thought they were worth a second look. The DRA ratings are more severely negative than UZR's. Chris Dial's 1999 and 2000 zone rating for Chipper match DRA almost exactly. Perhaps the DRA rating in 2001 is too low, but if you were the Atlanta Braves, and had a third baseman who hit like Chipper and who only cost you a handful of runs at third, would you really send him out to left field? I'm not sure DRA is clearly right about Chipper, but the evidence of Chris Dial's zone rating and Atlanta's personnel move suggests that DRA is not clearly wrong.

The other discrepancies in average ratings are attributable to the 2001 ratings for Cirillo and Koskie. Scott Rolen is the reference point for Cirillo, according to DM: "Scott Rolen's . . . . closest rivals in [2001] were Robin Ventura and Jeff Cirillo." For 2001, DRA rates Rolen at +16 and Cirillo at +4. UZR rates Rolen at +32 and Cirillo at +35. Now Jeff Cirillo is shown in the DRA historical survey in Part III to have been a consistently good to average third baseman, but it seems difficult to believe that at age 31 he saved 35 runs while playing in only 80% of his team's innings. DM says that Koskie's defense in 2000 was "very good" and that he was a runner-up for their AL "Gold Glove" in 2001, so DM agrees with UZR, not DRA.

5. Right Field

	'99		'00		'01		Avg		Notes
Player	**UZR**	**DRA**	**UZR**	**DRA**	**UZR**	**DRA**	**UZR**	**DRA**	
Bobby Abreu	-22	-19	9	0	-20	-11	-11	-10	
Shawn Green	27	21	-18	-3	-12	5	-1	8	
Vlad Guerrero	16	7	15	0	17	0	16	2	dm=uzr
Brian Jordan	6	3	6	17	14	19	9	13	
Paul O'Neill	-2	3	5	13	-13	-14	-3	1	
Magglio Ordonez	21	-4	-17	-13	-12	-4	-3	-7	
Sammy Sosa	10	-4	8	6	3	15	7	6	
Jermaine Dye	34	18	26	7			30	13	dm=dra
Jeromy Burnitz			5	1	-8	7	-2	4	
Raul Mondesi	-21	13			-9	2	-15	8	dm~dra

In right field, DRA is clearly wrong in failing to identify Vlad Guerrero as a very good to excellent fielder. In 1999, Diamond Mind described Guerrero as having "plus range" (presumably, better-than-average but not exceptional); in 2000, DM described his range as "very good"; in 2001, DM described it as "great", and argues that Guerrero should have been the Gold Glove winner in right field.

DM says nothing about Jermaine Dye's defense in 1999 or 2000. The UZR ratings are spectacular, far above a Gold Glove standard; the DRA ratings above average.

Regarding Raul Mondesi, DM seems to agree more with DRA than with UZR. UZR rates Mondesi as *very* poor in 1999 (-21) , but DM has nothing negative to say about his range; indeed, DM highlights the fact that in his prior season he had played *center* field, thus suggesting that he must have had good range for a right fielder:

> "I lost some respect for Mondesi in 1998 when he complained about being moved from RF to CF when Gary Sheffield was acquired from Florida. Seems he preferred RF because he won two Gold Gloves at that position. There's no question that it was the right move for the team, but Mondesi fought it every step of the way. Now he's Toronto's problem, as he was the key player in the trade for Shawn Green after the season."

DRA rates him at +13 in 1999, which would appear more consistent with DM.

In 2001, DM said:

> "[Mondesi h]as a very good reputation for defense, but that's mostly based on his great arm. In terms of range, our analysis shows that he's been slightly above average throughout his career. In the spring, it was reported that Mondesi came to camp carrying some extra weight, and his defensive numbers took a big dive. Coincidence? Maybe, but we felt a Fair rating was an accurate reflection of his 2001 performance. He could easily rebound next year.

DRA rates him 11 runs lower in 2001 than in 1999 (+2; down from +13), but the UZR rating (–9) might be more consistent with a "big dive" in performance. It's close, but my reading of DM favors DRA slightly more than UZR.

6. Left Field

	'99		'00		'01		Avg		Notes
Player	**UZR**	**DRA**	**UZR**	**DRA**	**UZR**	**DRA**	**UZR**	**DRA**	
Luis Gonzalez	17	-1	15	10	15	6	16	5	dm~dra
Gary Sheffield	-24	-7	-18	0	-17	-29	-20	-12	
Shannon Stewart	-6	-15	8	1	-13	1	-4	-4	
Ben Grieve	-24	-6	-26	-25			-25	-16	
Troy O'Leary	15	-6	9	2			12	-2	park
Barry Bonds			8	5	-3	-13	3	-4	
Bobby Higginson			19	6	28	12	24	9	dm~dra
Carlos Lee			-15	-5	4	-7	-6	-6	

UZR ratings for Luis Gonzalez suggest that he should have been a perennial contender for a Gold Glove in 1999-2001. DM says in its 1999 team comments that he "provides very good defense in left field" (the tone of the observation suggests it is based upon his historical ability; DM does not provide any specifics regarding that year). For

2000, when Gonzalez was 32 years of age, DM provides no description of defense. He is not mentioned at all in the 2001 DM GG review. I find that commentary consistent with slightly above-average fielding performance by an outfielder in his early 30s, but inconsistent with borderline Gold Glove performance. As we shall see in Part III.F, Gonzalez was a superb fielder in his twenties.

DM has not a word to say about Troy O'Leary's defense. From that I would infer that he was nothing special in the field. The DRA rating might be a little low, due to the park effect of Fenway. Over the course of 1974-2001, the average Bosox left field rating was –8 runs. MGL has found a significant park effect in Fenway's left field, which is so small that gross chances are reduced. MGL's published findings suggest that fielding park factors are generally fairly minimal, except in Coors and (left field in) Fenway. To keep DRA simple, I have not included any park factors, preferring instead to pay special attention to the clearly significant cases. O'Leary might be one of them.

DM expressly praises Higginson's arm in 2000, but says nothing about his range. UZR rates his range as potential Gold Glove-material; DRA rates him as slightly above average. In 2001, DM's GG review praises "Bobby Higginson and Jacque Jones [as] the two left fielders who separated themselves from the pack." That sounds like a "solid" Gold Glove performance, but not borderline-historical performance. UZR rates him +28; DRA rates him +12. Overall, I read DM as closer to DRA than UZR, but that might be just me

7. First Base

	'99		'00		'01		Avg		Notes
Player	UZR	DRA	UZR	DRA	UZR	DRA	UZR	DRA	
Jeff Bagwell	3	3	9	-2	11	17	8	6	
Carlos Delgado	-3	-2	-3	-9	-8	-4	-5	-5	
Mark Grace	6	-5	6	7	-9	-2	1	0	
Todd Helton	2	0	8	20	23	15	11	12	
Tino Martinez	9	11	11	-1	12	5	11	5	
John Olerud	17	7	9	19	-1	9	8	12	
J T Snow	-8	14	-13	-3			-11	6	dm=dra
Eric Karros	8	12	4	13			6	13	
Ryan Klesko			-6	-1	-4	-7	-5	-4	
Derrek Lee			-15	5	0	7	-8	6	?
Todd Zeile			20	7	6	13	13	10	
Sean Casey	-5	-18			-7	-10	-6	-14	
Jason Giambi	-11	-19			5	-2	-3	-11	
Lee Stevens	2	-11			-6	-3	-2	-7	
Kevin Young	11	-3			17	-4	14	-4	dm=uzr

I'm not going to spend a lot of time with first base ratings. They basically match up. DRA clearly got Kevin Young wrong. The DM 1999 commentary notes that J. T. Snow has won many Gold Gloves, and says nothing to contradict the commonly held view that he is a good fielder. In 2000, DM says that "our analysis has shown his range to be consistently average over the years". That assessment is closer to the DRA rating

than the UZR rating. Derek Lee's UZR and DRA ratings differ just enough to merit a check with DM, which has nothing to say in 2000 regarding his fielding.

8. Catcher

I don't have 1999-2001 UZR catcher ratings. As mentioned in Part I.D.4 above, TangoTiger has a great model for pre-UZR periods that I will use to evaluate historical catcher DRA ratings. As a quick "sanity" check for DRA, I'll provide the ratings for I-Rod and Piazza.

Here I will make an exception and "individuate" the ratings, because catchers miss so many games:

	"G"	R/162		91	92	93	94	95	96	97	98	99	00	01
Ivan Rodriguez	1268		"Games"	78	113	**130**	97	120	**137**	**138**	**135**	**136**	84	102
	85	**11**	DRA Runs	11	11	**-1**	-2	9	**22**	**12**	**5**	**18**	0	-1
Mike Piazza	1125		"Games"		15	**135**	97	106	**137**	**134**	**135**	**127**	117	122
	-18	**-3**	DRA Runs		-2	**15**	-10	-6	**-4**	**3**	**10**	**-7**	-11	-6

"Games" are estimated "full" games based on a new way of estimating catcher innings played: gross putouts. "Games" = 162*(individual catcher gross putouts/team catcher gross putouts) (adjusted for strike-shortened seasons). Gross putouts include pitcher strikeouts, so the "skill" portion is overwhelmed by the factor of playing time. Trust me; it works very well, at least for high-strikeout eras such as our own.

Anyway, I think the career ratings are basically right, though they probably understate slightly I-Rod's impact. I-Rod's career rating per 162 games is (+11) close to the career rates Tango found for the best catchers in his study (Jim Sundberg and Gary Carter), but is probably a little low because the DRA regression weight for SB allowed was about half of the UZR weight. The very high number of SB I-Rod "prevented" in 2000 and 2001 are therefore underweighted. I would put I-Rod's "true" career runs-saved per 162 games at about +14 or so. Piazza's ratings in 1999, 2000 and 2001 should probably be a little worse, for similar reasons.

Nevertheless, I think the results are impressive when you consider that CS data is not used. Obviously, we should use the CS data when we have it, but I wanted to show that DRA works without it, so that if I ever get around to rating catchers for pre-Retrosheet seasons, people will have reason to believe that the ratings are basically alright.

One last thought on stolen bases allowed. The regression weight for runs under DRA is virtually identical to the weight determined by John Jarvis under regressions of team offense over the course of the same time period used for the DRA study. It might

be worth additional study to determine whether SB are slightly overweighted by UZR and similar "change-in-state" models.

Unless I've misunderstood the explanation by MGL regarding the derivation of the SB weight under UZR, the SB weight is based not on the actual number of runs scored on average after a stolen base compared with similar situations in which no stolen base is attempted, but simply on the difference, *on average*, in the runs scored with a runner on second base rather than a runner on first base, given the same number of outs. (After all, that is the "change-in-state" effected by a stolen base, ignoring, for the sake of simplicity, steals of third.) As Bill James has observed, "[m]any more stolen base attempts occur when the team is ahead than when they are behind." *Abstract*, p. 357. What kinds of teams tend to have a lead? Teams that score more runs. Since run-scoring is determined *much* more by hitting than stolen bases, it is probably the case that runners who steal a lot of bases are on good *hitting* teams that *would have been more likely to drive them in anyway, whether or not the base was stolen*. Thus, the marginal *real* impact of stolen bases might be significantly less than their theoretical average value. The teams that would *most* benefit by stealing bases might steal too little, and teams that would *least* benefit by stealing bases might steal too much. Both dynamics would decrease the actual average impact of stolen bases on runs scored, thus explaining the lower regression weight.

C. Summary of 1999-2001 DRA-UZR Results

Based upon the detailed analysis above, there would appear to be only one clear DRA error among the 35 players with three years of data: right fielder Vlad Guerrero is rated only average under DRA, whereas he was clearly well above average over the three-year period, and a legitimate Gold Glove candidate in 2001. There were only four other clear errors among the 47 players with two years of data. Third basemen Scott Brosius, Eric Chavez and Corey Koskie are rated average by DRA, whereas all three were clearly above average. First baseman Kevin Young was rated slightly below average under DRA, whereas he was clearly very, very good for a first baseman.

Although there are a few other players whose UZR rating are probably a better match than DRA with DM (Castillo, Lofton, arguably a few others), none of these borderline ratings is meaningfully misleading.

Set forth below is a chart of DRA ratings at all positions, as well as *team*-level UZR rating ("UZRt"), so that you can see that team-level UZR ratings are substantively the same as the individual player UZR ratings ("UZRp") for the full-time players at all positions other than catcher.

