

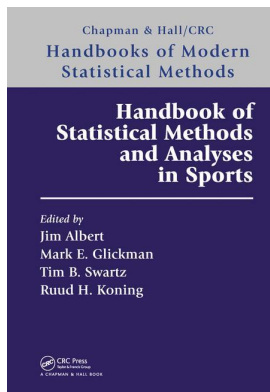
This article was downloaded by: 10.3.98.104

On: 18 Oct 2021

Access details: *subscription number*

Publisher: *CRC Press*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



Handbook of Statistical Methods and Analyses in Sports

Jim Albert, Mark E. Glickman, Tim B. Swartz, Ruud H. Koning

Defensive Evaluation

Publication details

<https://www.routledgehandbooks.com/doi/10.1201/9781315166070-3>

Mitchel Lichtman

Published online on: 21 Dec 2016

How to cite :- Mitchel Lichtman. 21 Dec 2016, *Defensive Evaluation from: Handbook of Statistical Methods and Analyses in Sports* CRC Press

Accessed on: 18 Oct 2021

<https://www.routledgehandbooks.com/doi/10.1201/9781315166070-3>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

3

Defensive Evaluation

Mitchel Lichtman

CONTENTS

3.1	Introduction.....	67
3.2	Why Defense Is So Difficult to Evaluate?.....	68
3.3	Early History of Defensive Metrics.....	70
3.3.1	Fielding Percentage.....	70
3.3.2	Range Factor.....	71
3.3.3	The First Play-By-Play or Batted Ball Defensive Metrics.....	73
3.3.4	Estimating Opportunities without Using Hit Location Data.....	75
3.3.4.1	Defensive Regression Analysis.....	75
3.3.4.2	Total Zone.....	76
3.3.4.3	With or Without You.....	76
3.4	Modern PBP Metrics.....	78
3.4.1	Ultimate Zone Rating (UZR).....	78
3.4.1.1	Parameters of Lichtman’s UZR.....	79
3.4.1.2	UZR Nuts and Bolts.....	80
3.4.2	Spatial Aggregate Fielding Evaluation.....	84
3.4.3	Other Advanced Batted Ball Metrics.....	84
3.5	The Future of Defensive Metrics.....	85
	References.....	86

3.1 Introduction

In baseball, *defense*, also known as *fielding* (technically, defense can include pitching), is one of the most controversial aspects of player evaluation. In this chapter, we will discuss exactly what we are attempting to measure when constructing a defensive metric, why it is important to any player or team evaluation system, and just how difficult it is to measure and quantify defense. We will also talk about why even the most robust modern metrics garner so little respect among some analysts and stat-friendly fans and commentators.

Attempts to measure defense have a storied history, from the traditional *fielding percentage* (FP) used in the early days of baseball (and somewhat remarkably, still used today) to the more modern and advanced play-by-play (PBP) metrics like *Defensive Runs Saved* (DRS) (Dewan, 2015), *Ultimate Zone Rating* (UZR) (Lichtman, 2010), and *Total Zone* (TZ) (Smith, 2008). In between we have metrics like Range Factor, Defensive Average, and simple Zone Rating (Basco and Zimmerman, 2010) which improved upon FP but were not

quite as good as the modern, advanced PBP metrics. In the last section of this chapter, we will briefly discuss how we might use the most recent and technologically advanced data sets like *Statcast* (Jaffe, 2014) and *Field f/x* (Sportvision, 2014) to create the putative *holy grail* of defensive metrics.

We will also see how the various defensive metrics rate two recently retired shortstops. One is a future first-ballot Hall of Famer considered by many traditionalists and baseball insiders to be an excellent defensive shortstop throughout most of his career. The other is a much less-heralded player who, for a period of seven consecutive years, was one of the best defenders in baseball history.

Not too long ago defense was all but ignored when evaluating individual players. While most people in and out of the baseball analytic industry were generally aware that defense was an integral part of a player's skill set and that it significantly impacted a team's ability to prevent runs, good methods of evaluating defense were not widely available. It was also difficult to scale defensive metrics on the same level as offense and pitching. Additionally, we were not sure how to compare players across the fielding spectrum. For example, how would a below-average shortstop compare to an above-average left fielder—that is, which one had more *position-neutral* defensive value?

Eventually, with the advent of advanced fielding metrics that use the same currency (runs) as most modern offensive stats like *linear weights* (Thorn and Palmer, 1984) or offensive WAR (Baseball Reference, 2015), and an understanding and quantification of the concept of *positional adjustments* (MacAree, 2015), one became able to combine offense, defense, and base running in order to arrive at a *total player evaluation*. In addition, we are now better able to understand the role of defense in pitching and run prevention. For example, some analysts have estimated that team wins are a combination of around 50% offense (batting and base running), 25% pitching, and 25% defense. Compare that to the traditional (and wrong) saw that “Baseball is 90% pitching,” or the careless assertion that baseball is half pitching and half hitting (with base running and defense completely left out of the equation).

3.2 Why Defense Is So Difficult to Evaluate?

Why is it that current defensive metrics, even the most advanced ones, are controversial and often mistrusted by fans, pundits, baseball insiders, and even the analysts who create them? In order to answer that, we can compare and contrast various methods of evaluating pitching, hitting, and defense. When that is done, we will see why evaluating defense is more difficult than, and in fact fundamentally different from, measuring offense and pitching, especially at the player level.

When a hitter completes a *plate appearance* (PA), he necessarily creates 1 of around 10 unique offensive events, defined by the rules of baseball and recorded by an *official scorer* (OS) (Major League Baseball, 2012). There is little or no ambiguity regarding the result of a plate appearance, at least as far as the record books and box scores are concerned. In order to evaluate a player with respect to a single PA, one can easily assign a theoretical run value to the resultant event. For example, using Pete Palmer's *linear weights* (Thorn and Palmer, 1984), the most widely accepted methodology for evaluating and quantifying offense, a single is worth around .47 more runs than a league-average PA, while a home run clocks

in at 1.40 runs above average (the exact value of each offensive event depends, in part, on the run environment—that is, the average number of runs scored per game).

To quantify the theoretical offensive value of a player, notwithstanding context such as the timing of those events (generally thought to be largely out of a player's control), all one has to do is compute the *average* run value of all those events combined. In doing so, one arrives at a single number that represents that player's theoretical offensive contribution *in runs above or below average*. This kind of metric is fairly easy to understand and *credible*.

The exact same methodology can be used for evaluating pitchers. At the end of every PA, one can credit or debit a pitcher with the marginal run value of that PA (e.g., .47 runs for a single or *minus* .26 runs for an out), just as was done for the batter. This type of metric when prorated to 9 innings or 27 outs is often called *component ERA* (ERC). As with the batter, while this number represents *theoretical* rather than *actual* run prevention, it reflects actual events for which the pitcher was at least partially responsible and which can be easily quantified.

One can also evaluate pitchers based on their runs or earned runs allowed per some number of outs or innings—for example, the traditional and ubiquitous *earned run average* (ERA) and its cousin *runs allowed per 9 innings* (RA9). Whereas ERC describes *theoretical* run prevention based on individual offensive events, and ERA and RA9 reflect *actual* runs allowed, both cases represent an accounting of actual events that occurred on the field.

