Systematic Review of Outcomes Reporting in Professional Baseball

A Call for Increased Validation and Consistency

Jelle P. van der List,*† MD, Christopher L. Camp,† MD, Alec L. Sinatro,† BA, Joshua S. Dines,† MD, and Andrew D. Pearle,† MD

Investigation performed at the Sports Medicine and Shoulder Service, Department of Orthopedic Surgery, Hospital for Special Surgery, New York, New York, USA

**Background:** Historically, treatment efficacy of professional baseball injuries has been determined by assessing the return-to-play (RTP) rate or using patient-reported functional outcomes scores; however, these methods may not be sensitive and specific enough for elite athletes. As a consequence, performance-based statistics are increasingly being reported in the medical literature.

**Purpose:** To (1) assess how treatment efficacy is currently reported in professional baseball players; (2) examine the variability in the reporting of these measures in terms of frequency, length of time followed, and units of measure; and (3) identify any attempts to validate these performance-based statistics.

**Study Design:** Systematic review.

**Methods:** All studies reporting treatment efficacy in professional baseball in PubMed, Embase, and Cochrane were identified. Data collected included frequency and method of reporting: RTP, functional outcomes, and performance-based statistics.

**Results:** Fifty-four studies met all inclusion criteria. Of these, 51 (94%) reported RTP, 12 (22%) utilized functional outcomes, and 18 (33%) provided baseball-specific performance-based statistics to assess treatment efficacy. Great variability was seen in how follow-up was defined (games, seasons, months), duration of follow-up, and which performance-based statistics were utilized. None of the studies validated these performance-based statistics, determined minimal time of follow-up needed, or assessed the baseline variability in these statistics among noninjured players.

**Conclusion:** Most studies reported RTP to determine treatment efficacy, but significant variability was seen in how players were followed. Similarly, great variability was noted in the type and number of performance-based statistics utilized. Additional studies are necessary to validate these measures and determine the appropriate length of time that they should be followed.

**Clinical Relevance:** This study provides a clear overview of the current methods that are used to determine treatment efficacy in professional baseball players.

**Keywords:** baseball; shoulder; rotator cuff; elbow

Injuries in professional baseball have been on the rise for quite some time.8,9,53 The most commonly injured body regions for baseball players include the shoulder, elbow, and knee; however, back, abdominal, hip, hand, and wrist injuries are also commonly reported.8,9,39 Historically, medical professionals have aimed to determine the efficacy of treatment of these injuries by assessing the rates at which these elite athletes are able to return to play (RTP). Although a uniform definition of RTP is lacking, it is generally defined as the percentage of players who are able to return to their previous levels of play.19,28,36 Other strategies used to assess treatment efficacy in professional baseball include the use of validated patient-reported functional outcome scores, such as the American Shoulder and Elbow Surgeons (ASES) score or the Kerlan-Jobe Orthopaedic Clinic (KJOC) score.57,48,65 Although these methods of assessment have provided valuable insight, they are certainly not without their limitations, especially for elite athletes. Although RTP rates are generally easy to determine, they lack sensitivity, as many players can return to the same levels of play but demonstrate inferior performance...
when compared with preinjury levels. Similarly, the high physical and functional demands of professional baseball players may diminish the sensitivity of patient-reported functional outcome scores that are mainly intended for the general population or recreational athletes. These limitations may introduce a significant ceiling effect among high-level athletes with scores that are validated for the normal population.

In recent years, there has been a dramatic increase in the development, utilization, and publication of statistics in professional baseball. In an attempt to overcome the limitations of standard outcome measures, many authors have tried to assess treatment efficacy by comparing preinjury performance-based statistics with postinjury statistics. For example, earned run average (ERA) and number of innings pitched (IP) are often compared before and after injury or treatment to determine if a pitcher has returned to his preinjury level of performance. This method of comparing pre- and postinjury (or pre- and posttreatment) performance statistics can be a valuable method to assess treatment efficacy among professional baseball players. It is important to note that these performance-based statistics (ERA, batting average [BA], IP, etc) are distinctly different from the functional outcome scores mentioned previously (KJOC score, ASES score, etc). For high-level athletes, it seems intuitive that the performance-based statistics may prove to be a more precise tool for determining whether they have returned to their preinjury performance levels than RTP rates (which are likely not sensitive enough) or functional outcome scores (which are not typically validated for elite athletes).