		Avg			Notes
Player	**Pos**	**UZRt**	**UZRp**	**DRA**	
Rich Aurilia	6	3	4	1	
Royce Clayton	6	4	5	-1	
Derek Jeter	6	-14	-15	-25	

Edgar Renteria	6	9	5	7	
Miguel Tejada	6	7	6	0	
Omar Vizquel	6	5	7	-10	dm=dra
Deivi Cruz	6	-3	-2	7	
Nomar Garciaparra	6	-6	-6	5	
Chris Guzman	6	-22	-20	-10	
Neifi Perez	6	12	11	8	
Rey Sanchez	6	23	26	22	
Alex S. Gonzalez	6	1	5	2	
Alex Rodriguez	6	13	13	2	?
Alex Gonzalez	6	-10	-10	-9	
Rey Ordonez	6	20	17	4	dm=dra
Roberto Alomar	4	-5	-4	-2	
Ray Durham	4	8	7	7	
Jeff Kent	4	7	4	3	
Edgar Alfonzo	4	4	5	5	
Jay Bell	4	-6	-5	-7	
Warren Morris	4	-6	-8	2	
Pokey Reese	4	26	26	31	
Luis Castillo	4	7	7	-6	dm~uzr
Mark Grudzielanek	4	-3	0	-1	
Adam Kennedy	4	16	13	8	
Eric Young	4	-1	-3	0	
Damian Easley	4	-9	-9	3	dm=dra
Mike Cameron	8	18	20	11	
Steve Finley	8	-3	-1	-4	
Doug Glanville	8	-18	-16	8	dm=dra
Andruw Jones	8	34	34	31	
Bernie Williams	8	-11	-9	-6	
Ken Griffey, Jr.	8	-8	-5	-1	
Marquis Grissom	8	-23	-22	-7	dm=dra
Ruben Rivera	8	38	29	8	dm~dra
Jose Cruz, Jr.	8	-32	-30	-17	dm=dra
Jim Edmonds	8	-13	-9	-4	
Kenny Lofton	8	27	24	9	dm~uzr
Carlos Beltran	8	-5	-2	-2	
Jeff Cirillo	5	12	16	-1	?
Troy Glaus	5	-4	-6	4	
Chipper Jones	5	-5	-6	-15	dial=dra
Joe Randa	5	0	0	4	
Robin Ventura	5	14	13	8	
Scott Brosius	5	18	15	0	dm=uzr
Eric Chavez	5	11	12	3	dm=uzr
Corey Koskie	5	13	11	0	dm=uzr
Mike Lowell	5	-7	-6	2	
Phil Nevin	5	-6	-4	1	

Bobby Abreu	9	-11	-11	-10	
Shawn Green	9	-1	-1	8	
Vlad Guerrero	9	17	16	2	dm=uzr
Brian Jordan	9	8	9	13	
Paul O'Neill	9	-1	-3	1	
Magglio Ordonez	9	-3	-3	-7	
Sammy Sosa	9	5	7	6	
Jermaine Dye	9	33	30	13	dm=dra
Jeromy Burnitz	9	-1	-2	4	
Raul Mondesi	9	-14	-15	8	dm~dra
Luis Gonzalez	7	16	16	5	dm~dra
Gary Sheffield	7	-16	-20	-12	
Shannon Stewart	7	-4	-4	-4	
Ben Grieve	7	-33	-25	-16	
Troy O'Leary	7	13	12	-2	park
Barry Bonds	7	2	3	-4	
Bobby Higginson	7	24	24	9	dm~dra
Carlos Lee	7	-4	-6	-6	
Jeff Bagwell	3	8	8	6	
Carlos Delgado	3	-5	-5	-5	
Mark Grace	3	1	1	0	
Todd Helton	3	10	11	12	
Tino Martinez	3	11	11	5	
John Olerud	3	7	8	12	
J T Snow	3	-12	-11	6	dm=dra
Eric Karros	3	5	6	13	
Ryan Klesko	3	-4	-5	-4	
Derrek Lee	3	-6	-8	6	?
Todd Zeile	3	15	13	10	
Sean Casey	3	-6	-6	-14	
Jason Giambi	3	-3	-3	-11	
Lee Stevens	3	-3	-2	-7	
Kevin Young	3	16	14	-4	dm=uzr

One way of assessing the accuracy of DRA is to regress UZR ratings onto DRA ratings, to see how well they match up. Such an approach assumes that UZR is a perfect measure of fielding impact. I think the discussion above gives us reason to believe that UZR might not always be right. In general, DRA seems to match up better with DM (which is based on zone data) than UZR, except at third base.

When we regress UZR ratings onto DRA team ratings, the regression equation obtained is UZRp = .95*DRA. The coefficient for DRA (0.93) is statistically significant ($p < .0001$). The r-squared is 47%. The square root of the r-squared, or the correlation coefficient, is 0.69, or nearly 0.7.

What this tells you is that the best way to estimate UZR, if you have two or three years of DRA ratings, is to take the DRA value, discounted by only 5%. Such estimate has the same "scale" as UZR, and a nearly 0.7 correlation with UZR.

It's a completely subjective guess based on the DM information discussed here, but I believe that two- or three-year average DRA ratings would have an approximately "one-to-one" scale correspondence and greater than 0.8 correlation with ratings under a "perfect" zone system, whatever that might be.

After making such guess, it occurred to me that there was a very imperfect but suggestive method for supporting it. What I show below is a comparison of UZR and DRA ratings, but with the UZR rating *adjusted to take into account DM commentary*, as well as the results of "regressing" the *adjusted* UZR rating "onto" DRA. If the DM commentary strongly supports UZR (dm=uzr), the UZR rating is left unchanged. If the DM commentary is mixed, but supports UZR more than DRA (dm~uzr), I "shift" the UZR rating one-quarter of the "distance" between it and DRA. If the DM commentary does not provide clear support for either system ("?"), I shift the UZR rating one-half of the distance toward DRA. If the DM commentary is mixed, but supports DRA more than UZR, I shift the UZR rating three-quarters of the way toward the DRA result. If the DM commentary strongly supports DRA, I equate the UZR rating to DRA. For the Fenway left field park effect, I added back the average –8 rating in the sample to the O'Leary rating.

Here's how the charts look, so you can see how the changes came out:

Player	Pos	UZR	DM Note	"DM'd" UZR	DRA
Rich Aurilia	6	4		4	1
Royce Clayton	6	5		5	-1
Derek Jeter	6	-15		-15	-25
Edgar Renteria	6	5		5	7
Miguel Tejada	6	6		6	0
Omar Vizquel	6	7	dm=dra	-10	-10
Deivi Cruz	6	-2		-2	7
Nomar Garciaparra	6	-6		-6	5
Chris Guzman	6	-20		-20	-10
Neifi Perez	6	11		11	8
Rey Sanchez	6	26		26	22
Alex S. Gonzalez	6	5		5	2
Alex Rodriguez	6	13	?	7	2
Alex Gonzalez	6	-10		-10	-9
Rey Ordonez	6	17	dm=dra	4	4
Roberto Alomar	4	-4		-4	-2
Ray Durham	4	7		7	7
Jeff Kent	4	4		4	3
Edgar Alfonzo	4	5		5	5
Jay Bell	4	-5		-5	-7

Warren Morris	4	-8		-8	2	
Pokey Reese	4	26		26	31	
Luis Castillo	4	7	dm~uzr	4	-6	
Mark Grudzielanek	4	0		0	-1	
Adam Kennedy	4	13		13	8	
Eric Young	4	-3		-3	0	
Damian Easley	4	-9	dm=dra	3	3	
Mike Cameron	8	20		20	11	
Steve Finley	8	-1		-1	-4	
Doug Glanville	8	-16	dm=dra	8	8	
Andruw Jones	8	34		34	31	
Bernie Williams	8	-9		-9	-6	
Ken Griffey, Jr.	8	-5		-5	-1	
Marquis Grissom	8	-22	dm=dra	-7	-7	
Ruben Rivera	8	29	dm~dra	13	8	
Jose Cruz, Jr.	8	-30	dm=dra	-17	-17	
Jim Edmonds	8	-9		-9	-4	
Kenny Lofton	8	24	dm~uzr	20	9	
Carlos Beltran	8	-2		-2	-2	
Jeff Cirillo	5	16	?	8	-1	
Troy Glaus	5	-6		-6	4	
Chipper Jones	5	-6	dial=dra	-15	-15	
Joe Randa	5	0		0	4	
Robin Ventura	5	13		13	8	
Scott Brosius	5	15	dm=uzr	15	0	
Eric Chavez	5	12	dm=uzr	12	3	
Corey Koskie	5	11	dm=uzr	11	0	
Mike Lowell	5	-6		-6	2	
Phil Nevin	5	-4		-4	1	
Bobby Abreu	9	-11		-11	-10	
Shawn Green	9	-1		-1	8	
Vlad Guerrero	9	16	dm=uzr	16	2	
Brian Jordan	9	9		9	13	
Paul O'Neill	9	-3		-3	1	
Magglio Ordonez	9	-3		-3	-7	
Sammy Sosa	9	7		7	6	
Jermaine Dye	9	30	dm=dra	13	13	
Jeromy Burnitz	9	-2		-2	4	
Raul Mondesi	9	-15	dm~dra	2	8	
Luis Gonzalez	7	16	dm~dra	8	5	
Gary Sheffield	7	-20		-20	-12	
Shannon Stewart	7	-4		-4	-4	
Ben Grieve	7	-25		-25	-16	
Troy O'Leary	7	12	park	12	6	dra+8
Barry Bonds	7	3		3	-4	
Bobby Higginson	7	24	dm~dra	13	9	

Carlos Lee	7	-6		-6	-6
Jeff Bagwell	3	8		8	6
Carlos Delgado	3	-5		-5	-5
Mark Grace	3	1		1	0
Todd Helton	3	11		11	12
Tino Martinez	3	11		11	5
John Olerud	3	8		8	12
J T Snow	3	-11	dm=dra	5	6
Eric Karros	3	6		6	13
Ryan Klesko	3	-5		-5	-4
Derrek Lee	3	-8	?	-8	6
Todd Zeile	3	13		13	10
Sean Casey	3	-6		-6	-14
Jason Giambi	3	-3		-3	-11
Lee Stevens	3	-2		-2	-7
Kevin Young	3	14	dm=uzr	14	-4

When you regress the DM-adjusted UZR ratings "onto" the DRA ratings, you obtain the following regression result:

DM-adjusted UZR = .94*DRA; r-squared = 67%.

The square root of the r-squared, or the correlation coefficient, is 0.82. Of course this is not a completely objective analysis. But I believe it is highly suggestive and, probably, conservative. Except at third base, DRA tends to match up better with DM than does UZR. However, I did not attempt a "DM adjustment" unless there was a clear discrepancy between DRA and UZR—in other words, even the "DM'd" UZR rating is not fully adjusted. And I wish to assure you that I didn't even think of performing this kind of regression analysis until weeks after I had performed the subjective comparison between UZR and DM. In other words, I did *not* "jigger" the subjective DM analysis in order to "generate" this nice .82 correlation result.

To recap, given two or three years of full-time play, DRA can give you a reliable estimate of whether a player is essentially average (+/- half a dozen runs), meaningfully above or below average (+/- a dozen runs), or exceptionally above or below average (+/- two dozen runs). Close to 95% of the DRA ratings in the study are basically right, or at least not meaningfully wrong.

III. Historical DRA Ratings: 1974-2001

DRA can be applied throughout baseball history, with one modification addressed in Part IV. Unfortunately, I only had conveniently formatted data to evaluate fielders from the 1974-2001 era, but I thought it would be fun to rate even just that subset of fielders.

Posted by baseballprimer as an Excel file is an Appendix providing position-by-position charts that show the DRA per-season ratings for each player who (i) is listed in *Win Shares* as having played at least 5,000 innings at his position and (ii) played at least five full-time (130+ games) seasons (not including seasons split between two teams) during the 1974-2001 time period. (For catchers, I only provided ratings for catchers rated by Tango's system, and I included them even if they had only three 130+ seasons.)

As was done in the 1999-2001 study, the team rating at the position is credited to the full-time player. Seasons in which the player played only 130-145 games (80-90%) are italicized. The Appendix has been formatted so that if you use the option under Excel to "shrink" the page to 80% and print out in "landscape" format, each chart for each position will print out on its own page for easy review. As the charts must be printed out in "landscape" format, they can't be shown in the "body" of this article. However, I'll summarize here, position-by-position, the overall ratings and discuss any surprises.