What about defense? On a team level, when a plate appearance is completed and a non-HR ball is put into play, there are only three primary outcomes, from a fielding perspective: the ball can be caught and turned into at least one out, it can fall for a hit, or it can be scored as an *error*. There are other more nuanced aspects to fielding, but these three primary outcomes are the basic events that can occur as a result of a PA and which are related to the skill and performance of the fielders. As with the hitters and pitchers, there is little ambiguity in this description, so why is evaluating defense so problematic? It isn't really, *at the team level*.

There are, in fact, team defensive metrics, like *defensive efficiency rating* (DER) (Baseball Prospectus, 2015), which measure the percentage of in-play batted balls turned into outs. Unfortunately, these metrics are not particularly robust. Among other things, they do not distinguish between singles and extra-base hits, nor do they account for the location, speed, and trajectory of each ball in play. These metrics also don't account for park effects, which can significantly affect the quality and distribution of hits, outs, and errors.

What about at the player level? After all, metrics are interesting mostly because they allow us to compare one player to another and to determine the identities of the most and least valuable players. How do we take those three basic fielding events, the hit, the out, and the error, and assign them to individual players? The outs and errors are easy of course. When a fly ball is caught, one player is assigned a *putout*. When a ground ball is fielded and turned into an out, one player (occasionally more) is typically credited with an *assist*, or a *putout* if he tags the runner or base by himself without a throw. Additionally when a routine batted ball is *not* converted into an out but *should* have been *with ordinary effort*, the offending player is charged with a *fielding error*. Again, there is no ambiguity in any of these results.

However, putouts, assists, and errors alone do not allow us to effectively evaluate defense. At best, they allow us to approximate a player's defensive skill or performance. In order to properly evaluate fielding, we *must* know the number of *opportunities* in which they occur, just like we *must* know how many singles or home runs a batter or team produces *per PA or AB* (offensive opportunities) in order to properly evaluate offense. Therein lies

at least one of the problems associated with almost all fielding metrics, past and present. How do we assign those opportunities to individual fielders?

3.3 Early History of Defensive Metrics

Let's rewind the clock and see how baseball measured and evaluated defense *back in the day*, based on *what actually happened on the field* with respect to each fielder, bypassing this sticky issue of *responsibility* for batted balls that fall in for hits.

3.3.1 Fielding Percentage

As early as 1876, putouts, assists, and errors were recorded for fielders exactly as they are recorded today. The sum of those three outcomes was called *chances*, again, exactly as it is today. (Until 1887, wild pitches and passed balls were counted as errors.) So in the dawn of professional baseball, almost 150 years ago, the prevailing statistic for measuring fielding was exactly the same as it is today in the mainstream media, and among most fans, commentators, and baseball insiders. That enduring, time-honored, and crude metric is called *fielding percentage* (FP), or *putouts plus assists divided by chances* (Basco and Zimmerman, 2010).

It does not take a mathematical genius to figure out some of the flaws in that simple methodology. For example, infield putouts are awarded for catching pop flies as well as tagging runners and bases, and assists are given not only when a player fields a ground ball and retires the batter or a runner, but when he relays a throw from an outfielder, resulting in an out. A fielder is also credited with an assist when he touches a batted ball (or the ball touches him) that another fielder turns into an out.

A first baseman is awarded a putout every time he catches a throw from a fielder and the batter is out at first. Errors are given to a fielder when he muffs a play in which the batter or a runner *should have* been retired, or a batter or runner advances an extra base but for a bad (physical) play by the fielder. For outfielders, putouts are awarded on caught fly balls and assists on throws in which a runner or the batter is retired. Errors by outfielders are relatively rare and occur more often on muffed plays than by dropping a routine fly ball.

In other words, the denominator of fielding percentage is a *mishmash* of all kinds of defensive events, some of them requiring lots of skill and others not so much. The key to fielding percentage is errors. Essentially fielding percentage is the number of errors a fielder makes divided by some approximation of the number of plays in which he is involved—namely *chances*, or putouts plus assists plus errors (it is actually *one minus that number*).

For what fielding percentage actually does, which is to tell us the rate at which a fielder makes an error, it does a good job and relies on accurate information. It also only uses events that *actually occurred*, with no inferences, approximations, or subjectivity, other than whether a muffed play should be scored as a hit or an error by the official scorer. The advantage of fielding percentage is that it is easy to compute, it is simple to understand, it uses readily available data, and it is *believable* to the general public due to its simplicity and transparency.

The significant downside to fielding percentage is that it completely ignores one important aspect of fielding, *range*, which is the ability to *reach* batted balls and turn them

TABLE 3.1

Career Fielding Percentages of
Two Recently Retired Shortstops

Derek Jeter	.976
Adam Everett	.976
League-average SS (1995–2014)	.972

into outs. As it turns out, error rate, or fielding percentage, only constitutes approximately 25% of fielding talent for infielders and a much smaller percentage for outfielders. The remainder of fielding talent is the aforementioned *range* (excluding other less salient aspects of defense).

Some fielders, by virtue of their speed, agility, instincts, positioning, and jumps, are more adept at reaching and turning batted balls into outs than other less-talented fielders. For example, Andruw Jones in his prime in center field was spectacular at getting to balls hit far from his starting position, due to his speed and agility, while Manny Ramirez in left field had limited range. Yet, fielding percentage tells us almost nothing about the difference between the defensive talents of these two players. In fact, Manny averaged only one or two more errors per season than Andruw, and many of those were misplays in fielding balls off the Green Monster at Fenway, which Andruw never had to deal with. For outfielders, fielding percentage is a particularly poor measure of fielding talent.

Even for infielders, where fielding percentage represents *some* aspect of defensive skill, many players with low error totals are not very good fielders because of their limited range. Similarly, a player with a high error rate might be excellent at *getting to balls* and thus preventing hits. Good range can easily make up for a poor fielding percentage and vice versa. During the late 1990s and early twenty-first century, the two poster boys at the shortstop position for great and poor ranges (according to most analysts), Adam Everett and Derek Jeter, had identical career fielding percentages of .976 (see Table 3.1). Yet, the advanced defensive metrics, and even many of the less robust ones, suggest that Jeter was a below-average defender due to his limited range and that Everett was one of the best fielding shortstops in baseball history due to his exceptional range.

3.3.2 Range Factor

Despite its lasting popularity and ubiquitous nature, fielding percentage, which is just *error rate*, tells us almost nothing about the defensive talent of an outfielder and only a little about the run prevention ability of an infielder. Several baseball insiders recognized this as early as the nineteenth century. In the late 1880s, shortly after fielding percentage was popularized, two notable baseball figures, Henry Chadwick, historian, statistician, and perhaps the grandfather of sabermetrics, and Al Wright, a player for the Boston Braves and manager of the Philadelphia Athletics, invented what Wright called *fielding average* (rather than *percentage*). Fielding average (FA) was an early measurement of *range*. It ignored errors completely, categorizing them as hits. FA is simply “putouts plus assists divided by games played,” equivalent to the modern *Range Factor* popularized by Bill James and Pete Palmer in the 1980s. Unfortunately, Wright and Chadwick’s *fielding average* never gained any traction and it took 100 years or so before it was resurrected and renamed by James (Basco and Zimmerman, 2010).

James and Palmer's *Range Factor* (RF), like Wright's *fielding average*, is a relatively simple metric that tracks how often an individual fielder creates or participates in an out, and is the *sum of putouts and assists divided by innings or games played*. Like most of the early defensive metrics, RF reflects exactly what happened on the field with little ambiguity. For outfielders, a putout is awarded for catching a fly ball and an assist for throwing out a base runner or the batter; thus it is relatively straightforward. However, RF conflates two separate and largely independent skills, catching fly balls and throwing out runners. Also, since most OF errors are not dropped fly balls, by ignoring errors RF excludes bad defensive play by an outfielder, like muffing a hit or making a bad throw and allowing a base runner to advance an extra base.