As functional demands, musculoskeletal injuries, and frequencies of surgical intervention continue to grow for these athletes, a better understanding of postinjury and postintervention performance is warranted. With the recent introduction of novel methods of assessing recovery and treatment efficacy in professional baseball players, it is important to understand which methods are currently utilized, how frequently these are reported, and the amount of variability used to define the measures. Although extensive work has gone into validating functional outcome scores and measures (eg, KJOC score), it is unclear if these newer, performance-based statistics have been validated. Therefore, this systematic review was performed to better understand the current status of performance outcome reporting on baseball players in the medical literature. Specifically, the purposes of this work were to (1) assess how treatment efficacy is currently reported in professional baseball players (RTP, functional outcomes, and performance-based statistics); (2) assess the variability in the reporting of these measures in terms of frequency, length of time followed, and units of measure; and (3) identify any attempts to validate these performance-based statistics.

Ultimately, validated performance-based statistics are needed to assess medical outcomes in professional baseball players. Although accomplishing this may be difficult given the vast number of statistics utilized and the significant variability in injury patterns, risk factors, and RTP rates based on player position, it is our hope that this review will serve as a first step in this important process.

METHODS

The PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-analyses) were followed for this systematic review. No funding was received.

Search Strategy

A systematic electronic query of the PubMed, Embase, and Cochrane Library databases was performed on April 4, 2016, to identify all studies reporting on RTP to professional baseball in any fashion. For the preliminary search, the search algorithm "major league baseball OR professional baseball OR elite baseball" was used. To maximize the capture of literature, the algorithm was not limited by study design, language, publication year, or specific injury terms. After removing duplicates of the obtained articles, 2 authors (J.P.L. and C.L.C.) independently reviewed the titles and abstracts of all studies. The full text was read for all studies meeting inclusion criteria. To reduce the chance of missing eligible studies, the references of all included full-text articles were scanned for additional eligible studies. Final determination of inclusion or exclusion of each study was agreed on by all participating team members.

Inclusion criteria consisted of studies (1) that reported outcomes of injuries in professional baseball (including Major League Baseball [MLB] and Minor League Baseball) and (2) that were at least evidence level 4 studies. Studies were excluded when they (1) did not report injuries; (2) reported RTP of other sports or amateur baseball; (3) reported only epidemiology or days missed after injuries without RTP; or (4) were reviews, case reports, podium presentations, or abstracts.

Risk-of-Bias Assessment

The evidence levels of all included studies were determined with the adjusted Oxford Centre for Evidence-Based Medicine's 2011 levels of evidence. Since no randomized clinical trials were identified in the final included studies, the MINORS criteria (methodological index for nonrandomized studies) were used to assess the methodological quality of studies. This tool has 8 criteria to assess the methodological quality of noncomparative studies and 4 additional criteria for assessing the methodological quality of comparative studies. Each criterion is given 2 points if it is reported and adequate, 1 point if reported but inadequate, and 0 points if it is not reported.

Data Extraction

Data for all studies included author names, year of publication, injury diagnosis, treatment for injury, number of players included, position or role in game (eg, pitcher, batter, fielder), report of RTP, functional outcomes reported, and performance-based statistics. All studies were evaluated to determine if RTP was reported. If so, further analysis of the criteria used to define RTP (ie, previous level,
any level, other criteria) and the interval in which players
were followed (eg, seasons, weeks/months, games) was also
analyzed. Similarly, the frequency in which functional
scoring systems and/or performance-based statistics were
reported was studied. The Conway Scale was not included
in the functional outcomes analysis, because this scoring
system primarily reports the athletes’ RTP rates.\textsuperscript{10} For
studies utilizing performance-based statistics, the follow-
ing variables were studied: which statistics were included
(eg, ERA, IP), how long they were followed, the unit of
measure for the length of follow-up (eg, games, innings,
weeks, seasons), and whether these statistics were com-
pared with preinjury statistics or with matched-pair con-
control groups.

Statistical Analysis

Statistical analysis was performed with SPSS 21.0 (IBM
Inc). All measures of player demographics, RTP, functional
outcomes, and/or performance-based statistics are reported
through descriptive statistics, such as mean $\pm$ SD, range,
and medians. Fisher exact test was used to assess differ-
ence in frequency of reported statistics between shoulder
studies and elbow studies.