Providing an "overall" assessment of each fielder is a complicated task. In *The New Bill James Historical Baseball Abstract* (the "*Abstract*"), each player's overall ratings is based upon their average performance under *several* different criteria, in order to try to reconcile the differences in career and peak performance and factors that escape quantification. Bill rates players based upon the following criteria: (i) the average of the player's three best seasons, whether or not consecutive, (ii) the player's five best consecutive seasons, (iii) the player's per-162 game performance over the course of his career, (iv) the player's grand total value (career Win Shares), rescaled so as to be of approximately the same scale as the other averages, and (v) a subjective element.

In general, the overall DRA assessments provided here (Excellent, Very Good, Solid/OK, Poor) are based upon the "Index", which is the average of (i) the average DRA rating for the player's five best consecutive seasons (which are "boxed" in the charts in the Appendix), excluding the severely strike-shortened 1981 and 1994 seasons, unless such seasons were necessary for the player to "reach" five seasons, (ii) the player's "career" DRA rating (i.e., the sum of all of his ratings for seasons in which he played 130 or more games at the position), *divided by five*, and (iii) the average of the player's *second*- and *third*-best ratings, whether or not consecutive or included in their top-five consecutive season ratings, but *only* including seasons of *146 or more* games and *only* if the rating is above-average (above zero). (At catcher I accepted the best and second-best 146+ ratings, if available, given the lack of data.) At times I will point out reasons for incorporating subjective considerations.

The rationale behind this approach will become clearer as we go through the ratings, but in essence what we're trying to establish is a good indication of the player's *significant* and *reliable* contribution with his glove. When I rate a fielder as "Excellent", that means the evidence strongly suggests he had the clear ability to save 20 or more runs a season over an extended period of time, say, five or more seasons. Such players should have a few Gold Gloves to their credit, assuming voters were well-informed and thoughtful (they very often weren't). "Very Good" fielders had the clear ability to save about 10 runs a season for an extended period of time. They usually deserve to win a

Gold Glove or two, depending on the competition in a given year. Solid/OK guys were basically average—actually, usually slightly above average, as players who play full-time are usually better-than-average fielders, as Bill James has observed and quantified. "Poor" fielders cost their teams more than 10 runs a season, compared to league-average fielders, over an extended period of time. (Yes, the "run-spreads" (runs per season) are "tighter" here than those I mentioned in the Introduction and Part II, because two- and three-year performances have wider swings that five-year-or-longer performances.)

The five-year "peak" rating is probably the best single estimate of the quality I'm trying to capture in the historical ratings. The "career" ratings are also important, however, even though the accuracy of the career rating is limited by the fact that part-time seasons are not included, and even though career ratings include "decline" phases, because data beyond the five-year peak provides additional evidence about the player's ability.

The second- and third-best seasons of over 90% playing time are a good confirmation of the player's true peak ability, independent of whatever residual distortion is left in the system by using the team-level rating. By throwing out the "best" (146+) season, we prevent fluke seasons from seriously distorting our estimates. (There is only one truly "weird" single-season rating in the study, but we're trying to be conservative here.) In addition, a player may sometimes have a good season outside of his five-year peak, and it seems fair to factor this in. For those players for whom we lack seasons of greater-than-90% playing time, that very lack of data is something that should be considered. Only above-average "peak" ratings are included, as I found that one or two bad seasons for some players significantly distorted their overall Index. The Index for poor fielders is therefore somewhat shifted toward the mean, but for purposes of all-time rankings I'm more interested in evaluating good-to-great players correctly than in assessing precisely the negative impact of the truly poor fielders

On average, the three "objective" ratings numbers are usually approximately the same. A typical pattern for a Very Good fielder would be a +10 five-year average rating, a "career" runs-saved rating (excluding seasons of part-time play) of +50, and a two-year average peak rating of +10. In other words, for the players who managed to put in more than five full-time seasons, they tended to be average (a "zero" rating) during their non-peak periods. But for those players who played well outside of their five-year peak, taking the career rating and dividing it by five gives them credit for sustained excellence. For players who failed to play two 146+ seasons, that "gap" in their record causes their Index to be shifted closer to an average Index rating, to reflect our slightly lower level of confidence in their individual performance, independent of the effects of their substitutes.

All of the per-season ratings are provided in the charts, so you can devise your own method of summing up the data. I just thought that this was a simple approach that worked reasonably well.

I'll provide a chart of the players listed in rank order by their Index number, and then discuss the notable results. (Remember, as MGL and Bill James have said, if we're

not surprised by the numbers some of the time, we've wasted our time putting them together!).

A. Shortstop

Index	Player	'Career' Total	Top-5 Consecutive Avg	Best 146+ 2d	3d
26	Ozzie Smith	193	21	20	20
23	Dave Concepcion	128	21	23	18
22	Ozzie Guillen	107	22	26	22
13	Mark Belanger [68-71,73]		24	20	10
12	Cal Ripken	64	10	13	13
10	Garry Templeton	42	8	17	11
10	Barry Larkin	40	8	17	11
9	Rick Burleson	57	10	8	6
9	Bill Russell	47	8	10	10
9	Jose Uribe	49	10	13	
8	Craig Reynolds	59	12		
8	Mike Bordick	39	8	9	6
7	Bucky Dent	47	9	2	0
7	Alan Trammell	37	7	7	7
7	Greg Gagne	44	10	2	
6	Tony Fernandez	26	7	8	5
6	Royce Clayton	34	5	12	
6	Jay Bell	33	5	11	
4	Alfredo Griffin	29	3	6	2
3	Dick Schofield	18	4	5	0
1	Chris Speier [71-73]		3		
1	Omar Vizquel	-24	5	4	4
0	Robin Yount	1	1	1	
0	Rafael Ramirez	-3	-3	8	
0	Larry Bowa [70-72]		-3	4	
-1	Spike Owen	-10	-2	3	
-1	Freddie Patek [69,71-73]		-3		
-1	Walt Weiss	-14	-2		
-2	Shawon Dunston	-16	-3		
-3	Alex Rodriguez	-22	-4		
-4	Frank Taveras	-40	-2		
-5	Ivan DeJesus	-63	-4		
-12	Derek Jeter	-102	-15		

In general, I'm happy with how the Index sorts out the shortstops.

The Excellent shortstops during 1974-2001 were Ozzie Smith, Dave Concepcion, Ozzie Guillen, and Mark Belanger. The Very Good shortstops were Ripken, Templeton, Larkin, Burleson, Jose Uribe and maybe Craig Reynolds and Greg Gagne. Bill Russell was not as good as he looks, for reasons I'll explain in discussing Steve Garvey's first

base rating. Russell was probably Solid/OK. The clearly Poor shortstops were DeJesus and Jeter. (Bowa would probably rate as Poor if we had all of his seasons and factored in his "Career" rating.) The rest were Solid/OK.

Ozzie Guillen won only one Gold Glove, but he should have won a few more during the late 1980s that Tony Fernandez won with his bat and solid but unspectacular fielding. Guillen's fielding reputation was probably harmed by his fairly high error rates, and the antiquated notion that errors should count *twice* against a fielder, once in reducing the number of plays credited to him, and a second time because . . . because that's how statistics were designed 150 years ago, when there were no fielding gloves, and simply being able to catch a ball cleanly was probably as important as range. In the modern era (or at least from 1974 through 2001), there has been no statistically significant connection between shortstop errors (or errors at any other position, except pitcher and right field) and runs allowed, *after you take into account context-adjusted plays made*. Evaluating major league fielders by their error rates or totals is a simply a bad habit we should just give up, except when we have highly accurate UZR data that can isolate the marginal detriment of particular kinds of errors; even then, the marginal impact is vanishingly small. (End of rant.) The *Abstract* has a wonderful essay about Ozzie's exciting brand of play, both as a fielder and a baserunner. When you look at the year-by-year chart, you'll see in dramatic fashion how Ozzie's outfield collision with Tim Raines in 1992, which, according to the *Abstract*, cost Ozzie his speed, also caused his DRA ratings to drop. But for that injury, Ozzie Guillen might have had as distinguished a career as a fielder as Ozzie Smith.

We're missing the heart of Mark Belanger's career, so his Index understates his value. Even towards the end of his career, he demonstrated that he was among the greatest fielding shortstops of all time.

Barry Larkin's Index is a little lower than it should be, because he failed to reach the 130-game threshold for a lot of his seasons. However, we have enough seasonal ratings of full-time play to conclude that he has been the kind of shortstop who saves about ten runs a season, and probably saved somewhat more in the best seasons in the very beginning of his career.

As indicated in Part II, Omar Vizquel has simply not been any better than an average shortstop. He won his first Gold Glove during his only genuinely good season (1993; DRA +13 runs). That season looked better than it was, because his pitching staff had very low strikeouts, generated a lot of ground balls, and therefore permitted Omar to rack up a huge number of assists. Apparently he is graceful, he avoids making errors, he played on the dominant non-Yankee team of the mid- to late-1990s, and the voters couldn't agree on who else to honor, so they just settled on him. Partly this was due to the fact that *nobody* was noticeably outstanding at shortstop during the 1990s. Royce Clayton was probably the best shortstop who played five full seasons after 1990. His Index is slightly distorted by one single season negative rating. The "compression" in shortstop ratings in the 1990s is generally reflected in the ratings at other positions. I'll discuss this trend at the end of this historical survey.

Greg Gagne's Index is probably too low. His plate appearance data suggests that he was playing a lot of partial games, so his ratings might be diluted by poor back-up shortstop fielding. Bill James has praised his fielding, particularly during the early 1990s, when he was with the Royals, and his single season DRA ratings are generally very good, sometimes exceptional, during that period.

A-Rod seems to be improving his fielding of late, but his ratings through 2001 are only basically average, on average. See the discussion of his 1999-2001 performance in Part II.

Although "Career" ratings are missing part-time seasonal data, only two or three of Ozzie Smith's seasons are missing. Given that, I'm very confident that Ozzie saved about 200 runs in his career, compared to a league-average shortstop. That is the highest career runs-saved number in the study. Likewise, the career number for Ripken at short is correct, for obvious reasons. He was genuinely very good for about his first four or five seasons, and average, on average, after that.

B. Second Base

Index	Player	'Career' Total	Top 5 Consecutive Avg	Best 146+ 2d	Best 146+ 3d
15	Ryne Sandberg	93	10	19	14
14	Harold Reynolds	62	14	17	15
14	Glenn Hubbard	104	21		
12	Frank White	81	10	12	9
12	Bobby Grich [73]	79	9	14	9
12	Lou Whitaker	71	15	13	
10	Jose Lind	42	10	19	3
9	Willie Randolph	68	10	8	
9	Dave Cash	40	8	19	2
7	Roberto Alomar	56	7	4	3
7	Bret Boone	29	7	10	5
7	Carlos Baerga	40	8	8	
6	Ray Durham	23	4	11	6
6	Jody Reed	54	7		
5	Manny Trillo	22	7	7	2
4	Robby Thompson	25	5	7	
4	Jeff Kent	31	6		
2	Johnny Ray	6	1	6	2
2	Davey Lopes [73]	18	2		
1	Bill Doran	-2	2	5	
1	Jerry Remy	-8	3	2	
0	Scott Fletcher	-2	0		
0	Damaso Garcia	1	0		
0	Chuck Knoblauch	-30	-1	13	1
-1	Jim Gantner	-9	-2	1	
-2	Joe Morgan [many]		-5		
-2	Julio Franco	-39	-1	2	1

-4	Marty Barrett	-30	-6
-5	Juan Samuel	-42	-5
-8	Craig Biggio	-89	-5
-8	Tom Herr	-52	-13
-12	Steve Sax	-136	-8

There are a few more surprises at second, but I believe the ratings are essentially correct. Oh, I forgot to mention . . . when I put bracketed numbers next to a player, it indicates missing seasons of full-time play. When only one full-time season was missing, I thought it better to include than exclude the "Career" rating.

At second, nobody dominated in the way that some dominated at short, except possibly Frank White, who possibly should be rated as Excellent. I would put about four players in a category between Very Good and Excellent: Ryne Sandberg, Harold Reynolds, and Glenn Hubbard. Grich, Whitaker, Lind and Randolph were Very Good. Sax and Herr were Poor. Biggio is better than he looks, for reasons I'll explain. The rest were Solid to OK.