Despite its weaknesses, a quick glance at the Range Factor leaderboards in the OF suggests that it is a pretty good measure of outfield range. For example, Paul Blair and Daren Erstadt, two of the best and quickest centerfielders of the modern era have a career RF of 3.046 and 3.033, while Matt Kemp and Carl Everett, not known for their defense, have career RFs of 2.286 and 2.217. Those numbers suggest that the speedy Blair and Erstadt made 3/4 more outs per game on defense than the less-skilled and slower Kemp and Everett.

For measuring infield defense, RF is a bit messier. As explained earlier, for infielders, assists and putouts are ambiguous. An assist can be a ground ball turned into an out or a relay throw from an outfielder that nails a runner. Occasionally it can be a ball that glances off an infielder's glove or body that is turned into an out by another fielder. It can also be awarded to a player or players involved in a successful rundown.

A putout is given to any infielder who retires a batter or runner without a throw by catching a popup or fielding a ground ball and tagging a runner or a base. Additionally, a putout is given to a fielder who receives a throw and tags a base or runner. Consequently a first baseman can amass hundreds of putouts per season with most of those plays requiring little skill. Pete Palmer at some point only used assists (i.e., he ignored putouts) for first basemen in his *Total Player Rating* (Thorn and Palmer, 1984).

For the rest of the infielders, there is still a good deal of noise in RF because of the various ways in which a fielder can be credited with a putout or an assist. Some infield plays have a large skill component, like fielding a ground ball, while other plays are relatively skill-free and routine like catching a pop fly on the infield area or receiving a throw from another fielder and tagging a base or a runner. Still, if one looks at a list of shortstop career RF, one finds players at the top of the list who were known for their outstanding defense, like Mark Belanger, Ozzie Smith, and Rey Sanchez, at 5.24, 5.22, and 5.14, per nine innings. At the bottom of the list, we see players like Derek Jeter, Tony Womack, and Hanley Ramirez, none of them known for their range, at 4.04, 4.05, and 4.17, respectively, more than one out per game worse than the great ones. [Table 3.2](#) displays the career Range Factors for our two signature shortstops, Derek Jeter and Adam Everett.

TABLE 3.2

Career Range Factors for
Two Shortstops

Derek Jeter	4.04
Adam Everett	4.63
League-average SS (1995–2014)	4.46

As one can see, despite Jeter and Everett having exactly the same career fielding percentages, .976 (see Table 3.1), Everett was almost half a “play” (putouts plus assists) per game better than Jeter. That corresponds to a savings of around .31 runs per game (the difference between an out and a hit or error is approximately .73 runs). Interestingly, even though Jeter was a much better hitter than Everett, the difference in their batting linear weights (a precise measure of offensive value) was only .27 runs per game. According to Range Factor, Everett saved more runs on defense compared to Jeter (.31) than Jeter produced on offense compared to Everett (.27). In other words, once we combine offense, defense (using RF as our defensive metric of choice), and base running (both players were excellent base runners), they were roughly equivalent players *per game* (Jeter played more than four times as many games as Everett) throughout their careers. Most baseball fans, commentators, and those who played, coached, and managed the game would find that statement hard to swallow. Perhaps Range Factor is not accurately representing these players’ defensive value. We will see shortly what the other, more advanced, metrics have to say about Jeter and Everett.

Two problems with Range Factor, for both outfielders and infielders, are that we don’t know how many opportunities each fielder has had in order to compile his putouts and assists, and we don’t know the difficulty of those opportunities. The latter issue tends to even out over large samples, especially in light of what we know about pitchers’ balls in play (BIP)—most pitchers tend to allow around the same quality of batted balls, and any fluctuations that one observes in small samples tend to be as a result of chance. However, the number of opportunities per inning or per game can vary significantly from fielder to fielder at each position depending on the pitchers that each fielder plays behind (as well as other variables).

For example, over the course of a season or even several seasons, shortstop A might play behind predominantly ground-ball pitchers while shortstop B might play behind mostly fly-ball pitchers. In this scenario, shortstop A would get more opportunities to field ground balls than shortstop B, such that even if they had exactly the same skill at fielding those ground balls, shortstop A would necessarily have more assists and thus a better Range Factor than shortstop B. The same is true for outfielders in reverse.

Also *all* fielders will get more opportunities when “pitch to contact” pitchers are on the mound than when strikeout pitchers toe the hill. Pitcher handedness affects opportunities as well. For example, if a shortstop plays behind predominantly LH pitchers (thus more RH batters), he would get more ground balls hit in his direction than if he played behind mostly RH pitchers. If one can find a way to identify or even infer opportunities, one can refine RF using actual chances rather than innings or games as the “denominator,” thus enabling analysts to create a more accurate defensive metric.

3.3.3 The First Play-By-Play or Batted Ball Defensive Metrics

Play-by-play (PBP) or *batted ball* data is typically comprised of the result of every plate appearance (as well as events that don’t end a PA, like stolen bases, wild pitches, and passed balls), including the type and location of every ball in play (some PBP databases do not include type and/or location data). Prior to the advent of PBP databases, all of the defensive metrics, including those discussed earlier, only used information available in the standard box score or statistical compilation, namely, putouts, assists, and errors. In the mid-1980s, a group of *statheads* started a program called *Project Scoresheet*, whereby volunteer “stringers” recorded the result of every PA in every game including, eventually,

the type and location of every batted ball. A few years later, a company called STATS also began recording PBP data from every game, selling and licensing this information to teams, the media, and occasionally the general public.

With the availability of this PBP data, it became relatively easy to create a metric that added an important piece to the defensive puzzle—opportunities. In fact, PBP data opened up a whole new world of offensive, defensive, and pitching metrics and greatly accelerated the pace of baseball analysis (sabermetrics) in general. Several analysts dove into this treasure trove of batted ball data, creating “zone-based” defensive metrics which, for the first time, included some semblance of *real opportunities* for fielders.

In the mid-1980s, Sheri Nichols and Pete DeCoursey invented *defensive average* (DA), which was defined as the number of ground balls for infielders and fly balls for outfielders turned into outs, divided by balls that were hit in the vicinity of each fielder and deemed “potentially catchable.” A few years later, STATS devised a similar fielding measure and published the results in their annual *STATS Scoreboard*. They called their new metric *Zone Rating* (ZR), which was similarly defined as *ground balls or fly balls turned into outs divided by all balls hit within a predefined “area of responsibility” for each fielding position*. An “area of responsibility” was defined as every location in which at least 50% of the balls were turned into outs by a player at that position (Basco and Zimmerman, 2010). One can easily see why Nichols’ *defensive average* and STATS’ *zone rating* were called “zone-based” metrics. They are outs divided by opportunities, where the denominator is all balls hit within a “zone” or area of the field surrounding each fielder and defined by the creator of the metric.

These zone-based defensive metrics were very good at the time—much better than the previous attempts to measure defense such as *fielding percentage* and *Range Factor*. They represented a quantum leap in defensive evaluation, adding the missing element of true opportunities, and disentangling the *mishmash* of assists, putouts, and errors. Most of these metrics used only ground balls for infielders and “air balls” for outfielders.