RESULTS

Literature Search

After removal of duplicative works, 753 studies were ini-
tially reviewed by title and abstract. Of these, 82 were
reviewed by full text. A total of 54 studies reported RTP
rates, functional outcomes, and/or performance-based sta-
tistics and were included in this systematic review.\textsuperscript{\dag}
Fifty-one studies (94\%) reported RTP rates,\textsuperscript{\S} while 12
(22\%) reported functional outcomes.\textsuperscript{k} Of the 54 studies,
15 reported performance-based statistics for pitchers,\textsuperscript{\{}
2 reported them for hitters,\textsuperscript{21,68} and 1 reported on all players
stratified by position.\textsuperscript{17} A flowchart of inclusion and exclu-
sion of studies is shown in Figure 1.

Methodological Quality

There were 2 evidence level 2 studies,\textsuperscript{14,25} 14 with evidence
level 3,\textsuperscript{8} and the remaining 38 studies were evidence level
4.\textsuperscript{\textsuperscript{**}} For all included studies, the methodological quality of
level 4 noncomparative studies was a mean 10.4 (range: 9-
12) out of 16 possible points (64.8\% of maximum score). For

\textsuperscript{\dag}References 2-7, 10-18, 20-26, 30-35, 37, 40-52, 54-60, 64, 65, 67.

\textsuperscript{\S}References 2-7, 10-18, 20-26, 29-34, 36, 37, 39-51, 53-59, 63, 64,
66, 67.

\textsuperscript{k}References 5, 14, 18, 37, 44, 47-49, 51, 54, 58, 65.

\textsuperscript{\{}References 6, 12, 20, 25, 32, 34, 35, 38, 40-43, 46, 55, 56.

\textsuperscript{8}References 15, 17, 20, 30, 32, 35, 40-43, 46, 47, 55, 68.

\textsuperscript{**}References 2-7, 10-13, 16, 18, 21-24, 26, 31, 33, 34, 37, 38, 44, 45,
48-52, 54, 56-60, 64, 65, 67.

Figure 1. Flow diagram of inclusion and exclusion of studies.

level 2 and 3 comparative studies, the mean score for meth-
dological quality was 17.8 (range: 15-21) out of 24 possible
points (74.2\% of maximum score).

Return to Play

To define RTP, 49 (91\%) used the criterion of return to pre-
vious level (MLB) (Table 1). In 23 of these studies, players
were followed for 1.1 $\pm$ 0.3 seasons (range: 1-2, median: 1),
while 6 studies followed players for 8.8 $\pm$ 11.3 games
(range: 1-30, median: 5.5) to determine if they returned
to their previous levels of play. Twenty studies did not
specify how they determined that players returned to their
previous levels of play. Two studies used a different
method of determining RTP.\textsuperscript{22,48} Neuman et al\textsuperscript{48} defined
RTP by asking players their perceptions of their current
levels of play as a percentage of their preinjury levels.
The researchers did not define how long or for how many
games patients were followed. Fedoriw et al\textsuperscript{22} deemed
that pitchers had successfully returned to play if they dem-
strated an ERA within 2.00 of the preinjury level and
walks plus hits per IP (WHIP) within 0.500. Batters were
considered to have returned to play if the BA was within
0.100 of the preinjury level. These patients were followed
for 1 season.

Functional Outcomes

Twelve studies (22\%) utilized previously published func-
tional outcomes to assess treatment efficacy in MLB play-
ers. The most commonly reported scoring systems were the
KJOC and ASES scores (5 and 4 studies, respectively).
Other commonly used functional outcome scoring systems
included the Timmerman-Andrews Score and the Athletic
Shoulder Outcome Rating Scale (Table 1).

Performance-Based Statistics

An increase in studies publishing performance-based sta-
tistics was noted over time, with no studies published
before 2007, 2 studies between 2007 and 2009, 5 studies
between 2010 and 2012, and 10 studies between 2013 and 2015 (2 studies were published in the first 3 months of 2016) (Table 2). All studies that reported performance-based statistics compared them with the players’ own pre-injury statistics. Sixteen studies of pitchers reported on a total of 1063 patients. Pre- and postinjury statistics were reported for medial ulnar collateral ligament (MUCL) reconstruction (6 studies), MUCL revision surgery (3 studies), glenoid labral tear repair (2 studies), rotator cuff repair (2 studies), disk herniation (2 studies), and all shoulder injuries (1 study).