Frank White's rating is probably too low, because DRA does not give infielders credit for putouts, for reasons explained in Part I.D.1. Although this is the correct approach to take in the vast majority of cases (UZR and DM take this approach as well), in the case of Frank White, this is probably inappropriate, and illustrates why, no matter how good a system is for evaluating fielders, it is important to incorporate subjective observations of *unusual patterns of play* in order to arrive at the best possible assessment:

> "Artificial turf had become common in the early 1970s; the Royals moved into Royals Stadium, which had turf, in 1973. White was the first second baseman to fully grasp how the turf had changed the game, and the most successful second baseman at implementing adjustments. With the artificial turf, White realized that the ball would not roll dead in front of him, thus he could set up deeper than second basemen traditionally did. White probably stationed himself further from second base than any other second baseman who ever played the game, playing shallow right field 50, 60, or even 70 feet from second base. He often caught pop ups in foul territory. I saw him many times catch a dying quail over first base, a ball that might have rolled into the corner for a double had White not been playing where he was. More routinely, he would catch the soft single into right field, while with his great quickness and playing deep he was still able to get to balls hit behind second. Of course he would edge toward second with a right-handed hitter at the plate or a man on first, but even so, for White, cheating toward second base meant setting up 20 or 25 feet behind the bag." *Abstract*, p. 504.

I have no way of quantifying what Bill James has observed, but, if we assume conservatively that Frank White was genuinely saving, say, ten hits a season by catching short fly balls that none of his teammates would have been able to reach, his career runs saved total could easily be higher by another 75 runs, maybe more.

There's been a tendency of late to discount Ryne Sandberg's fielding. DRA does succeed in taking a lot of the "starch" out of some of his ratings. But in 11 seasons of 130+ game play, he has only four mildly negative ratings, none of which is worse than –6, and several very good to excellent ratings. He certainly did not deserve all of his Gold Gloves. But for sustained excellence over a career, he was the best, excluding Frank White.

Harold Reynolds didn't play that long, but for his first four full-time seasons, he was Excellent, and then tapered off.

About a third of Glenn Hubbard's "Career" rating should be discounted. He obtained the only freakish single-season rating in the entire study, at any position (+49 runs, in 1985). I'm proud of the fact that, except for that one rating, DRA succeeds in keeping his other ratings around +14, well below the excessive ratings he receives under some other systems. He only had one 146+ game season. His rating for that year was +16.

The rating for Trillo may seem a little low, but I'm open to any analysis that shows how and why it should be higher. Alomar's rating is correct, notwithstanding all his Gold Gloves. "An overrated fielder, in my opinion; a good fielder, even a very good one, but no better than some guys who don't win Gold Gloves, like Fernando Vina." *Abstract*, p. 491. DM agrees, at least with regard to his fielding in 1999-2001.

We're missing most of Joe Morgan's career. He was not a good fielder; at best, he was OK. Charlie Saeger has data indicating that Joe might also have been unusually good at catching fly balls, as Frank White was. Joe's putout rates declined when his stolen base attempts dropped in 1978. Unfortunately, I just don't see how to provide reliable single season ratings for putouts. I'll talk more about Joe when we discuss some "No Joe Morgans" in the Center Fielder section.

Craig Biggio is neither as bad as his "Career" rating nor as good as his four Gold Glove awards would suggest. Craig Biggio is the only middle-infielder in major league history of which I'm aware who began his career as a *catcher*. He caught almost 400 games—and made the All Star team as a catcher—before moving to second base. His first DRA rating at second is –28 runs, which is not surprising when you consider that he was moving from the position requiring possibly the *least* agility to the position requiring, at least for purposes of turning the double play, the *most* agility. He was basically OK by the following season, and gradually improved his performance to a point at which he arguably deserved the Gold Glove he received in 1997 (DRA +14). He suffered some sort of injury in 2000 (causing him miss over a third of the season); in 2001, his DRA rating basically collapsed (-25 runs), right in sync with his stolen base attempts. His 2001 ESPN zone rating is dead last among full-time major league second basemen. I don't have DRA ratings for 2002. If you give Biggio a "pass" for the seasons he spent learning the position and coping with injuries, he's slightly below average, but not bad. In 2003 he was moved to center field, ostensibly to make room for Jeff Kent. Biggio's zone rating this year, as reported by ESPN, is second-to-last.

If you don't believe me (or DRA) that making the move from catcher to second is difficult, ask Biggio:

> "I would never say this when I was still playing second, but going from catcher to second base was like telling me I'm going to be the president of Wal Mart tomorrow. That's how out of place I felt. . . . That was one of the hardest things I ever had to do in my life. I can't express how difficult that was. I was a catcher my whole life. Then I had to learn a new position in the big leagues. And everything about it was so different. As a catcher, everything goes away from you. As a second baseman, everything is coming at you. So you're in the field, and you've got to concentrate every second: "Who's up? Fast guy? Slow guy? If a ball is in the gap, where do I go?" Cutoffs. Relays. Turning the double play. I had to learn all that. It took me two-and-half to three years until I actually got comfortable, where I didn't have to think about it, where it was second nature." Interview reported by Jayson Stark for ESPN, March 12, 2003.

I have great respect for Craig Biggio, who is a first-ballot Hall-of-Famer in my book. He just wasn't a great fielding second baseman. Simply turning himself into an adequate second baseman was a truly historic accomplishment.

Tommie Herr had famously low error rates but poor range. The Cardinals rating at second improved when he left. Steve Sax was terrible; all 11 of his single-season ratings are negative.

One last point on the second base ratings. Second base is probably the only position at which a significant number of players (White, Sandberg and Alomar) seemed to be able to maintain high ratings (and even reach sustained performance peaks) later in their careers, as you'll see in the Appendix. I don't have any theories why this might be the case. Perhaps turning the double play is a skill that can be maintained longer than range. Mazeroski's assists totals remained high into his thirties.

C. Center Field

Index	Player	'Career' Total	Top 5 Consecutive Avg	Best 146+ 2d	3d
22	Andruw Jones	143		43	31
14	Garry Maddox [73]	97	21		
13	Lance Johnson	64	10	23	12
13	Omar Moreno	61	12	17	14
12	Devon White	53	10	19	14
11	Chet Lemon	97	14	2	
10	Marquis Grissom	36	13	16	6
6	Brett Butler	17	5	11	7
5	Steve Finley	24	6	9	
4	Dwayne Murphy	17	3	13	
3	Kenny Lofton	20	5		

1	Brian McRae	-23	3	5	5
1	Willie McGee	9	2		
1	Cesar Cedeno [72,73]		2		
1	Bernie Williams	3	2		
0	Andre Dawson	0	-3	7	
0	Ken Griffey, Jr.	-30	3	5	2
0	Ray Lankford	-11	2	2	
0	Lloyd Moseby	-19	0	4	1
-1	Kirby Puckett	-7	-4	5	
-3	Amos Otis [70-73]		-9		
-4	Rick Manning	-28	-7		
-6	Robin Yount	-50	-8		
-10	Dale Murphy	-72	-14		

At first glance, there seem to be a lot of mistakes. All of them can be explained as quirks of the "Index" caused by missing data, as well as common misperceptions of player value.

Before proceeding with a discussion of outfielders, I should add that the 1974-2001 historical outfielder ratings also include credit for assists and, in the case of right fielders, "debits" for errors. Believe it or not, DRA runs-saved estimates based on team outfield assists do have a very strong correlation with team UZR outfielder "arm" ratings, even though assists only track baserunners "killed" and not baserunners "held", which UZR can do. The DRA runs-saved based on outfielder assists were generally not very significant in the 1974-2001 study, except for Jesse Barfield. Also, because outfielders are generally either centerfielders or left/right fielders, I included players who had five 130+ seasons at left and right, on a combined basis, but only included centerfielders who had at least five 130+ seasons in center. If a player satisfied either criterion, I also kept track of their other seasons: e.g., in the Left Fielder section, I report Ricky Henderson's center field runs-saved because he had more than four 130+ seasons in left.

Based on DRA information, there are two Excellent center fielders: Garry Maddox and Andruw Jones. I made an exception in including Andruw; we only have four seasons of DRA ratings: +43, +48, +21 and +31. As I mentioned in Part II, I believe his ratings, which match up with UZR, overstate his value, as he has almost certainly been taking discretionary chances away from Atlanta's middle infielders. Nevertheless, he's been excellent. We only have one 146+ rating for Garry Maddox: +24. The lack of a second- and third-best 146+ rating causes his Index to be too low. I think we can be very confident that both Garry and Andruw had the clear ability to save 20 runs a season over an extended period of time.

The Very Good centerfielders are Lance Johnson, Omar Moreno, Devon White, Chet Lemon, and Marquis Grissom. Andre Dawson and Kirby Puckett *were* only OK centerfielders. Robin Yount was *at best* an OK centerfielder. Dale Murphy *was* a Poor center fielder.

Amos Otis and Cesar Cedeno score too low almost certainly because we're missing the first three or four of their full-time seasons. One interesting finding of the 1974-2001 study is that outfielders, and in particular center fielders, simply do not maintain high ratings after five seasons or so. The legs almost always give out. In contrast, several infielders obtain high ratings well into their 30s (e.g., Ozzie Smith, Ryne Sandberg, Mike Schmidt, Greg Gagne, Mike Bordick).

Lance Johnson never won a Gold Glove. He did steal over 300 bases in his career, with a 76% success rate. He led the league in triples *five* times.

Devon White (seven Gold Gloves) and Marquis Grissom (four Gold Gloves) have had a few genuinely good years in the field, by any reckoning.

Chet Lemon and Omar Moreno are probably examples, in Gold Glove voting, of "No Joe Morgans" (as in, "he's no Joe Morgan"). Joe Morgan started winning Gold Gloves he didn't deserve when he was past thirty (Rennie Stennett should have won them, though he is not in the study because he didn't have five 130+ game seasons), because he had developed into being one of the very best *hitters* in baseball at a position where good hitters were rarely found, avoided making errors, was a heads-up player in *all* aspects of the game, and a leader of one of the greatest teams in major league history. (Sometimes Gold Gloves are really "honorable mentions" for MVP. Recall how A-Rod got his first Gold Glove in 2002 when Tejada was given the MVP.) Lemon was probably denied a Gold Glove because he was widely (and correctly) perceived as a poor "percentage" player; Moreno never won a Gold Glove in all likelihood because he was a terrible hitter for an outfielder, even for a center fielder. Lemon was fast, but made *visually obvious* fielding mistakes:

> "Chet Lemon is an enigma . . . . Baserunning is not Lemon's strong suit, despite his above average speed. His instincts are poor . . . . He has a bad habit of diving headfirst into first base. That has produced more nagging injuries than base hits . . . . His judgement on throws from the outfield has been questioned and he is prone to missing the cutoff man." *The Scouting Report: 1987* (as cited in the *Abstract*).

Lemon's career "arm rating" under DRA is exactly average. Moreno, of course, had no power and notably poor strikeout-to-walk ratios. But both Lemon and Moreno were Very Good center fielders.

Dwayne Murphy was probably a Very Good centerfielder before some sort of injury took away his speed. As can be seen in the Appendix, his ratings went negative after an injury in 1983 caused him to miss significant time and reduce his stolen base attempts by about 75%.

Bernie Williams started out as a very good centerfielder in his three full-time seasons, but has gradually lost his speed and throwing arm, thus resulting in an average career rating through 2001.

The Gold Glove voters were simply wrong about Andre Dawson, Dale Murphy and Kirby Puckett.

Andre's teams had fly ball pitching staffs, so he caught a lot of balls, plus he was graceful and always put in an excellent effort. (See his profile in the *Abstract*.) He was OK; just nothing special. Montreal moved him to right field, where he immediately obtained one very good rating before his knees completely failed him.