Since the boundaries of these “zones” could vary from one metric to another, a player’s ZR (e.g., .792) does not really mean anything unless one compared it to another player or to the league average at the same position. Using some simple math, one could even convert these ratios (outs divided by balls in zone) into runs saved or cost, much like the more modern defensive metrics that will be discussed later.

One issue with these metrics was what to do with balls fielded outside of a player’s zone. On the average, an “out of zone” (OOZ) out was a very good or even a great play, yet in a simple zone rating system, this type of out wasn’t counted at all—obviously a mistake. Eventually some of these metrics gave credit for a ball fielded OOO by adding it to the numerator (sometimes the denominator as well) and perhaps giving it more weight than a ball fielded in zone (IZ). Still other metrics reported two numbers—an in-zone ratio and an out-of-zone one.

Using a metric that accounts for opportunities, for example, Revised Zone Rating (RZR), [Table 3.3](#) shows that Everett is still a much better fielder than Jeter. It is easier to interpret these ratings by converting RZR ratios into runs saved or cost. Jeter averaged only 1.44 balls in zone (BIZ) per game while Everett averaged 2.14, which suggests that Range Factor really did exaggerate the difference between the two shortstops, since Jeter had many fewer opportunities than Everett. Jeter would therefore convert $(.816 - .792) * 1.44$, or .035 *fewer* balls per game into outs, as compared to a league-average SS. Everett fielded $(.871 - .816) * 2.14 = .12$ *more* balls than a league-average SS. According to RZR, Everett is only .155 outs per game better than Jeter, or .113 runs, which is only one-third of the difference one gets using Range Factor.

TABLE 3.3Revised Zone Rating^a 2003–2014 for
Two Shortstops

Derek Jeter	.792
Adam Everett	.871
League-average SS	.816

^a Revised Zone Rating (RZR) is a single ratio that combines in-zone and out-of-zone balls fielded. It is reported only since 2003.

Although assigning “zones of responsibility” within these metrics is somewhat arbitrary, and the handling of OOB plays is anything but elegant, these zone-based systems are still far better than those that do not count opportunities at all, like Range Factor. One can clearly see how RF greatly exaggerated the difference between Jeter and Everett’s fielding skill by virtue of the fact that Everett had many more opportunities *per game* to field a ground ball than did Jeter. Zone Rating accounts for this while Range Factor does not.

3.3.4 Estimating Opportunities without Using Hit Location Data

There are several excellent defensive metrics that are able to estimate opportunities without using any batted ball data at all. Three of these are Michael Humphreys’ *Defensive Regression Analysis* (DRA), Sean Smith’s *Total Zone* (TZ), and Tom Tango’s *With or Without You* (WOWY). Despite the absence of such granular information as the type, location, and speed of every ball in play, these metrics can be quite accurate, especially in large samples when these parameters tend to “even out,” and the size of the data set enables the methodology to yield substantial statistical power.

3.3.4.1 Defensive Regression Analysis

DRA and TZ ratings are similar in that they essentially estimate fielder opportunities at each position using commonly available data, although they utilize different methodologies. After normalizing some of the data to disentangle cross-correlations, DRA uses a standard regression formula to determine which of the traditional statistics are helpful in explaining run prevention. Like the more modern PBP metrics (e.g., UZR and DRS) discussed later, DRA results in a number that represents runs saved or cost at each defensive position. In tests performed by Humphreys, DRA correlates well (close to a .9 correlation) with UZR for single season values, and the standard deviation of individual player ratings, which reflects the putative ability of the metric to identify small differences in skill, is similar for both methodologies. The advantage of DRA is that it can be used with historical data where no batted ball or PBP information is available. DRA, like some of the other excellent defensive metrics over the years, never gained much popularity (Humphreys, 2005).

Table 3.4 displays career DRA ratings for our two shortstops. For any defensive metric that is presented in runs above or below average, or saved/cost, such as DRA and most of those discussed later, the value of a league-average defender at any position is *zero* by definition. DRA, which uses a regression equation to account for parks, batters, pitchers, etc., has the difference between Jeter and Everett at .22 runs per game defensively, somewhere

TABLE 3.4

Defensive Regression Analysis (DRA) for
Two Shortstops

Derek Jeter (1995–2009)	–.13 runs per game
Adam Everett (2001–2009)	+ .09 runs per game

TABLE 3.5

Total Zone (TZ) for Two Shortstops

Derek Jeter (1995–2010)	–.059 runs per game
Adam Everett (2001–2010)	.099 runs per game

between Range Factor and Revised Zone Rating. DRA also suggests that Jeter and Everett were much closer in *overall* talent than most people think.

3.3.4.2 Total Zone

Total Zone (TZ) also uses commonly available data and yields very good results. This measure is more transparent than DRA, as it does not use a multivariable regression formula. Basically, TZ uses league-wide PBP information when it is available to determine how often each batter normally makes an out on a ball hit to each position, such that it can estimate the defensive contribution of each fielder on a batter-by-batter basis. If this kind of batted ball information is not available, TZ estimates the number of outs that a batter makes to each position based on out-of-sample data using the handedness, batted ball rates, and ground/fly tendencies of the batter and pitcher. TZ is a good metric given large samples of data and can be used when no or limited PBP or batted ball data is available (Smith, 2008).

Table 3.5 displays career TZ ratings for our two shortstops. TZ is probably less accurate than DRA and tends to “shrink” extreme defensive performances toward zero, typical for a less robust metric. Consequently, TZ suggests that the difference between Jeter’s and Everett’s defensive performance per game is only .16 runs per game, around 25% less than DRA.

3.3.4.3 With or Without You

WOWY, or *With or Without You*, is an ingenious method devised by long-time sabermetric researcher and coauthor of *The Book: Playing the Percentages in Baseball* Tom Tango (Tango et al. 2006). WOWY basically accounts for the fact that not all fielders at each position get the same distribution of balls hit in their vicinity because their contextual parameters, such as parks, pitchers, batters, etc., are likely to be different, even in large samples. WOWY accounts for this by comparing the number of batted balls fielded by a particular fielder to the number fielded by all other players at that position, holding each of these parameters constant. Each data pair is weighted by the number of outs recorded by the player in question (Tango, 2008).

In Tango’s article in the 2008 *The Hardball Times Annual*, it is reported that Derek Jeter played behind 124 different pitchers in his career and from 1993 to 2007 “his” pitchers pitched in front of a total of 308 different shortstops. For each of those pitchers, WOWY takes the number of batted balls turned into outs with Jeter at shortstop and

compares that percentage with the number of balls turned into outs with all other players at shortstop (Tango, 2008).

Tango tells us that, for example, with Clemens on the mound and Jeter at short, 10.6% of all balls in play (anywhere in the park) were converted into outs by Jeter, while 12.2% were turned into outs by 22 other shortstops. So there is a 1.6% difference with Clemens on the mound, weighted by 1966, the number of BIP with Jeter on the field. The same calculation is made for every pitcher that Jeter plays behind and each difference is weighted by Jeter’s number of BIP with that pitcher. A weighted average is then calculated for all of these pitchers. In Jeter’s case, the final tally using more than 39,000 BIP while he is on the field, is 11.6% for Jeter and 12.5% for all other shortstops. That is around 38 fewer plays (or around 28 fewer runs prevented) for Jeter per season, a number that comports with the advanced defensive metrics like UZR and DRS, to be discussed later (Tango, 2008).