For these pitchers, the most commonly reported statistics were ERA (16 of 16, 100%), IP (16 of 16, 100%), and WHIP (14 of 16, 88%) (Table 3). After injury, ERA and IP were followed for 2.6 ± 0.7 seasons (range: 1-3, median: 3) in 14 studies, 67 months in 1 study, and 45 games in 1 study. WHIP was followed for 2.6 ± 0.7 seasons (range: 1-3, median: 3) in 14 studies. Strikeouts per game were reported in 9 studies (56%) and followed for 2.6 ± 0.8 seasons (range: 1-3, median: 3) in 7 studies, 67 months in 1 study, and 45 games in 1 study. Walks per game (38%), games played (38%), pitch velocity (31%), total pitches (25%), percentage fastballs (25%), wins (25%), win percentage (19%), strikeout:walk ratio (19%), and losses (19%) were other commonly reported statistics. No statistical difference in frequency of reported statistics was noted between shoulder studies and elbow studies (for all statistics, P > .40) (see Appendix Table, available online).

For these pitchers, the most commonly reported statistics were ERA (16 of 16, 100%), IP (16 of 16, 100%), and WHIP (14 of 16, 88%) (Table 3). After injury, ERA and IP were followed for 2.6 ± 0.7 seasons (range: 1-3, median: 3) in 14 studies, 67 months in 1 study, and 45 games in 1 study. WHIP was followed for 2.6 ± 0.7 seasons (range: 1-3, median: 3) in 14 studies. Strikeouts per game were reported in 9 studies (56%) and followed for 2.6 ± 0.8 seasons (range: 1-3, median: 3) in 7 studies, 67 months in 1 study, and 45 games in 1 study. Walks per game (38%), games played (38%), pitch velocity (31%), total pitches (25%), percentage fastballs (25%), wins (25%), win percentage (19%), strikeout:walk ratio (19%), and losses (19%) were other commonly reported statistics. No statistical difference in frequency of reported statistics was noted between shoulder studies and elbow studies (for all statistics, P > .40) (see Appendix Table, available online).

Validation of Performance-Based Statistics

In the 18 studies reporting performance-based statistics, none mentioned validation of the statistics. Similarly, none of the studies determined minimal time of follow-up required for these statistics, baseline variability of these statistics in noninjured players, or minimal clinically important differences. However, 12 of 18 studies (67%) did utilize matched-pair control group analysis to better understand the effect of injuries and treatment (Table 2).

DISCUSSION

Historically, RTP and functional outcomes have been used to assess treatment efficacy for professional baseball players. Since these tools may not be sensitive or specific enough for these high-level athletes, baseball-specific performance-based statistics are increasingly being used.
The purpose of this review was to determine how treatment efficacy is assessed for injuries of professional baseball players, quantify the variability in how these measures are reported, and identify and attempt to validate these statistics. Ultimately, 92% of studies reporting outcomes utilized RTP rates, 20% provided previously published functional outcomes, and 33% used performance-based statistics to assess treatment efficacy among professional baseball players. With regard to performance-based statistics, there was significant variability in the type of statistics reported, the length of time they were followed, and the units of measure for that time. No attempts to validate these outcome measures were identified, although 67% of studies utilizing performance-based statistics did perform matched-pair analyses.

A number of limitations in this study are worth noting. First, this is a review of all outcomes reporting in professional baseball and does not focus on a specific position or type of injury. This may contribute to the variability of time with which the athletes were followed. The primary goal of this study, however, was to assess not the differences in length of follow-up but the different methods being used to define the RTP and functional outcomes of professional baseball players. A secondary goal was to assess any attempts to validate performance-based statistics (ERA, BA, etc) but not functional outcome scores (KJOC score, ASES score, etc), as these scores are generally validated before their implementation. Additionally, this study did not analyze the efficacy or validity of pre- and postinjury performance-based outcomes. Therefore, no conclusion can be drawn on which statistical measures are more accurate and clinically relevant. Accordingly, future studies are necessary to validate these performance-based statistics, to define which statistics are most valuable in determining return to previous level of play for different injuries, and to determine the minimal time frames that these statistics should be followed. Doing so will likely prove to be a complex endeavor, as different positions and tasks within baseball (ie, pitchers, hitters, fielders, base runners) require unique physiologic demands that portend different injury profiles, risk factors, and time out of play. Successful performance is measured by a multitude of unique statistics that are specific to that position or task (eg, ERA is not relevant for fielders, just as BA is not relevant for most pitchers). These factors will each have to be taken into consideration, and ultimately, a number of position- and injury-specific statistics will need to be identified. Distinction will also have to be drawn between general statistics (eg, games played and at bats) and specific measures of performance (eg, BA and ERA).