Dale Murphy exemplified the Craig Biggio and Joe Morgan phenomena discussed above. He was a catcher/first baseman who was moved by his team to a position *per se* inappropriate for him, but was voted Gold Gloves because he made few errors, hustled, was respected/not flakey, and was possibly the most dominant hitter of his time at a demanding fielding position. He is missing a five-year rating because one of his full-time seasons was the strike-shortened 1981 season. All five of his ratings are negative. He just missed getting a sixth negative rating (–12, in 1980) because he played only 129 games in center. While it is true that his teams had ground ball pitching staffs, DRA takes that into account, and there are no unusual infield or outfield DRA ratings for Atlanta (except Glenn Hubbard's 1985 season) during Dale's tenure in center. When Dale's playing time in center dropped temporarily in 1982 (he played 162 games, 118 in CF and 65 in LF, thus suggesting he was moved to LF for defensive purposes late in games), the team CF rating rebounded from –12 in the strike-shortened 1981 season (approximately –17 on a full-season basis) to –9. Dale stole a fair number of bases, but his success rate was only average, even during his peak years. When his team moved him to right, his rating shot right up. He was, for a brief time, a Very Good right fielder. There might be a slight park effect, but when the Braves got a real speedster at center—Otis Nixon, in 1991-93, before other NL parks "turned over"—the Braves CF rating was well above average.

Kirby Puckett established his Gold Glove reputation by throwing out a lot of baserunners in his first two seasons and by enthusiastically chasing down the unusually high number of fly balls generated in the late 1980s by the Minnesota pitching staff, which had low strikeout totals and generated few ground balls. Kirby recorded a huge number of putouts (465) in his first 130+ game season. However, his pitching teammates ranked 12th out of 14 in strikeouts, and DRA indicates that they were a fly-ball-oriented staff—and not just by reference to Kirby's numbers. The simplest way of seeing that Kirby was only average that year is that Kirby had an almost exactly average percentage of his team's outfield fly outs. Center field putouts equaled just under 40% of total outfield putouts in the AL from 1974-1992. Kirby played 161 games in center. He exceeded the 40% level by only a few putouts. The next season, Kirby had 429 putouts (still a fairly high total), but his pitchers were 8th in strikeouts and *first* in home runs allowed—a strong indicator of fly ball pitching. (Bert Blyleven gave up *fifty* homers.) In the following season, Twins pitchers were above-average in strikeouts and only fourth in home runs allowed—and Kirby's putouts dropped to 341 (although he played in "only" 147 games; projected out to 162 games, his putouts would still have been clearly below 400). By this time, however, his reputation was established, he was awarded his first Gold Glove and, as has happened so many times before in Gold Glove voting, once the

voters decide they like someone, they stick with him until someone spectacular comes along to take his place. Two last pieces of indirect evidence—Kirby was *not* a good base-stealer (63% career success rate) and was among the top-ten in triples only once, even though he was always among the top ten in doubles, and other Twins players at the time succeeded in cracking the top ten in triples, so park effects were probably not at issue. In light of all these facts, as well as the fact that he was among the top ten in On-Base-Plus-Slugging ("OPS") only three times in his career, I view his Hall of Fame selection as a mistake, an especially unfortunate one considering the recent discovery that his "charity" work was really the result of the efforts of his long-suffering wife.

Robin Yount's single-season numbers at center, as shown in the Appendix, show that he was barely average or significantly below average in five out of his six full-time seasons in center. Just because you're adequate at short won't make you great in center; center is a less demanding position, but it requires a very different skill set.

Having had so much fun confounding conventional wisdom at center, I reluctantly turn to third, where the ratings make almost *too* much sense.

D. Third Base

Index	Player	'Career' Total	Top 5 Consecutive Avg	Best 146+ 2d	3d
25	Terry Pendleton	126	22	28	25
23	Mike Schmidt	166	17	20	17
22	Tim Wallach	124	20	25	20
20	Buddy Bell [73]	101	21	20	19
14	Ken Caminiti	82	13	13	10
12	Gary Gaetti	65	11	14	12
12	Darrell Evans [73]	55	11	20	8
11	Graig Nettles [70-73]		14	23	14
9	Robin Ventura	53	5	10	10
8	Ron Cey [73]	11	11	11	11
7	George Brett	38	10	6	2
7	Doug DeCinces	44	10	2	
5	Jeff Cirillo	22	5	10	
5	Matt Williams	37	7	2	
4	Travis Fryman	15	3	10	
1	Wade Boggs	-6	2	5	1
1	Kevin Seitzer	3	1	4	1
0	Todd Zeile	-10	-2	5	
0	Larry Parrish	0	0		
-1	Enos Cabell	-11	-2		
-1	Sal Bando [68-73]		-4		
-1	Charlie Hayes	-8	-1		
-2	Ray Knight	-13	-3		
-2	Vinny Castilla	-16	-3		
-3	Ed Sprague	-21	-4		
-3	Jim Presley	-26	-5	4	

-4	Brook Jacoby	-30	-6
-4	Ken Reitz [73]	-46	-4
-5	Carney Lansford	-42	-5
-6	Toby Harrah	-48	-10
-10	Chipper Jones	-72	-14

The Index appears to have rounded up the usual suspects in an orderly fashion. Pendleton, Schmidt, Wallach and Buddy Bell were the Excellent third baseman. Had we had all of Graig Nettles seasons, he would likely have achieved an Excellent rating. Pendleton had the highest peak; Schmidt the most sustained excellence; Wallach may have had three years of ratings slightly inflated by handling a lot of what were really shortstop chances for Hubie Brooks, but he was at least Very Good. Caminiti, Gaetti, Evans, Ventura and Cey demonstrated that they were Very Good, although Cey had a few poor ratings at the end of his career with the Cubs. Depending on one's focus on peak v. career value, Brett and DeCinces could be rated as Very Good as well.

In general, there is a wider "spread" in performance at third, because third base is where you put some phenomenal fielders, as well as some truly terrible fielders, when you need to get their bats in the lineup and other good hitters have to play first or left.

Are there any surprises? Charlie Hayes started out with a couple of pretty good seasons, but quickly declined. Ken Reitz was a classically overrated low-errors, low-range guy. As mentioned in Part II, Cirillo has been Solid/OK, not Gold Glove material. Chipper Jones, as mentioned in Part II, seems to have been consistently well below average. The Braves, at least, believe so, as they sent him to left field. (And I tend to trust the judgment of Atlanta management in this area; Atlanta had the best sustained team-level fielding DRA ratings in the '90s, including seasons before Andruw arrived.) Toby Harrah started life in The Bigs as a shortstop; all five of his DRA ratings at third are negative.

Perhaps the most significant finding at third base is the magnitude of Mike Schmidt's fielding accomplishments. He did not have a single negative rating. When he moved to first for significant portions of the 1985-86 seasons, the Phillies team rating went negative; when he returned to third full-time, the rating went positive again. Allen Barra has written persuasively that Schmidt's *hitting* accomplishments, in the context of their times, may approach those of the very greatest players of all-time. Adding approximately 170 runs of fielding value—at a tough fielding position—may eventually lend support to the idea that Schmidt dominated the 1970s and 1980s nearly as much as Mays dominated the 1950s and 1960s. I would place Schmidt among the ten greatest major league players of all time—just outside the "inner circle" of Wagner, (Oscar) Charleston, Ruth, Mays, and Bonds. I might even place him within that inner circle, if one considers factors such as the likely concentration of talent in the 1970s and 1980s, when baseball was fully integrated and the number of teams relative to the population was (I think) lower than it is now. Consider this: Mike Schmidt may be the only player in major league history who was both the best hitter and the best fielder (on a career basis) at his position. (I suspect that Brooks Robinson had less range.)

E. Right Field

Index	Player	'Car' Total	Top 5 Consecutive Avg	Best 146+ 2d	3d
30	Jesse Barfield	149	29	33	32
10	Sammy Sosa	41	8	16	12
9	Tony Gwynn	25	10	11	10
6	Harold Baines	15	8	7	6
6	Dwight Evans	43	8		
5	Paul O'Neill	25	6	7	4
5	Manny Ramirez	23	5	12	1
5	Dave Parker	6	10	6	4
1	Dave Winfield	-47	6	14	
0	Darryl Strawberry	-14	2		
-2	Ruben Sierra	-22	-5	3	2
-2	Andre Dawson	-15	-3		
-2	Tom Brunansky	-16	-3		
-6	Ken Singleton [73]	-54	-7		
-7	Jeff Burroughs	-109			
-10	Jay Buhner	-86	-14		

The sample of players who played five full season in right field was small. In the left field section, I will discuss the players who played five or more seasons in left or right, as well as the corner outfield ratings of centerfielders.

There was only one Excellent right fielder during the period (Jesse Barfield) and only two Very Good fielders (Sosa and Gwynn). Barfield saved about 50 runs with his arm, and about 100 with his surprisingly good range. His ratings were very consistent, albeit over a fairly short period of time. Tony Gwynn was a Very Good right fielder (his team moved him to center for a while, which explains the 1988-89 "gap" in his single season ratings in the Appendix) until he got fat. Sosa has been consistently Very Good to Solid throughout his career thus far.

There's not much to say about the Solid/OK fielders. Andre Dawson had one positive rating in right field (+13, in 1984), had knee problems that caused him to miss significant playing time for two years, and had negative ratings in right thereafter. Dale Murphy was +14, +8 and +1 in his years in right field after moving from center. Dave Parker was pretty good, arguably Very Good, until he started having troubles with drugs and his weight.

Buhner won one Gold Glove. I have no idea why. All of his DRA ratings, except the first, are negative. His Gold Glove might have been an honorable mention for MVP. (See the "No Joe Morgan" discussion in the Center Fielder section.) He finished in the top five in MVP voting in 1995, made the All-Star team in 1996, and was voted a Gold Glove in 1996, but did not finish in the top ten in MVP voting in 1996.

Ken Singleton was a fine hitter but an immobile outfielder.

Dave Winfield had a few strong ratings, including one +5 rating in left (1982) in the early part of his career, but also a few truly poor ratings towards the end. The Gold Glove voters formed their opinions and did not change them. That he was a great hitter and extremely graceful no doubt helped him.

Jeff Burroughs was a consistently horrible fielder for four full seasons in right ("career" –109), moved over to left, and immediately obtained a –19 rating before his career effectively ended.

F. Left Field

Top 5

Index	Left Fielder	'Career' Total	Consecutive Avg	Best 146+ 2d	3d
25	Barry Bonds	134	22	26	26
17	Jose Cruz	106	15	17	15
16	Luis Gonzalez	108	16	19	4
8	Rickey Henderson	84	8		
5	Albert Belle	21	5	8	6
5	George Foster	40	5	2	
1	Jim Rice	-4	-1	6	6
1	Tim Raines	8	2		
-2	Gary Sheffield	-35			
-4	Greg Luzinski	-65			
-5	Dusty Baker	-40	-7	1	
-8	George Bell	-74	-10		
-8	Mike Greenwell	-62	-12		
-9	Gary Matthews [73]	-69	-12		

Barry Bonds was an Excellent left fielder for several years, and a pretty good left fielder thereafter. He doesn't really have a negative single season DRA rating until 2001. Were it not for the fact that Ted Williams also played left field, it might be possible to say of Barry, as we can probably say of Mike Schmidt, that he was both the best offensive and defensive player at his position in major league history.

Jose Cruz and Luis Gonzalez were (have been) Excellent-to-Very Good. Jose belongs in the Hall of Fame, but of course will never get there. As Bill James has pointed out, his batting statistics were "ruined" by playing in the Astrodome. If you include his right field ratings, his career runs-saved in the outfield was +115. I thought there might be a park effect helping him in the field in a manner analogous to the park effects that hurt him at the plate, but the left field ratings in the Astrodome are almost exactly average in the years before and after he played, with one significant exception.

Namely, Luis Gonzalez. He started out like gangbusters as a young left fielder in the Astrodome, and had one very high rating +44, at age 25. I would've been inclined to consider that a "wrong" rating, and perhaps it should be discounted a bit, though UZR

shows similarly high outfield ratings sometimes, and DRA shows Andruw Jones and Jesse Barfield reaching numbers that high, although on a more consistent basis. In Part II, in the discussion of 1999-2001 left fielder ratings, I note that DM says that Gonzalez has generally provided "good defense in left field." His DRA ratings are only slightly above average during that period, but by that time he was in his thirties. All of his ratings outside of the Astrodome are positive (except one –1 rating with Detroit in 1998), though not as good as those in the Astrodome. I suppose it is possible that a player who plays in any unusual environment for a fairly long time probably can develop an unusual skill in exploiting that environment. (Think Yaz and the Green Monster.) But, to repeat, there were at least a dozen "average" ratings at left field in the Astrodome during the seasons neither Cruz nor Gonzalez were playing, so I don't believe they were obtaining an "unfair" advantage through a park effect.