The same calculations are done for parks, batters, base runners, etc., and all the results are combined in a weighted average. In every case, Jeter is at the bottom of the list of shortstops, further cementing the fact that Jeter rates, defensively, as one of the worst everyday shortstops according to virtually every metric that incorporates range and not just error rate (Tango, 2008).

The attractiveness of WOWY is that it does not rely on inherently less-than-perfectly accurate and often biased estimates of the characteristics of batted balls and the inferred position of the fielders. For example, it assumes the distribution of batted balls allowed by each pitcher, as well as the position of the fielders, is essentially the same, especially in large samples, regardless of who is on the field defensively. The same assumption is made with regard to the batters, base runners, parks, and any other parameter that might affect a fielder’s catch rate other than his defensive skill. It is also quite transparent and easy for a reader to understand and accept. For example, WOWY essentially says, “Here is a large sample of batted balls; Jeter turned 11.6% of them into outs. Here is another large sample of likely similar batted balls (since they were allowed by the same pitchers, hit by the same batters, occurred in the same parks, and with the same configuration of base runners and outs); all other shortstops turned 12.5% of them into outs.” From that perspective, it is clear and believable that Jeter is probably a less mobile and thus less valuable shortstop than the rest of the field, in terms of turning batted balls into outs.

Using the WOWY methodology, we are able to quantify a fielder’s defensive performance or skill in runs above/below average by taking the difference between the “with” percentage of balls fielded and the “without” percentage, multiplying it by the average number of BIP per game, and then again by the run value of the difference between a hit near the SS position and an out (around .73 runs). If we do that for our two shortstops, Jeter and Everett, we get the results presented in Table 3.6. The downside to WOWY is that it requires a large sample of both *with* and *without you* data (the latter type of data is often lacking) in order to be meaningful; thus it is generally only useful for long careers or many seasons worth of data.

WOWY may be quite accurate in large samples such as we have for our two shortstops. If that is true, one can argue that on a *rate or per game* basis at least, Everett was the *more*

TABLE 3.6
Career WOWY for Two Shortstops

Derek Jeter	–.19 runs per game
Adam Everett	.16 runs per game

valuable player overall once offense and defense are combined. Table 3.6 shows a .35 run difference per game in defense between our two shortstops. This can be contrasted with a difference in hitting of only .27 runs per game. According to these numbers, Everett has an overall .08 runs per game advantage in skill over Jeter. Keep in mind that Jeter played many more games than Everett and thus had far more career value.

3.4 Modern PBP Metrics

The principal weakness of metrics like DRA and TZ that do not directly use batted ball data is that they don't consider the *exact* location of each ball in play, as well as its type and speed. Additionally, these metrics don't consider the speed and power of the batter, the number of outs, or the location of any base runners, in order to infer fielder positioning. These parameters can significantly influence the chance of each ball being caught. This is especially problematic for small samples of data. In larger samples, many of these variables tend to "even out," especially with those metrics that adjust for pitcher and/or batter handedness, G/F tendencies, and in some cases, such as with WOWY, the exact identities of the pitchers, batters, and parks, and configuration of the base runners.

The most popular current advanced defensive metrics, like UZR and DRS, attempt to use all of these parameters. In order to compute the results of these metrics, very granular game-level data are required. Much of that information is provided by the nonprofit group *Retrosheet* (a progeny of *Project Scoresheet*) and companies like STATS, Inside Edge, BIS, and MLBAM, the media arm of Major League Baseball. The requisite data used by most of these modern *batted ball defensive metrics* include the type, location, and speed (or "hang time") of every batted ball in play, the identity and handedness of the pitcher and batter, the park, base runners, outs, and outcome of the play, including errors that allow base runners to advance. Not all advanced defensive metrics use every single one of these parameters, and some parameters are more important than others in the development of these metrics.

3.4.1 Ultimate Zone Rating (UZR)

One of the first attempts to move away from a single zone system like STATS Simple Zone Rating and Nichols' and DeCoursey's Defensive Average was also designed by STATS and was called *Ultimate Zone Rating*. One of STATS' founders, John Dewan, expanded the ZR methodology to account for the actual difficulty of each play. The difficulty of a play, and thus the amount of credit or debit given to a fielder, was based on how often *all* players at that position fielded a similar ball hit to the same location (Basco and Zimmerman, 2008). That seems now like an obvious upgrade, but it was revolutionary at the time and spawned the modern era of advanced PBP defensive metrics.

STATS and Dewan presented their results in the 2001 *STATS Scoreboard*, but Dewan left STATS shortly thereafter and the original version of UZR went dormant. Years later, Dewan went on to develop a series of similar defensive metrics, Revised Zone Rating (RZR), Defensive Plus-Minus (PM), and Defensive Runs Saved (DRS), for his new company, Baseball Info Solutions (BIS) (Basco and Zimmerman, 2008). In 2002, Mitchell Lichtman developed his own version of Ultimate Zone Rating (UZR) using first Retrosheet data, then STATS data, and currently BIS data.

3.4.1.1 Parameters of Lichtman's UZR

The basic idea behind most of the batted ball defensive metrics is to determine the league-average catch rate at each defensive position for every batted ball that is put into play, given its type, location, and quality (e.g., speed or hang time), as well as the various parameters that can help us to infer the approximate initial position of each fielder and in some cases their ability to retire the batter based on other factors. Once that is done, for every ball put into play with Fielder A on the diamond, we can compare his result with that of an "average fielder" at his position, given the characteristics of that batted ball and the context in which it was hit.

UZR currently uses four classifications of batted balls, ground balls and bunt ground balls for infielders, and line drives and fly balls for outfielders. Note that line drives and popups on the infield are ignored, and *air balls* (any line drive, fly ball, pop fly, etc.) must be of a minimum distance in order to be included for outfielders. Keep in mind that these "type" classifications, as well as distances and locations, are according to the person or persons who record the data and are subject to human error and bias. In addition to the batted ball type, UZR uses the direction and location of every batted ball as well as a three-prong description of its relative speed (slow, medium, or fast). For direction and location, UZR splits the field up into sections rather than relying on the exact coordinates or vectors recorded by the video or "at-game" observers.

Since the initial position of each fielder affects his catch rate for any particular batted ball, one can improve the accuracy of the metric by attempting to infer that position *to some extent*. One of the primary drivers for fielder positioning is the side of the plate in which the batter stands. Most batters pull at least 2/3 of their ground balls and hit slightly more air balls to the opposite field. In addition, batted balls from RH and LH batters have different characteristics, including speed and spin, which can affect their chances of being caught at each fielding position. Thus, UZR treats left- and right-handed batters separately by creating two *buckets* for every batted ball type.

The UZR *engine* also uses outs and base runners to estimate fielding position. For example, with no runners on base or a runner on second base, we assume that, on the average, every first baseman is playing maybe 10 or 12 ft from the first base line. With a runner on first, and second base empty, however, the first baseman usually starts out on the bag and ends up at a position closer to the line. When the double play is in order, the middle infielders typically play shallower and closer to the second base bag. In a potential bunt situation, the first and/or third baseman may be playing up. With a runner on third and less than two out, some or all of the infielders may play in.

The speed of the batter is another variable that affects the infield "catch" rates on ground balls. For fast batters, the infielders must play a little shallower and are often forced to make a quicker and harder throw to first. For slow batters, they can play back, thus giving them more range, and take their time on throws to first, enabling them to turn more ground balls into outs. UZR creates three categories of batters—fast, medium, and slow—and calculates baseline catch rates separately for each.