The percentage of players able to successfully return to sport and their previous levels of play is an important and valuable measure to assess efficacy of treatment and recovery. This is especially important for highly motivated athletes. In this study, nearly all studies (91%) used the criterion of returning to previous level of play (MLB level) to define RTP rates after treatment of baseball injuries. However, RTP may not be specific enough for high-level athletes. Fedorow et al recently looked at professional baseball players with superior labral tears and reported RTP rates via 2 methods. In the first, they used the general method of classifying patients as non-RTP when they did...
In the second method, however, they used more specific measures by assessing return to prior performance. With this method, they classified players as failing to return to prior performance when (1) they did not return for 1 full season, (2) they did not return to the previous level (MLB), (3) they changed from being a starting pitcher to a relief pitcher, or (4) their postinjury statistics were not similar to their pre-injury statistics (ERA difference of 2.00 or WHIP difference of 0.500 for pitchers and a BA difference of 0.100 for hitters). Interestingly, the authors noticed that despite an RTP rate of 62% after superior labral tears, only 27% returned to “prior performance level” according to these criteria. This raises the question of which outcome (RTP or return to prior performance) can more accurately and validly assess treatment efficacy. Although only 2 pitching statistics were used, the authors should be applauded for their work. This is, to our knowledge, the first and only study that has integrated pitcher statistics as criteria for return to previous level of play. Furthermore, the authors showed that these outcomes were useful in assessing treatment efficacy among professional baseball players. It is however, not without limitations. For instance, the statistics are not validated for this use; the minimal clinically important differences remain unknown (largely because the authors chose performance-based statistical values