Rickey Henderson has also been a Very Good fielder. He had +24 runs in two seasons in center, and +84 runs saved in left field. Although we're missing his many seasons of less-than-130-games-at-one-outfield-position fielding, I feel comfortable concluding that he saved more runs in the field than he created with all of his basestealing.

Due to the Fenway effect on left fielder ratings, Jim Rice should rate somewhere between Very Good and Solid, and Mike Greenwell should probably rate as Solid/OK. As mentioned before, the average Red Sox DRA rating in left field was –8 during 1974-2001.

Although he had a poor reputation as a fielder, Albert Belle was OK both in left and in right (+5 in 1999). Bill James has written that he was an underrated baserunner.

I included Greg Luzinski for fun, even though he had only four 130+ seasons during the 1974-2001 study. He was terrible, of course, and almost certainly *much* worse than his team rating, as the Phillies had a defensive replacement in left field for portions of as many as a half or third of the games he played in left. It is probably the case that Greg almost never finished a game that his team was leading after his last likely at-bat.

Gary Matthews has two negative rating in left (-13, –8), as well as the significantly negative rating in right shown above. George Bell had one positive rating (+7) in his first full season and five negative ratings thereafter. There is an amusing essay about his fielding in the *Abstract*. He apparently led or nearly led the league in errors three years in a row. Sometimes players have bad hands *and* bad range.

Gary Sheffield has been a consistently Poor outfielder. His "Career" rating in right is –30, and his "career" rating in left through 2001 is –35. His UZR ratings in Part II indicate he might be even worse than that, and he's not getting better with age.

Kevin McReynolds didn't have five ratings in either center or the corners, but his ratings tell a story: in chronological order, he had two positive ratings in center field (+7,

+4) and the following ratings in left: +4, -1, 9, -19. "He was a top NL outfielder for a few years, put on weight and disappeared." *Abstract*, p. 683.

G. First Base

Index	Player	'Career' Total	Top 5 Consecutive Avg	Best 146+ 2d	3d
20	Keith Hernandez	124	17	20	18
16	Bill Buckner	99	20	15	
13	Eddie Murray	66	12	14	14
12	Mark Grace	60	12	18	8
12	Jeff Bagwell	80	9	14	11
11	Eric Karros	70	7	13	12
10	Pete O'Brian	54	7	15	11
10	Rafael Palmeiro	35	9	16	13
10	Wally Joyner	61	12	13	
7	John Olerud	22	8	9	7
6	Tino Martinez	29	6	10	5
4	Mike Hargrove	32	6		
3	Willie Upshaw	11	4	6	2
3	Chris Chambliss	12	3	5	2
2	Fred McGriff	-7	0	9	6
2	J. T. Snow	1	4	3	
0	Will Clark	-2	1		
1	Cecil Cooper	-3	2	2	
-2	Tony Perez	-13	-2		
-3	Kent Hrbek	-28	-2		
-3	Don Mattingly	-28	-3		
-4	Jason Thompson	-38	-6		
-5	Mo Vaughn	-51	-4		
-5	Andres Galarraga	-50	-4		
-5	Dan Driessen	-34	-7		
-6	Mark McGwire	-55	-7		
-8	Steve Garvey	-113	-3	1	

As I did in Part II, I will avoid a lengthy discussion of individual first base ratings. Keith Hernandez was so outstanding that his fielding makes him a legitimate Hall of Famer. I was at first surprised that there were as many Very Good first basemen as there were, as the standard deviation in team first basemen ratings is about 9 runs per season, whereas it is about 12 for the other positions. I think a dynamic similar to the one we saw at third is at play here: there are many awful fielders at first base who bring the average down, so the very small number of good fielders tends to stand out.

There are two obvious errors, and possibly two other mistakes. As Bill James first noticed, Bill Buckner racked up a tremendous number of discretionary assists by forcing his pitchers to cover the bag on virtually every ground ball to first, even if he was able to make the putout himself, whereas Steve Garvey, who had no confidence in his throwing arm, would avoid making a throw to the pitcher whenever possible, and

recorded many more unassisted putouts instead. DRA ratings at first are based solely on context-adjusted assists, so Buckner is overrated and Garvey is underrated.

Buckner and Garvey at first (and Frank White at second) are excellent examples of how important it is to take into account subjective observations of unusual patterns of play. They are also excellent examples of how care should be taken in considering unusual patterns of play by a small number of players when determining the standard methodology for rating other fielders at their position.

Bill James' approach to dealing with the problem of Buckner/Garvey elections is to estimate the total number of unassisted putouts at first ("EUPO-3") and add such plays made to first base assists to determine total plays made. (I believe that Charlie Saeger first had the idea of estimating unassisted putouts at first.) When I added EUPO-3 with first base assists, I found that the resulting ratings had little or no correlation with UZR ratings of first basemen, which are based on ground balls fielded, and which *do* include unassisted ground ball putouts.

I think there are two reasons for this. *Most* of the variation in unassisted putouts at first base is due to pop-ups, and most first baseman do *not* have a severe bias towards a Buckner or Garvey approach to discretionary assists. There are two pieces of evidence in support of these assertions. First, the results of the comparison with UZR. As the DRA-UZR-DM analysis in Part II shows, "assists only" DRA ratings at first match up well with UZR-DM. Second, when I *excluded* EUPO-3 from estimated team infield fly outs (i.e., *treated* them as ground balls), I found that team infield fly outs had no statistically significant correlation with descriptions of the size of the foul territory in the team's ballpark (as reported on the ballparks.com website), but when I *included* EUPO-3 among estimated infield fly outs, there was almost always a very significant correlation with foul territory. (In general, very large foul territories increase infield fly outs by about 15 per season.) This either/or analysis strongly suggests (practically proves) that most unassisted putouts at first base are fly balls, or at least that most of the *variation* in unassisted putouts at first reflects fly ball pitching and discretionary pop-up chances.

As discussed in Part I.D.1 above, infielder put outs on fly balls do not reliably correlate with fielding ability, as they often reflect discretionary chances. Then why did Bill James find that EUPO-3, when added to assists, generally correlate so well with first baseman fielding reputations? Because first base is the position where the *very* worst fielders on a team are found, and the many truly bad fielders at first will record hardly *any* EUPO-3, thus making the competent/good first basemen stand out. On the one hand, if you are a second baseman (or catcher or right fielder) who is playing with Mo Vaughn, you *know* the moment a pop-up or short fly ball is hit to the right side of the field that Mo is not going to be of any help, and you *always* prepare to take the chance and almost always take the chance. On the other hand, if you are a second baseman (or catcher or right fielder) who is playing with a confident, basically competent or good fielder such as Steve Garvey or Keith Hernandez or Don Mattingly, you figure, "He can handle it." (The fact that Garvey, Hernandez and Mattingly were also actual or *de facto* team "captains" may also have led their teammates to defer to them in taking such discretionary chances.)

The *team* isn't any better off if the play is made by the first baseman. Therefore, it is inappropriate to give first basemen credit for these chances. As is the case with other infielder fly outs, there are obviously some chances that a good fielder makes that nobody else could. Try finding them in non-zone data.

Though the "assists-only" DRA ratings for Garvey and Buckner are incorrect, I'm not sure that they're very far off. Garvey's fielding game had to be compromised by his poor arm. Most, though not all, of exceptional ground ball fielding plays at first base—e.g., balls hit into the hole between second and first, slow rollers down first, balls hit sharply down the line when there isn't a runner on first and the first baseman is playing back—require the first baseman to throw to a pitcher (or second baseman) covering the bag. If you can't throw, you're going to make fewer of those plays. As Bill has shown, Garvey started very few double plays, which also cost his team runs. Garvey's EUPO-3 totals were probably inflated by the deference paid to him by his teammates on discretionary pop-ups, as well as the fact that he played in two of the ballparks with the largest foul territories near first base: Los Angeles and San Diego. Interestingly, Garvey may have acquired more confidence in his arm during the second half of his career, as his ratings over that period are essentially average. *Win Shares* rates Buckner as a "B" fielder at first, and that rating averages in his total career performance, including his decline phase, when his knees practically immobilized him, thus suggesting he was probably a B+/A- fielder at his peak. If it were me, I would take Buckner (as a fielder, not a hitter) at his peak over Garvey at his peak.

The other ratings that might be incorrect are the essentially average rating for Don Mattingly and the slightly below average rating for Dan Driessen. I didn't see them play that much, so I don't know how the Buckner/Garvey analysis would apply to them. Don Mattingly did win a lot of Gold Gloves, but he was also very much a "Joe Morgan": low errors, a very heads-up player, a great hitter, a true team leader on a team that always attracted a lot of media attention. (By the way, before some paternity suits ruined his image, Garvey benefited from similar factors in Gold Glove voting.) If we had zone ratings back in those days, we could confirm whether or not Don recorded an above-average number of unassisted ground outs. Don's back problems might also have hurt him in the field as they eventually did at the plate. The Abstract has some nice things to say about Dan Driessen's fielding, but his Win Shares rating is only B-. I don't think his DRA rating is significantly wrong.

Having made the argument that it is better *in general* to use assists rather than assist plus EUPO-3 to rate first basemen, I do, however, want to clarify that EUPO-3 is a *very* useful statistic for identifying bad first baseman who make a lot of Buckner elections. (Bill James includes Dick Stuart (Dr. Strangeglove), Zeke Bonura and Marvelous Marv Throneberry in this category.) *Bad* first basemen *will* have low EUPO-3. But I don't see how we can draw reliable inferences about the contribution of *high* EUPO-3 to team fielding success. The statistical analyses I have performed using UZR and pop-up/foul territory information strongly suggest that it is not possible to incorporate EUPO-3 systematically into ratings.

No rating formula is perfect, at any position. At first base, all of the systems that are out there—UZR, Win Shares, DRA—are unable to address what might be the *most* important factor: preventing throwing errors by other infielders by catching bad throws. In time we may be able to use Retrosheet play-by-play data to solve this problem, using the same approach Tango developed to rate catchers. Right now it is probably impossible to separate out the throwing ability of the infielders from the catching ability of first baseman, except over the course of very long careers. So here again we need to rely upon subjective impressions.

One last point. Very extreme Buckner/Garvey cases may distort the estimates under DRA of the ground ball versus fly ball tendency of a pitching staff, for reasons you can probably guess. However, even the worst case won't cause individual third base, middle infielder or outfielder ratings to be off by more than a few runs a season. The ratings for Ron Cey, Bill Russell, and Davey Lopes, Garvey's Dodger teammates, should accordingly be discounted slightly. Of those three, only Russell's rating should be meaningfully different—he should probably rate Solid/OK, not Very Good.

H. Catcher

What a tough position, both to play and to rate. I'll admit right up front that I've taken some liberties with the data, because the inability of catchers to play without a lot of back-up support drags down the team catcher ratings per season and drastically cuts the number of 130+ seasons per catcher.

First, I'll simply show a chart of catchers rated by both Tango and DRA. To be counted in the DRA sample, a catcher had to play at least *three* seasons of at least 130 games at catcher (and no more than one before 1974). The resulting "population" was 11 catchers—nothing to bet the bank on, but enough to get the "nine degrees of freedom" usually considered the minimum necessary for running a regression analysis. (I excluded Mike Scosia from the sample because his team's performance was significantly impacted by many partial games played by his superb back-up catcher, Rick Dempsey.) I'll compare the Tango estimate of runs/148 games and the DRA *Index* for the catcher, and then the Tango estimate of career runs saved with the DRA estimate, based on the best information I could easily compile. The DRA career rating is based on the "individuated" runs-saved estimates for Carter and Sundberg, along with the team totals for the other catchers who qualified for inclusion in the sample. I've modified Tango's ratings to exclude non-PB errors, which were not shown to have a statistically significant effect on runs.

Second, in the Appendix, I'll provide "individuated" ratings for Jim Sundberg, Gary Carter, Carlton Fisk. Third, also in the Appendix, I'll provide the regular format of ratings for the Tango/DRA sample.