For the outfield, batter power is another factor (besides handedness) that affects positioning. As with batter speed, UZR uses three power categories. Of course, one would prefer to know the average position of each fielder *for every batter and in every game situation* (or their actual starting position when the ball was put into play). Unfortunately, fielder positioning is not available within the traditional batted ball and PBP databases (some of them do include whether a "shift" occurred—UZR currently ignores all plays in which a shift influenced the result), so batter "hand," power, and speed must suffice.

The following “contextual” buckets are created—or adjustments are made—when determining the baseline catch rates in UZR:

- Batter hand
- Batter speed (slow, medium, fast)
- Batter power (small, medium, large)
- Base runners and outs
- Pitcher G/F ratio
- Park effects

These affect the positioning of the fielder and/or the speed, spin, etc., of the batted ball. These are the characteristics of each batted ball for which UZR creates separate buckets.

- Type (ground ball, bunt ground ball, fly ball, line drive)
- Ball speed (slow, medium, fast)
- Location (“slices” of the infield for ground balls and small sections of the OF for air balls)

3.4.1.2 UZR Nuts and Bolts

First, every batted ball is “bucketed” using some of the variables listed earlier. A ground ball is put into a bunt or nonbunt bucket and an air ball is put into a fly ball or line drive bucket. Those buckets are then subdivided into batter hand and batter speed (for infield GB) or power (for outfield air balls) buckets. We now have 36 possible ground ball buckets and an equal number of air ball buckets. In addition, there are eight buckets in the infield, representing the direction or vector of the ground ball, and several dozen outfield sections indicating the approximate landing zone of each air ball. In total, we have more than a thousand possible unique buckets. (Outs, base runners, parks, and pitcher G/F rates are not “bucketed.” Mathematical adjustments are made to the baseline catch rates for these variables.)

The UZR algorithm looks at all batted balls that fall into each of those buckets over several years of data, in each league separately (both leagues can be combined), and computes the fraction that falls for a hit as well as the average value of that hit. The percentage of outs and errors is also computed, for each of eight fielding positions (catchers are excluded). For example, a fast ground ball hit by a speedy LHB in direction X may have resulted in a hit 60% of the time, an out by the shortstop 30% of the time, an out by the second baseman 8% of the time, an error by the shortstop 2% of the time, and an error by the second baseman 1% of the time. Adjustments are made to some or all of those numbers based on the base runners, outs, pitcher G/F tendencies, and park factors.

We now have baseline league-average numbers (hit rates and position-specific out and error rates) for each of those 1000-odd buckets as well as the adjustments explained earlier. We have to be careful with the number of buckets created. If the league-wide, multiyear sample size of any one bucket is small enough, we end up introducing lots of noise into the base-line rate for that bucket. Given so many buckets, even with relatively large numbers of batted balls in each bucket, our baseline rates tend to be a little noisy anyway. Statistically, it is likely that a few of our buckets will be very noisy. Our only solace is that within a large sample of opportunities for an individual player, the noisy buckets tend to cancel one

another out, and a few spurious base-line rates (out of over 1000) won't significantly affect the final results. When we discuss SAFE, we'll introduce a powerful method developed by its creator, Shane Jensen, by which the noise associated with the location buckets can be significantly reduced by use of a *smoothing algorithm*.

Once these baseline numbers are created from several years of data (UZR uses six seasons prior to and including the season being evaluated, but it can be any number), the UZR engine goes through the database again season by season to create the results for each player or team (team UZR can be calculated using the same basic methodology, but with no regard for *which* position made an out or an error, only *whether a ball was fielded or not*). For every ball in play, it establishes the bucket to which it belongs, including of course, the most important parameters, its type and location. Then it notes the result, which can fall into three categories: a hit, an out, or the batter reaches on an error (ROE).

If the ball falls for a hit, we have to determine which fielders are going to be charged with some fraction of the run value of the difference between a hit and an out. To do that, UZR checks the baseline hit/out rate for that bucket. Every position that occasionally makes an out on an equivalent batted ball gets docked according to the proportion of outs that it makes in the league-wide database. For example, suppose a ground ball in bucket B (after applying all contextual adjustments) is normally converted into an out by all shortstops 27% of the time, and by all second basemen 9% of the time, based on the six seasons prior to and including the season in question. First, the UZR engine determines the average run value of a batted ball in bucket B by multiplying .36, the fraction that are turned into outs, by the average value of an out, which is normally around $-.26$ runs, and adding that to .64 (the hit fraction) times the average value of a hit in that bucket (depending on the proportion of singles, doubles and triples), in this case probably around .5 runs. That gives us an average *ball value*, in runs, for a batted ball in bucket B of .2264. (The calculations are displayed in [Table 3.7](#).)

On a hit, all fielders combined must be charged with a total of .2736 runs, or .5 (the average run value of a *hit* in that bucket) minus .2264 (the average run value of all balls hit into that bucket). That .2736 is divided among all fielders based on the proportion of their "responsibility." In our example, the shortstop makes $.27/.36$, or 75% of the outs, and the second baseman, $.09/.36$, or 25% of the outs. So the shortstop gets charged with $.75 * .2736$ runs and the second baseman, $.25 * .2736$ runs. For that one play, that would reflect a UZR of $-.2052$ for the shortstop and $-.0684$ for the second baseman. (The calculations are summarized in [Table 3.8](#).)

If that same batted ball were turned into an out—by the shortstop *or* the second baseman—the calculations are a bit more nuanced. First of all, when a ball is caught, no fielder is charged with negative runs, even if that fielder didn't actually make the catch. In our example, if the shortstop fields the GB, which it normally does 27% of the time, the

TABLE 3.7
Run Value of a GB in Bucket B in the UZR Rating System

Out value	$-.26$
Hit value	$.5$
Fraction of balls that are hits	$.36$
Fraction of balls that are outs	$.64$
Ball value ($-.26 * .64 + .5 * .36$)	$.2264$

TABLE 3.8

UZR Calculations for the SS and 2B on a Hit in Bucket B

Hit value	.5
Ball value	.2264
All fielders combined must be charged with the difference (.5 - .2264)	.2736
<i>SS fields 75% of balls in this bucket</i>	
<i>2B fields 25%</i>	
SS UZR is <i>minus</i> (.75 * .2736)	-.2052
2B UZR is <i>minus</i> (.25 * .2736)	-.0684

TABLE 3.9

UZR Calculations for the SS and 2B on an Out in Bucket B

Out value	-.26
Ball value	.2264
SS fields the ball, he gets <i>plus</i> (-.26 - .2264)	.4864
2B fields the ball, he gets <i>plus</i> (-.26 - .2264)	.4864
The fielder who does not field the ball	0

second baseman is *not* debited any runs. The reasoning behind that decision is twofold: one, since we don't know the *exact* location and difficulty of each batted ball or the *exact* starting position of each fielder, if a ground ball in bucket B is caught by the shortstop, we must assume that the ball was probably closer to him and further from the second baseman than the average ball in that bucket (that is a *Bayesian inference*), and two, the second baseman *may* have also had an opportunity to catch the ball. In any case, we want individual UZR's to add up at the team level, and docking one player on a caught ball without giving the other player extra (underserved) credit would be problematic. If the shortstop does indeed catch our hypothetical ball in bucket B, he gets credit for $-.26$ runs minus the average value of that batted ball, $.2264$, or $-.4864$ runs. In other words, his catch created $.4864$ fewer runs on the average, than an average batted ball from bucket B. His UZR for that one play is $+.4864$ runs. These calculations are summarized in Table 3.9. (The *signs* of the numbers can be confusing. The normal convention is that *minus* is good for the defense and *plus* is good for the offense; however, UZR for a good fielder or a good play is always presented as *plus* and for a bad fielder/play it is presented as *minus*.)