<table>
<thead>
<tr>
<th>Baseball Statistic</th>
<th>Abbreviation</th>
<th>No. of Studies (%)</th>
<th>Unit of Follow-up</th>
<th>Length Followeda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earned run average</td>
<td>ERA</td>
<td>16 (100)</td>
<td>Seasons 14</td>
<td>2.6 0.7 1-3 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Months 1</td>
<td>67 67 45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Games 1</td>
<td>45 45 45</td>
</tr>
<tr>
<td>Innings pitched</td>
<td>IP</td>
<td>16 (100)</td>
<td>Seasons 14</td>
<td>2.6 0.7 1-3 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Months 1</td>
<td>3 3 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Games 1</td>
<td>45 45 45</td>
</tr>
<tr>
<td>Walks plus hits per IP</td>
<td>WHIP</td>
<td>14 (88)</td>
<td>Seasons 14</td>
<td>2.6 0.7 1-3 3</td>
</tr>
<tr>
<td>Strikeouts per game</td>
<td>K/9</td>
<td>9 (56)</td>
<td>Seasons 7</td>
<td>2.6 0.8 1-3 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Months 1</td>
<td>67 67 45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Games 1</td>
<td>45 45 45</td>
</tr>
<tr>
<td>Walks per game</td>
<td>BB/9</td>
<td>6 (38)</td>
<td>Seasons 4</td>
<td>2.5 1.0 1-3 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Months 1</td>
<td>67 67 45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Games 1</td>
<td>45 45 45</td>
</tr>
<tr>
<td>Games played</td>
<td></td>
<td>6 (38)</td>
<td>Seasons 5</td>
<td>2.6 0.9 1-3 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Months 1</td>
<td>67 67 45</td>
</tr>
<tr>
<td>Pitch velocity</td>
<td></td>
<td>5 (31)</td>
<td>Seasons 5</td>
<td>2.4 0.9 1-3 3</td>
</tr>
<tr>
<td>Total pitches</td>
<td></td>
<td>4 (25)</td>
<td>Seasons 4</td>
<td>2 1.0 1-3 2.5</td>
</tr>
<tr>
<td>Percentage fastballs</td>
<td></td>
<td>4 (25)</td>
<td>Seasons 4</td>
<td>2 1.0 1-3 2.5</td>
</tr>
<tr>
<td>Wins</td>
<td></td>
<td>4 (25)</td>
<td>Seasons 4</td>
<td>3 0.0 3-3 3</td>
</tr>
<tr>
<td>Win percentage</td>
<td></td>
<td>3 (19)</td>
<td>Seasons 3</td>
<td>3 0.0 3-3 3</td>
</tr>
<tr>
<td>Strikeout to walk ratio</td>
<td>K/BB</td>
<td>3 (19)</td>
<td>Seasons 3</td>
<td>2.7 0.6 2-3 3</td>
</tr>
<tr>
<td>Losses</td>
<td></td>
<td>3 (19)</td>
<td>Seasons 3</td>
<td>3 0.0 3-3 3</td>
</tr>
<tr>
<td>Fielding independent pitching</td>
<td>FIP</td>
<td>2 (13)</td>
<td>Seasons 2</td>
<td>3 0.0 3-3 3</td>
</tr>
<tr>
<td>Pitches per plate appearance</td>
<td>P/PA</td>
<td>2 (13)</td>
<td>Seasons 2</td>
<td>2.5 0.7 2-3 2.5</td>
</tr>
<tr>
<td>Batting average against per game</td>
<td>BAA/9</td>
<td>2 (13)</td>
<td>Seasons 2</td>
<td>2.5 0.7 2-3 2.5</td>
</tr>
<tr>
<td>Batting average against</td>
<td>BAA</td>
<td>2 (13)</td>
<td>Seasons 2</td>
<td>2 1.4 1-3 2</td>
</tr>
<tr>
<td>Strikeouts</td>
<td>K</td>
<td>2 (13)</td>
<td>Seasons 2</td>
<td>3 0.0 3-3 3</td>
</tr>
<tr>
<td>Pitches per inning</td>
<td>P/N</td>
<td>2 (13)</td>
<td>Seasons 2</td>
<td>2 1.4 1-3 2</td>
</tr>
<tr>
<td>Home runs per game</td>
<td>HR/9</td>
<td>2 (13)</td>
<td>Seasons 2</td>
<td>2 1.4 1-3 2</td>
</tr>
<tr>
<td>Hits</td>
<td>H</td>
<td>2 (13)</td>
<td>Seasons 2</td>
<td>2 1.4 1-3 2</td>
</tr>
<tr>
<td>Saves</td>
<td></td>
<td>2 (13)</td>
<td>Seasons 2</td>
<td>3 0.0 3-3 3</td>
</tr>
<tr>
<td>Percentage in strike zone</td>
<td></td>
<td>2 (13)</td>
<td>Seasons 2</td>
<td>2.5 0.7 2-3 2.5</td>
</tr>
<tr>
<td>Percentage strike</td>
<td></td>
<td>2 (13)</td>
<td>Seasons 2</td>
<td>2 1.4 1-3 2</td>
</tr>
<tr>
<td>Runs against per game against average pitchers</td>
<td>RA/9avg</td>
<td>1 (6)</td>
<td>Seasons 1</td>
<td>3 3 3</td>
</tr>
<tr>
<td>Runs against per game</td>
<td>RA/9</td>
<td>1 (6)</td>
<td>Seasons 1</td>
<td>3 3 3</td>
</tr>
<tr>
<td>Inning pitched per game</td>
<td>IP/9</td>
<td>1 (6)</td>
<td>Seasons 1</td>
<td>3 3 3</td>
</tr>
<tr>
<td>Walks</td>
<td>BB</td>
<td>1 (6)</td>
<td>Seasons 1</td>
<td>3 3 3</td>
</tr>
<tr>
<td>Home runs</td>
<td>HR</td>
<td>1 (6)</td>
<td>Seasons 1</td>
<td>3 3 3</td>
</tr>
<tr>
<td>Hits per game</td>
<td>H/9</td>
<td>1 (6)</td>
<td>Seasons 1</td>
<td>3 3 3</td>
</tr>
<tr>
<td>Shutout</td>
<td>SHO</td>
<td>1 (6)</td>
<td>Seasons 1</td>
<td>3 3 3</td>
</tr>
<tr>
<td>Runs above replacement</td>
<td>RAR</td>
<td>1 (6)</td>
<td>Seasons 1</td>
<td>3 3 3</td>
</tr>
<tr>
<td>Losses (%)</td>
<td></td>
<td>1 (6)</td>
<td>Seasons 1</td>
<td>3 3 3</td>
</tr>
<tr>
<td>Complete games played</td>
<td></td>
<td>1 (6)</td>
<td>Seasons 1</td>
<td>3 3 3</td>
</tr>
<tr>
<td>Seasons played</td>
<td></td>
<td>1 (6)</td>
<td>Months 1</td>
<td>67 67 67</td>
</tr>
</tbody>
</table>