Catcher	Tango Runs/148 Games	DRA Index	Tango Career Runs	DRA "Career" Runs
Bob Boone	6	9	80	61

Gary Carter	11	14	147	80
Jody Davis	4	7	26	25
Bo Diaz	0	-5	0	-41
Terry Kennedy	3	-3	23	-21
Lance Parrish	6	7	62	45
Tony Pena	3	0	29	-22
Darrell Porter	2	-3	16	-19
Benito Santiago	2	1	12	-7
Jim Sundberg	13	14	151	93
Butch Wynegar	5	0	40	-2

The Index for catchers is the same as for other positions, except that it includes a five-to-three year peak (e.g., if the player only had four or three 130+ season) and includes their *first* and second best (if available) positive 146+ game seasonal ratings. The Index is thus designed to give every catcher the best opportunity to demonstrate excellence given the significant loss in playing time even during their "full-time" seasons, which results in significant dilution of their per-season value as measured by the team rating. In addition, I used the "individuated" career runs-saved estimates for Carter and Sundberg. As the "individuation" method is very laborious, I didn't get around to performing it for the other catchers.

The Index has an over 0.9 correlation with Tango's runs-saved per 148 games. Though this is obviously an "apples and oranges" comparison, it does demonstrate that DRA, even without CS/SB/WP/PB data per catcher, should enable analysts to identify truly good catchers throughout major league history.

The correlation between Tango career ratings and DRA career ratings is also comfortably above 0.9. However, as you can see just by eyeballing the data, the DRA career ratings are significantly understated, and a regression of DRA "career" ratings onto Tango career ratings indicated that about 35 runs of career value was lost on average per catcher under DRA. The regression equation was:

Tango Career Runs = 34.84 + 1.06*DRA "Career" Runs.

"Individuating" the ratings for Sundberg and Carter helped some (see the Appendix charts for catcher), even without individual SB/WP/PB data, thus suggesting that if I had done the same for the other catchers the number of "lost" runs would be lower. Sundberg's ratings were hurt by catching a lot of innings for knuckleballer Charlie Hough. When Sundberg left Texas and Hough remained, the WP/PB/BK runs (i.e., catcher runs excluding SB/"CS") were, on average, –17 for a few seasons. So I assumed that Charlie "cost" Sundberg 15 runs per 250 inning pitched /162 games caught and provided a Hough-adjustment for Sundberg, pro-rated to take into Hough's actual innings pitched and Sundberg's estimated innings caught. (By the way, this finding may put into better perspective Voros McCracken's finding that Hough and other knuckleballers have unusually good rates of preventing non-HR hits. They *have* to have better Hits Allowed rates to keep their jobs (thus resulting in "selective sampling"), because of the significant number of runs they allow by causing more WP and PB.) One

interesting finding from the individuation analysis is that Gary Carter seemed to lose his ability to cut off the running game in the season before the Expos traded him to the Mets.

In spite of these numerous caveats, DRA does manage to show the best catchers saving 15 to 20 runs in their best seasons. And without CS or individual WP/PB or SB allowed data.

I. Trends in Fielding Performance

One of the most significant findings made possible by DRA is that fielding performance has been converging towards average during the 1990s. I'll summarize these finding and then offer some explanations.

At short, Greg Gagne and Royce Clayton have had a few very good seasons, but nobody is dominating the position the way that Concepcion, Ozzie Smith and Ozzie Guillen dominated in the 1970s, '80s and '90s, respectively.

At second, Ryne Sandberg and Jose Lind had a few good seasons in the early 1990s, but nobody has had a record of sustained excellence since then (though Pokey Reese was outstanding in 1999-2000).

At center, there actually have been a few Excellent to Very Good fielders: Andruw Jones, Marquis Grissom and Steve Finley.

At third, Ken Caminiti and Robin Ventura have been Very Good, but nobody has dominated as Nettles, Schmidt, Pendleton and Bell did.

At the corners, Sosa has been Very Good, but no one has been Excellent, and there are fewer of the Solid-to-Very Good fielders, at least as measured by sustained performance over several years.

At catcher, I-Rod is at least as good as Sundberg and Carter were, but there doesn't appear to be anyone else remotely as dominant. In the 1970s, you had Bench, Yeager, Sundberg, Rick Dempsey—so many great catchers.

I think there are five reasons for this striking pattern. First, major league baseball may be more competitive than ever, and as Bill James and the late Stephen J. Gould have written, it's generally more difficult to dominate in a more competitive environment. Second, zone data on *hitters* has been publicly available for about a decade, so everybody at least in theory knows how to position themselves for optimal fielding. Third, I think Bill's introduction of Range Factor (i.e., gross plays made per game or per nine innings, not error rates) has, despite its limitations, properly focused teams on evaluating fielders more for their range than for their surehandedness. There are many fewer Larry Bowas out there bringing down the average. Fourth, since the late 1980s, teams have better data for evaluating fielders. Caught stealing data began to be collected for catchers on a

systematic basis in 1986, and zone ratings were available for purchase by major league teams starting in the late 1980s.

Perhaps the most significant reason, however, is that fielding performance is dramatically affected by injuries and physical condition, and players are generally in much better physical condition these days. As the DRA ratings in the Appendix so dramatically demonstrate, the vast majority of fielders peak in their twenties (with the possible exception of second basemen). DRA also shows how many fielders had their fielding performance harmed by injuries and weight problems. Ozzie Guillen. Ozzie Smith (his 1986 rating dipped when he hurt his arm and had not yet made the adjustments Whitey Herzog has written about). Larkin has battled injuries that have prevented him from approaching the performance in his first few seasons. Craig Biggio. Dwayne Murphy. Andre Dawson. Excess weight ruined the fielding performance of Tony Gwynn, Dave Parker and Kevin McReynolds.

Though fewer players are dominating fielding than before, I believe it is still important for teams to focus on fielding. It's true that very few players are +20 runs relative to the league average, but many staffing decisions faced by teams involve choosing between two players who differ by 20 runs or more per season in their fielding ability. *That* is still very significant.

And in terms of honoring players with MVP awards, fielding should still be counted much more than it is. Were Bonds and Pujols not having outstanding seasons, Gary Sheffield might be the MVP this year. He is *not* that good a player—his fielding is so bad that it absolutely should be considered.

IV. Miscellaneous

A. Current Practical Uses for DRA

Zone ratings are almost certainly more reliable than DRA, at least in theory, because they are based upon better quality data. Why should we use DRA for anything other than historical ratings for players who played before zone data was collected?

Because DRA provides a very simple way to double-check surprising zone ratings, and because only DRA can be used to evaluate minor league fielders.

As MGL has shown, there are many, many factors to consider in using zone data. Without having access to all of the calculations, I believe that the interaction of such factors may contribute to some of the more extreme ratings under UZR, particularly in the outfield. As shown in Part I.A, the format of DRA equations is simple and, though for the time being you'll have to take my word for it, the equations make immediate intuitive sense, after a brief explanation. If an extreme UZR result pops up (even MGL admits that some UZR ratings might be incorrect—see his comments in Part II.A.4), the

DRA result may help an analyst diagnose whether the myriad zone factors that can't be tracked under DRA were appropriately taken into account for a particular player by the zone system in question. In short, DRA can *complement* and improve zone-based ratings. In fact, it was because of my work with DRA, as well as Mike Emeigh's fielding articles, that I was able to help MGL improve the "persistency" of UZR ratings.

Zone ratings have been around for about 15 years, and there still remain many controversies in how to use the underlying data. (I believe that Bill James is currently working on this issue for the Red Sox. Given that the Red Sox owner made a nine-figure fortune as a "quant" investor and is probably paying James a considerable amount of money for his advice, there would appear to be at least two very smart people who seem to think that zone systems can be improved.) To take just one example, should players get "credit" for making plays in unusual zones? The first versions of zone ratings ignored such plays. Mike Emeigh has argued that all batted balls should be considered in evaluating fielding, and I agree with him. Chris Dial has noted, however, that a fielder who faces batters who generate unusual distributions of batted balls (remember how much team schedules differ these days), or who plays next to an exceptionally strong or weak fielder, might re-position himself to reach balls in a zone in which other fielders never have an opportunity to make plays, thus possibly resulting in a rating that is too high. (Mike and Chris: I hope I've accurately represented your positions (pun intended); please let me know if I haven't.) DRA delegates the mind-bendingly complicated task of "modeling" positioning to the *fielders*, without giving any fielder unusual credit for particular plays. We let the *fielders* sort out where they should position themselves so that, individually and collectively, they make the most plays. Their individual and collective success in making plays is captured by DRA. In a baseball environment in which well over 25% of BIP result in hits, there are plenty of opportunities for each fielder to excel.

DRA can also have a significant practical role in enabling major league organizations to maximize the value of their talent. Both UZR and DRA show that fielding performance diminishes with age even more than batting ability. DRA can help organizations identify outstanding minor league fielders (for whom zone data is unavailable) while they are still young. Thanks to Bill James (who considered it his most important single discovery), we know that it is possible to project major league *hitting* performance using minor league statistics. *Of course* there will be complications in running DRA on minor league stats, and in "projecting" major league performance—but they are almost certainly worth sorting out, particularly for small-market teams who must make the best use of their young talent before it is poached by the Yankees.

Baseball America recently published the results of an interesting study by Mike Koblish, who compiled minor league data for the last five completed seasons (1998-02). Mike measured the percentage of BIP that fall in as hits (i.e., singles, doubles or triples) at the various levels of organized minor league play, and compared that to the major league rate. His findings suggest that although there is a meaningful difference between major and minor league fielding quality, the quality of minor league performance at all but the Rookie levels might be fairly consistent, thus suggesting that aggregate minor

league records could be a fairly reliable source of information for projecting major league fielding ability:

Level	H/BIP
Majors	.302
Triple-A	.314
Double-A	.310
High A	.314
Low A	.314
Short-Season	.318
Rookie Advanced	.337
Rookie	.324

In other words, it should be possible to compile sufficiently large aggregate samples of individual fielder performance, i.e., by using the player's entire non-Rookie data, to determine a projected DRA major league rating for the player. I'm *not* suggesting that Mike Koblish's data *proves* that this can be done, but the findings are encouraging.

B. Hall of Fame Debates

At the risk of repeating myself (somewhat) in an already long article, allow me to summarize the possible impact DRA ratings could have in stirring and settling Hall of Fame debates. If I ever reveal the DRA system's details in a book, I'm very confident that, in time, DRA runs-saved numbers (or those of similar systems) will eventually have enough credibility among Hall of Fame voters to be factored in.

At shortstop, I believe that Dave Concepcion has a good case for induction by the reformed Veterans Committee. His batting numbers are roughly equivalent to Ozzie Smith's (which are surprisingly solid, especially now that we're learning that On-Base Percentage is more important than Slugging Percentage), and Dave's fielding numbers are right up there. On a "rate" basis, considering offense and defense, I'm pretty sure that he was about as good as Ozzie, and might have had a higher sustained peak value, as Ozzie's best hitting and fielding years generally did not coincide, with the exception of 1987 and 1988. Dave's outstanding fielding *effectiveness* was, somewhat unfairly, overshadowed by Ozzie's *acrobatics*. Ozzie played more, so I agree with his first-ballot election, but Dave should be included, eventually. His chances will be harmed by the recent hitting explosion among major league shortstops, but since none of the slugging shortstops these days is dominant in the field, I very much hope the voters will be able to look past that.

Larkin will get in regardless of his (good) fielding. Jeter's career hitting and baserunning exploits will probably be sufficiently great to overcome his truly Poor fielding, perhaps the worst of any shortstop with a long career. But it should be a *lot* closer than most fans appreciate. I hope Hall of Fame voters ignore Omar Vizquel's

unmerited Gold Gloves. Ozzie Guillen was a Hall-of-Fame fielder, but a minor-league hitter, even for a shortstop.

At second, Grich, Sandberg and Whittaker deserve to be in the Hall of Fame. I'm only restating the obvious because their chances have also been harmed by the recent hitting "numbers" among middle-infielders (Alomar; Kent). When you consider both their fielding and hitting ability, they are *much* better than *many* other second basemen in the Hall. Biggio belongs in spite of his below-average fielding performance.

In center, Dale Murphy's borderline case is not supported by DRA. I'm somewhat reluctant to "penalize" a player for taking on a challenging position, presumably at the request of his team, but Dale was a Poor centerfielder and not a good hitter after age 31. Andruw Jones will probably get in the Hall someday, but he won't need a DRA report. The other good fielding center fielders didn't hit enough to get into the Hall. It's too late to keep Kirby out.