One way to test whether these computations are correct is to make sure that, for average fielders, hits and outs sum to zero for each bucket. If 100 balls were hit from bucket B, remember that 27 would be fielded by the shortstop, 9 by the second base, and 64 would fall for hits. Of the 64 hits, the shortstop would be charged with $-.2052 * 64$, or -13.1328 runs. The second base would be docked $-.0684 * 64$, or -4.3776 runs. For the 27 balls fielded by the shortstop, he would get credit for $.4864 * 27$, or 13.1328 runs. An average second baseman would field 9, for a UZR of $.4864 * 9$, or 4.3776 runs. You can see that each fielder gets a total debit/credit or UZR of exactly zero.

Remember that in UZR a fielder never gets debited any runs when another fielder catches a ball in play. You may also notice that when a fielder catches a ball in a particular bucket, he gets exactly the same amount of credit as any other fielder, regardless of

how often that position normally catches a ball in that bucket. In our example, both the shortstop and second baseman (one *or* the other) receive a credit of .4864 runs when they catch a ball in bucket B, even though the shortstop catches three times as many balls. Other similar batted ball metrics use a slightly different methodology.

A common method for crediting or debiting a fielder in many of the other advanced defensive metrics that use batted ball locations is to simply use a baseline catch rate for each bucket or location and apply that to whether a particular fielder at a position caught the ball or not. In these systems, errors are treated exactly the same as hits. In our example, if a batted ball results in a hit or error, the shortstop would be debited with .27 of a “catch” since he normally makes 27% of the outs in that bucket. The second baseman would be debited with .09 “catches.” If the shortstop makes the play, he gets credit for .73 of a catch, the difference between how often he made the catch (1 of course) and how often a league-average fielder makes the catch (.27). The second baseman, meanwhile, gets debited .09 catches, even though the ball was caught by the shortstop—the same as if the ball fell for a hit. Docking the second baseman when the shortstop makes a catch in a bucket in which the second baseman also converts balls into outs is an example of where many of the other advanced batted ball metrics differ from UZR.

Using this alternative methodology, let’s see if the numbers “add up.” In this scheme, a “catch” is worth .76 runs—the difference between an average hit in that bucket (.5) and an out (−.26). So for the 64 hits, the shortstop is debited $64 * .27 * .76$, or 13.1328 runs, the same as in UZR. The second baseman gets charged with $64 * .09 * .76$, or −4.3776 runs, again, the same as in UZR. When the shortstop makes his 27 outs, he is credited with $27 * .73 * .76$, or 14.9796 runs. At the same time, the second baseman will get debited $27 * .76 * .09$, or −1.8468 runs. When the second baseman makes the play nine times, he receives a total of $.91 * .76 * 9$, or 6.2244 runs, and the shortstop gets subtracted 1.8468 runs.

In total, the shortstop’s result is −13.1328 (for a hit), +14.9796 (for a catch), and −1.8468 when the second baseman makes the catch, for a total of zero runs. The second baseman gets $-4.3776 - 1.8468 + 6.2244$, or zero as well. So while the individual numbers vary slightly from UZR, the overall results are similar and everything adds up at the team level. Both methods are justifiable. The reason we don’t have one single, optimal methodology is because we simply don’t know the exact parameters of every batted ball, the precise starting location of each fielder, and the dynamics involved when one fielder makes a play and others do not on a ball hit in the vicinity of two or more players.

We mentioned that in some of these metrics, an error is treated in the same manner as a hit. In UZR, it is not. The reason is that we have more information on an error. A hit in a certain bucket implies that the ball was difficult to catch, on the average, regardless of the defensive prowess of the fielders and regardless of the average catch rate in that bucket. An error, on the other hand, is a ball that is deemed to be *catchable with ordinary effort*, according to the official scorer. While there is obviously some overlap between a hit and an error, and not every decision by a scorer is justifiable, clearly the average error is a much more catchable ball than the average hit. Treating a hit and an error the same is a mistake.

Table 3.10 shows that UZR, currently one of the more popular defensive metrics, is kind to Jeter compared to DRA and WOWY. The difference between Jeter and Everett, according to UZR, is once again only .14 runs per game, similar to less rigorous measures like RZR and TZ. Although UZR and many of the other modern metrics are complex and robust, they can only approximate defensive performance (and extreme values tend to over- or underrate performance), and that is why you will sometimes see fairly large differences in their outputs.

TABLE 3.10

UZR for Two Shortstops

Derek Jeter (2002–2014)	–.029 runs per game
Adam Everett (career)	.106 runs per game

3.4.2 Spatial Aggregate Fielding Evaluation

It was mentioned earlier that there is a defensive metric that uses smooth functions to model the probability of a “catch” given the various parameters of a batted ball, rather than using a discrete zone or vector-based methodology. This is an excellent solution to the problem of having relatively small samples in many of the buckets. Not only is sample size problematic with respect to these buckets but also is the notion of treating each bucket independently.

For example, let’s say that the shortstop catch rate for hard hit ground balls is 60%, 50%, 40%, 30%, 35%, 20%, and 10%, for successive vectors on the field moving further and further away from the normal shortstop position. It is likely that the 35% is an outlier due to statistical noise in that “location bucket,” particularly if the sample of opportunities is not especially large. If it is assumed that the further we go from the shortstop’s normal starting position the harder it is to field a hard hit ground ball (which is a good assumption), then we probably want to find a way to “smooth out” these numbers such that they resemble a more reasonable sequence. Ideally, one might want to create a function that estimates the catch rate for each bucket, based on the angle of the batted ball vector relative to a line from home plate to the average or inferred starting position of each fielder. That is exactly what SAFE (Spatial Aggregate Fielding Evaluation) developed by Shane Jensen et al. does.

For outfielders and infielders, SAFE estimates the league-average starting position on the field for each fielder based on the maximum catch rates for each type and speed of batted ball. From there the system self-learns using historical batted ball data to create smooth model functions based on the angle of each ground ball for infielders, and the distance and direction (from the fielder’s initial position) of air balls for outfielders (Jensen et al., 2009). This method can be adapted to any park and game situation (batter, outs, base runners, etc.) where the initial starting position of each fielder might be different from the standard one, and the characteristics of the batted balls might not be captured by the quality of the data. Like some of the other excellent metrics in their time, SAFE never gained much traction.

3.4.3 Other Advanced Batted Ball Metrics

Other batted ball metrics utilizing similar methodologies to UZR include Probabilistic Model of Range (PMR) (Pinto, 2003) and John Dewan’s Defensive Runs Saved (DRS). Unlike UZR, which tracks “arm rating” and infield “GDP turned” separately, DRS includes those aspects of defense. The methodologies for evaluating outfield arms and infield double play defense used by DRS and UZR are quite similar. For outfield arms, run values are computed by recording how often outfielders throw out runners at the various bases or prevent/allow them to advance, as compared to league averages rates, given the type and location of the batted ball, the park, the base runners, and outs. The speed of the baserunner and even the game state, such as inning and score, can be used to fine tune these numbers. Infield GDP merely credits or debits infielders according to the number of double plays turned per double play opportunity, as compared to league averages.