aBlank cells indicate not applicable or not available.
that they thought were representative, though not evidence based); the baseline variability in these statistics among uninjured players is unclear; and the move from a starting to a relief role does not always indicate a decline in performance. Future studies are necessary to validate these statistics and determine which are the most valuable and least confounded for determining how players respond to injuries and the associated interventions. This is of particular interest in contemporary baseball, as the utilization of performance-based statistics has increased significantly among coaches, scouts, general managers, and front office personnel in MLB in recent years.

Similar to RTP, functional outcomes may not be specific enough for high-level athletes and are generally validated only for the normal population or recreational athletes.1,27,62 In this review, 12 studies utilized 9 scoring systems, of which most were joint-specific scores validated for only the normal (nonelite athlete) population (Table 1).27,61,66 Alberta et al1 also identified that current scoring systems for shoulder and elbow injuries may not be sensitive to subtle changes in performance of high-level athletes; therefore, they developed and validated a functional assessment tool for upper extremity in the overhead athlete. Specifically, the authors designed the 10-item KJOC questionnaire and found that it had high correlation with the DASH score (disabilities of the arm, shoulder and hand) and was able to stratify overhead athletes by injury category. In a second study by this group,14 the KJOC was validated for athletes undergoing MUCL reconstruction. The authors found that this score was sensitive in detecting small changes after MUCL reconstruction, which suggests that these types of “athlete-specific” functional outcome scores may be used in the future to assess treatment efficacy for professional baseball players.

Eighteen studies were identified that reported performance-based statistics after injury and/or treatment, and the majority of these were published over the last 3 years. In review of Tables 2-4, it becomes clear that great variability existed between the methods and duration of follow-up to determine if players returned to previous performance. A significant portion of this variability is likely attributable to the different pathologic conditions in the studies, such as glenoid labral repairs versus MUCL reconstruction surgery or disk herniation treatment. However, significant variability remains when the most commonly studied surgical intervention is examined in isolation: MUCL reconstruction. Interestingly, these methodological differences resulted in disparate conclusions on treatment efficacy while the data were seemingly similar in several studies. For example, the interval of collecting preinjury statistics may significantly influence the interpretation of data and conclusions on treatment outcomes. Keller et al35 reported an 87% RTP rate (to MLB) after MUCL reconstruction in MLB pitchers and looked at several pitching-specific performance statistics. They concluded that although MUCL reconstruction allows most players to return to MLB baseball, there is a statistically significant decline in pitching performance when compared with the 3 years preceding surgery. Erickson et al20 reported an 83% RTP rate after MUCL reconstruction in MLB pitchers. More important, however, they also looked at the pitching performance and noticed that patients had statistically significant better pitching performances after returning versus their 1-year presurgery statistics. Differences in the interpretation of results can likely be explained by the differential duration of preinjury statistics included. Several studies have shown that preinjury statistical performance may drop the year before ulnar collateral ligament reconstruction, possibly