At third, DRA provides additional support for Bill James' argument that Darrell Evans is the most underrated player of all time, and Hall-of-Fame-worthy by a large margin. DRA might justify the induction of Buddy Bell or Graig Nettles, but first we have to get Ron Santo in. As mentioned before, DRA provides strong support for Allen Barra's contention that Mike Schmidt (who, of course, is already in the Hall) has been significantly underrated.

In right/left field, I've already made the hopeless case for Jose Cruz. Neither Jesse Barfield nor Luis Gonzalez has the batting numbers to get in. I hope that voters think a very long time about the real and substantial damage that Gary Sheffield has done in the outfield (and, presumably, in the infield, where he started playing) over the course of his career, in the event he accumulates impressive career batting statistics. In contrast, I hope that Albert Belle is not penalized for being (unfairly) considered a poor fielder. His career-shortening injury may keep him out, anyway.

At first, Keith Hernandez's fielding had such a large impact that I believe he belongs, but Mark Grace's fielding, though good, should not get him in.

At catcher, DRA does not make a compelling case for any borderline candidate in the 1974-2001 sample. DRA might someday contribute good numbers supporting Bill Freehan (for whom a lot of CS data might still be missing).

C. Adapting DRA to Changing Pitching and Fielding Data

DRA is as much a *process* as it is a *result*. Through the use of regression analysis, we *discover* the relationships between various pitching and fielding events and runs allowed, so that we can evaluate players in their proper context. As such relationships almost certainly have evolved over the past century, I would want to re-run DRA regressions using data at different times in history before evaluating pre-1974 fielders.

There is one relationship I am certain has changed through time: the effect of left-handed pitching on BIP distributions. In general, left-handed pitchers don't "cause" ground ball opportunities to be shifted to the left side of the diamond and fly ball opportunities to the right; rather, left-handed pitchers, due to platooning and pinch-hitting, tend to *face more right-handed batters*, who shift BIP opportunities in such fashion generally *independently* of whether they are facing right- or left-handed pitchers. Since platooning has gone in and out of fashion over major league history, we should expect the *observed* relationship between left-handed pitching and BIP distributions to change as well.

Another relationship that might have changed over the course of major league history is the relationship between errors and runs allowed. As noted repeatedly, there was no statistically significant relationship in 1974-2001 between errors and runs allowed, after taking into account plays *successfully* made, except at pitcher and right field. The result at pitcher is intriguing. Since pitchers by definition can't position themselves before making a pitch and have limited mobility after making a pitch, "sure-handedness" becomes relevant, although "range" (plays made) still remains important. Before baseball gloves began to resemble Jai-Lai traps, sure-handedness had to have had more relevance in fielding success. The DRA method would allow us to find out the extent to which this was the case.

Finally, the fact that DRA is a process as much as a result will help DRA to cope with a statistical problem I discovered only days before releasing this article: we don't have separate left-, center-, and right field putout data per team (still less per individual) for seasons before approximately 1970 or so. The reason I had such data was that John Jarvis used "Retrosheet" data to calculate team putouts at left, center and right each season in his 1974-2001 database that I used in developing DRA.

Retrosheet data includes new statistics (though not zone data) gleaned from the close examination of box score records of major league games, which have been coded and saved in computer files. Retrosheet data can yield good estimates of innings played at each position by each player, CS, possibly even data evidencing runners "held" by outfielders with good arms. As mentioned in Part I.D.IV, TangoTiger has used Retrosheet data to provide excellent ratings for catchers. Retrosheet was conceived by Bill James and made possible by the efforts of countless analysts, each of whom deserves a spot in sabermetric heaven for the monk-like devotion they have shown to this immense task.

Although team-level putout data is not currently available in left, center and right for seasons before 1970, this does *not* have *any* impact on the DRA process for rating pitchers or infielders, and it *is* possible to derive reliable DRA ratings for pre-1970 players who played full-time at one or more of the outfield positions, in a manner similar to that shown in Part III. In other words, although it *might* not be possible to generate reliable DRA ratings for each and every outfielder who ever played the game (as it will be possible to generate reliable DRA ratings for every *non*-outfielder), it *is* possible to generate reliable DRA ratings for the only outfielders in distant history whose fielding

exploits any significant number of fans care about: outfielders who managed to put together five or more seasons of at least 130 games at one outfield position.

Before explaining how this can be done, let me just mention a few questions of perennial interest that DRA can answer. Was Speaker essentially as good as Cobb, after taking into account fielding? Was DiMaggio just another fielder overrated due to a by now mythic reputation for gracefulness, one that will never be sullied by re-runs of ESPN Classic televised games? Was Mays at his peak better than Mantle at his peak, after taking into account fielding? Can we compile more evidence (not that any is needed) to get Minnie Minoso into the Hall? Has Barry Bonds been greater than Ted Williams, after taking into account fielding? (As you might have guessed, I currently believe the answer to all of these questions is, "Yes.")

Now (some) detail about how this can be accomplished. Recall that in Part III we rated full-time fielders by crediting them with the *team* rating at their position. Recall that we used certain arithmetic and regression-based techniques to determine the context-adjusted plays made by the team, given various factors with statistically significant relationships to such plays.

To rate, say Mickey Mantle in 1956, all we would have to do is use DRA's arithmetic and regression techniques *backwards*, and then apply certain conservative assumptions, in the following manner:

(i) Using regression analysis results from, say, 1946 through 1960 (post-War; pre-expansion), as well as certain relationship revealed in the 1974-2001 data set, it will be possible to generate a good estimate of the number of putouts the 1956 Yankees *should* have recorded in center field if the Yankee centerfielders were completely *average*.

(ii) Then we make a conservative estimate of the number of putouts the Yankees *actually* made in center. We know that Mantle played 144 games in center, some of which for partial games, as Joe Collins and Bob Cerv played parts of 22 games in center, and 144 + 22 > 154. We also know that Mantle did not play any games in left or right and recorded 370 putouts. We assume *conservatively* that (i) Mantle was no better or worse than Collins and Cerv and (ii) Mantle *did* play 144 *full* games, so that his rate of putouts per nine-inning "game" is *understated*. We then annualize such "conservative" estimate of the Mantle putout rate to 154 games, thus estimating that Yankee centerfielders had 396 putouts.

(iii) We then multiply the difference between (i) and (ii) and the estimated run-weight for context-adjusted center field putouts in 1946-60. Such run-weight would be based on the run-weight for *all* outfield putouts (revealed by the 1946-60 regression), adjusted to put the outfield weights in the same proportion they have in the 1974-2001 data set. We would "shrink" the rating back to a 144-game rating for Mantle.

Obviously, this method yields only estimates, but they would be conservative ones, at least when we rate outfielders that are presumptively well above average (Speaker, Mays, Mantle before his knees really became a problem). We could also improve the accuracy of single-season ratings by choosing to rate players only for seasons in which they played 90% of their teams' games and shared games were below a certain level. There are very few shared games for non-injury-prone stars such as Speaker, Cobb and Mays.

I'm also working on more sophisticated methods for coping with the limited outfield data we have. I just wanted to provide a glimpse of one way we could deal with the outfielder data problem, so that DRA could be used to provide good ratings of the best fielders at every position throughout major league history, should the opportunity for such a book-length project present itself, as mentioned in the Introduction.

D. Replacement-Level Win Shares and Loss Shares

DRA provides fielding ratings by reference to league-average performance. But that does *not* mean that DRA cannot be used to provide ratings tied to Bill James' profoundly important idea that players should be rated by reference to *replacement level* performance, rather than "average" performance.

I'm going to assume that you are familiar with the concept of replacement level and why it is the correct *ultimate* reference point for player ratings.

The way to evaluate players using DRA it to (i) take their DRA rating, (ii) add it to their Linear Weights batting/baserunning runs created above or below the *league* average *without a position adjustment* and (iii) calculate the *excess* of the resulting sum over the average *combined* defensive and offensive ratings of replacement level players at *that* position, pro-rated for the portion of the season that the player played. The resulting number tells you how many runs the player created at the plate/saved in the field above the replacement level player at his position. As a practical matter, the runs-saved estimate tells you the *total* runs the team would've lost if the player had been replaced by a replacement level player for the period of time the player was in the lineup. Yes, runs saved have an infinitesimally larger impact than runs scored, but it's just that, infinitesimal, so it's acceptable to lump runs created and saved into one pool.

For example, let's say that that Joe Shortstop in a particular year had a +15 DRA (or UZR) rating and was –10 runs at the plate and on the basepaths compared to players at *all* positions in his league. Let's assume, to keep it simple, that Joe played every inning of every game. You would then have the challenging task of computing the same numbers for all shortstops in the league, annualizing those numbers, and making a judgment of what the net annualized run rate was for replacement-level shortstops. I'm pretty sure that such rate is probably something like –15 runs at shortstop. Subtract –15 from Joe's +5, and *voila*, you have Joe's rating, in terms of runs: +20.

The annualized replacement-level run rate would, of course, differ by position. But it makes sense to compare players with replacement level players at their position. Why? Because you can only replace a shortstop with another shortstop.

Note that it would be *incorrect* to calculate *separate* replacement-level "rates" for fielding and batting/baserunning, and then add the separate amounts by which Joe exceeded *each* rate. The very poor hitters who have been used to define replacement level hitting at each position are generally (though not always) above-average fielders—in other words, they haven't been replaced because they compensate in the field for what they lack at the plate. Full-time replacement-level shortstop hitting (ignoring fielding) might very well be 25 runs below the league average rate—but only because the shortstops who keep their jobs while hitting so poorly are above average fielders, even for shortstops.

How should the replacement-level "runs" rating be converted into "wins"? Multiply the runs rating by an estimate of the number of marginal wins per marginal run. Pete Palmer and others have estimated that 10 marginal runs, on average, result in one additional win, over the course of a season. Pete has formulas for customizing such rating for different run-scoring environments. Assuming Joe Shortstop played in an average run-scoring environment, his marginal wins would be +20 * (1 win / 10 runs), or 2.0 "wins" above replacement level, or "Win Shares". In other words, Joe's team would probably have lost a couple of more games if he had been injured on Opening Day and out for the season. If Joe had been below replacement level, we'd charge him with "Loss Shares".

Of course this process begs the question of defining replacement level. I'm not saying it's easy or free of all sorts of subjective considerations. In fact, it's fiendishly difficult and fraught with subjective considerations. I could write another article nearly as long as this one about the factors to consider in defining replacement level. For purposes of this article, I just wanted to outline one simple method to show that DRA is compatible with replacement level value measurements.

CONCLUSION

Thanks for reading. I hope it's been fun and informative. Please be sure to look at the Appendix showing the single season ratings for all of the players in the 1974-2001 study. It may provide the most persuasive evidence that DRA works, though I also believe that the DRA-UZR-DM comparison in Part II is compelling.

It will be a while before the concepts behind DRA are widely accepted—probably not until a major league organization decides to acquire the rights to DRA, or until such time as I publish the DRA equations in a book. The equations would clinch it for even the most skeptical reader. I'm almost concerned that publication of the equations might be a disappointment—the equations are so simple and make so much sense that the

response could very well be: "What's the big deal? *Of course* that's how pitching and fielding should be modeled!" Yet DRA is truly novel, beginning with the very idea of using regression analysis *at all* (or at least so comprehensively) on traditional pitching and fielding statistics. Though some of the techniques used in DRA have been used before (including, for example, adjusting SO for BFP and fielding plays for BIP and left-handed pitching), most of the techniques and resulting formulas are *completely* new.

Nevertheless, I am confident that DRA will someday be accepted as a satisfying solution to a longstanding problem. I began with a quote from Michael Lewis, and I'd like to close with one as well:

> "Intelligence about baseball ha[s] become equated in the public mind with the ability to recite arcane baseball stats. What [Bill] James' wider audience [has] failed to understand [is] that the statistics [are] beside the point. The point [is] understanding; the point [is] to make life on earth just a bit more intelligible; and that point, somehow, ha[s] been lost." *Moneyball*, p. 95.

The pitching and fielding statistics settled upon in the 19th century turned out not to be very well designed, and have confused baseball fans ever since, but they actually *do* have an underlying order, and if you're willing to apply some 19th century statistical techniques with the help of late 20th century computing power, you can uncover even more of that embedded order and derive valuable information. Working on DRA has been for me a wonderful opportunity to experience what a famous scientist once called "the pleasure of finding things out." I hope in reading this you've experienced some of that as well.