TABLE 3.11

DRS for Two Shortstops

Derek Jeter (2003–2014)	–.066 runs per game
Adam Everett (2003–2011)	.162 runs per game

DRS also adds (to range/errors, outfield arms and infield GDP) what they call *good fielding plays* (GFP) and *defensive misplays* (DM). GFP and DM include things like outfielders making “over the wall” catches to prevent a home run, infielders making bad relay throws, first basemen handling difficult throws, and outfielders holding runners to a single on likely doubles, doubles on likely triples, or allowing runners to advance extra bases (e.g., slowly fielding a ball, misplaying a bounce off a wall, or overthrowing a cutoff man) without being given an error. As well, DRS uses more granular buckets for air balls (such as “fliners”—a combination of a fly ball and line drive) and in some versions uses batted ball “hang time” (for air balls, and “time through the infield” for ground balls) rather than simply three categories of speed, as UZR does. Generally, DRS is a very comprehensive defensive metric based on intensive video scouting (Dewan, 2015).

DRS in runs saved/cost per game for our two shortstops are displayed in Table 3.11. DRS sees Jeter as a worse defender than does UZR, but not nearly as poor as DRA or WOWY. It estimates the difference between the two shortstops at .23 runs per game, which once again suggests that they are equivalent players overall (combining offense and defense), a notion that is anathema to most baseball purists.

3.5 The Future of Defensive Metrics

All of the batted ball defensive metrics described here, even the most comprehensive ones like DRS and UZR, are limited in their accuracy due to the difficulty in identifying and recording precise, unbiased batted ball parameters and the exact starting location of each fielder. The latter is especially problematic with teams employing more and more infield shifts of varying degrees. The *holy grail* of defensive evaluation, at least with respect to fielding ground balls and air balls and turning them into outs, requires near-perfect information on essentially two things: one, the exact location and physical characteristics of every batted ball, and two, the starting position of every fielder who fielded or could have fielded the ball in question. With that, one can either create smooth models or functions, as SAFE does, or empirically compute the league-average probability of catching each type and location of batted ball from various discrete starting locations (for each fielding position), much like UZR, DRS, and PMR do now, but with more accurate data.

This kind of detailed and precise information is now available through MLB’s advanced camera and tracking systems dubbed *Field f/x* (Jaffe, 2014) and *Statcast* (Casella, 2015). Unfortunately, at the time of this writing, the complete set of data generated from these complex and highly technological systems are not available to the general public or even nonteam affiliated researchers and writers. In addition to the location, speed, and trajectory of each batted ball, and the position of the fielders when the ball is put into play, *Statcast* provides precise measurements for such variables as the fielder’s *first step*, *acceleration*, *top speed*, and *route efficiency*.

It is not exactly clear how to incorporate this kind of information into a defensive metric. For example, if fielder A, starting at position X, converts a particular ball in play into an out, and so does fielder B for an identical batted ball and from the same starting position, do we care which fielder had the better route, first step, or speed to the ball? An out is an out, given identical contexts. On the other hand, this kind of “data-driven scouting” might allow us to reduce the uncertainty associated with results based on small samples of data.

Perhaps this new technology will also enable us to treat misses (balls that fall for a hit) differently. For example, a player who misses a ball after taking an efficient route might be given “more credit” for a noncatch than a player who takes a poor route to the same ball, all other things being equal. In other words, a “near miss” with maximum effort and skill might be treated differently than a completely botched play.

There is also the philosophical question of how much credit to give a fielder for his initial positioning. For example, if fielders A and B turn identical batted balls into outs, but fielder A does it with a faster and more efficient route, while fielder B starts out closer to the ball, do we give them equal credit? How much is fielder positioning a function of the team’s advance scouting and/or coaching skills and how much is it a function of the fielder’s defensive prowess and instincts?

One thing is clear: Given the spectacular granularity and accuracy of the data that are now available, there will soon be (or perhaps there already is in some cutting-edge front office) a method of evaluating defense that may prove to be better than any metric that has ever come to light in the history of the game. This could transform defensive evaluation from a sabermetric stepchild to a veritable *wunderkind*.

References

- Basco, D. and Zimmerman, J. (2010). Measuring defense: Entering the zones of fielding statistics. *Baseball Research Journal*, 39(1). <http://sabr.org/research/measuring-defense-entering-zones-fielding-statistics>.
- Baseball Prospectus (2015). Glossary of terms. View details for defensive efficiency rating. Retrieved from https://www.baseballprospectus.com/glossary/index.php?search=DEF_EFF, accessed May 12, 2016.
- Baseball Reference (2015). WAR explained. Retrieved from http://www.baseball-reference.com/about/war_explained.shtml, accessed May 12, 2016.
- Casella, P. (2015). Statcast primer: Baseball will never be the same. Retrieved from <http://m.mlb.com/news/article/119234412/statcast-primer-baseball-will-never-be-the-same>, accessed May 12, 2016.
- Dewan, J. (2015). The Fielding Bible Awards. Retrieved from <http://fieldingbible.com/Fielding-Bible-FAQ.asp>, accessed May 12, 2016.
- Humphreys, M. (February 14, 2005). Defensive regression analysis: Complete series. *The Hardball Times*. Retrieved from <http://www.hardballtimes.com/defensive-regression-analysis-complete-series/>, accessed May 12, 2016.
- Jaffe, J. (March 3, 2014). MLB unveils revolutionary next-level tracking technology on defense. *Sports Illustrated.com*. Retrieved from <http://www.si.com/mlb/strike-zone/2014/03/03/mlb-advanced-media-defense-tracking-sloan-sports>, accessed May 12, 2016.
- Jensen, S., Piette, J., Shirley, K., and Wyner, A. (August 14, 2009). SAFE: Spatial Aggregate Fielding Evaluation. Retrieved from <http://www-stat.wharton.upenn.edu/~stjensen/research/safe.html>, accessed May 12, 2016.

- Lichtman, M. (May 19, 2010). The FanGraphs UZR Primer. Retrieved from <http://www.fangraphs.com/blogs/the-fangraphs-uzr-primer/>, accessed May 12, 2016.
- MacAree, G. (n.d.). Positional adjustment|FanGraphs Sabermetrics Library. Retrieved November 21, 2015, from <http://www.fangraphs.com/library/misc/war/positional-adjustment/>, accessed May 12, 2016.
- Major League Baseball (2015). *The Official Rules of Major League Baseball* (2012). Chicago, IL: Triumph Books.
- Pinto, D. (September 19, 2003). A probabilistic model of range. *Baseball Musings*. Retrieved from <http://www.baseballmusings.com/archives/004765.php>, accessed May 12, 2016.
- Smith, S. (January 10, 2008). Measuring defense for players back to 1956. *The Hardball Times*. Retrieved from <http://www.hardballtimes.com/measuring-defense-for-players-back-to-1956/>, accessed May 12, 2016.
- Sportvision (2014). Field f/x. Retrieved from <https://www.sportvision.com/baseball/fieldfx%C2%AE>, accessed May 12, 2016.
- Tango, T. (2008). With and Without...Derek Jeter. *The 2008 Hardball Times Annual* (p. 147).
- Tango, T., Lichtman, M., and Dolphin, A. (2006). *The Book: Playing the Percentages in Baseball* (Chapter 1, p. 27). Florham Park, NJ: TMA Press.
- Thorn, J. and Palmer, P. (1984). *The Hidden Game of Baseball: A Revolutionary Approach to Baseball and Its Statistics*. Garden City, NY: Doubleday.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>