<table>
<thead>
<tr>
<th>Performance-Based Outcomes</th>
<th>No. of Studies (%)</th>
<th>Unit of Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseball Statistic</strong></td>
<td><strong>Abbreviation</strong></td>
<td><strong>Type</strong></td>
</tr>
<tr>
<td>Batting average</td>
<td>BA</td>
<td>Seasons</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weeks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Games</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Seasons</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Games</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Games</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weeks</td>
</tr>
<tr>
<td>Stolen bases</td>
<td>SB</td>
<td>Seasons</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Games</td>
</tr>
<tr>
<td>Home runs</td>
<td>HR</td>
<td>Seasons</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Games</td>
</tr>
<tr>
<td>On base plus slugging</td>
<td>OPS</td>
<td>Seasons</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Games</td>
</tr>
<tr>
<td>Home runs per at bat</td>
<td>HR%</td>
<td>Seasons</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weeks</td>
</tr>
<tr>
<td>Runs</td>
<td></td>
<td>Seasons</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weeks</td>
</tr>
<tr>
<td>Runs batted in</td>
<td>RBI</td>
<td>Seasons</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weeks</td>
</tr>
<tr>
<td>Slugging percentage</td>
<td>SLG</td>
<td>Weeks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weeks</td>
</tr>
<tr>
<td>On base percentage</td>
<td>OBP</td>
<td>Weeks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weeks</td>
</tr>
<tr>
<td>Walks</td>
<td>BB</td>
<td>Weeks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weeks</td>
</tr>
<tr>
<td>Strikeouts</td>
<td>K</td>
<td>Weeks</td>
</tr>
<tr>
<td>Plate appearance</td>
<td>PA</td>
<td>Weeks</td>
</tr>
</tbody>
</table>
because players attempt to compete in an injured state or progressively worsening of a chronic injury eventually worsens acutely and requires surgery.35,46 Erickson et al35 followed preinjury statistics for 1 year before injury and noted improved postsurgery performance, while Keller et al35 followed them for 3 seasons preceding injury and noted decreased postsurgical performance. Although we cannot know for certain, this discrepancy may play a role in the different conclusions drawn by the authors in each study. To overcome this phenomenon, several authors have suggested either excluding statistics from the year before injury or compiling preinjury data from multiple seasons.38,46

Regarding postsurgical statistics, it was also noted that different conclusions could be drawn when players were followed for different lengths of time. Ricchetti et al35 assessed pitching statistics for the 3 seasons prior to injury and after surgery for glenoid labral tears. They noted that IP returned to normal in the second postoperative season but that ERA did not return to baseline until the third postoperative season. Keller et al35 assessed pitching statistics after UCL reconstruction and compared these statistics with matched controls. They found that IP was also lower at 1 year postoperatively versus controls but that this difference was eliminated if the pitchers were followed for 2 years postoperatively. They saw a similar trend for winning percentage, while they did not find differences between groups for ERA and WHIP. Interestingly, some studies followed postsurgical performances of athletes for only a single season42,56 or a specific number of games.21,34 Based on the findings in the previously mentioned works, this reduced period of follow-up may not be sufficient to allow full recovery and return to peak performance. Future studies are needed to determine the minimal length of follow-up required for each of these now commonly reported statistical measures.

Finally, in the 18 studies reporting performance-based statistics in professional baseball players, there was no mention of validation of any of these measures. To overcome this, several studies utilized control groups to correct for some of the potential confounders. This is very important since performance-based statistics are also influenced by a multitude of factors that can be independent of physical health, such as mental acuity, position in lineup, ability of surrounding batters in lineup, pitching role, quality of competition, handedness of opposing batters or pitchers, traveling schedules, and so on. To demonstrate this important concept, we compared 2 studies reporting on pitch velocity after UCL reconstruction. Lansdown and Feeley38 found that fastball velocity was significantly decreased postoperatively (i.e., 91.3 miles per hour [mph] preoperatively to 90.6 mph postoperatively for young players and from 91.7 mph preoperatively to 88.8 mph postoperatively in older pitchers) and concluded that UCL reconstruction leads to decreased pitch velocity. Jiang and Leland32 also compared pre- and postoperative pitch velocity and noticed a decline in fastball velocity (i.e., 91.5 mph preoperatively to 89.7, 88.7, and 87.7 mph at 1, 2, and 3 years postoperatively, respectively). However, when comparing pitch velocity with that of a matched healthy control group, they found no differences pre- and postoperatively. This demonstrates the benefit of matched-pair analysis to reduce the influence of the many confounders associated with these statistics.

In conclusion, this study demonstrates that most studies reported RTP rates based on the criterion of "return to previous level of play," but these players were followed for a variable amount of time to determine if they were ever able to achieve RTP. Despite promising use of performance-based statistics, great variability is seen in which statistics are used, how long these statistics are followed, and whether they are compared with preoperative performance data. Although we believe that performance-based statistics can be valuable for assessing treatment efficacy among MLB players, minimal attempts at validation have been performed. In the future, it would be beneficial to quantify the normal variability of these statistics among healthy, noninjured players as a way of quantifying the degree in which they are confounded by factors not related to injury. Similarly, it is necessary to determine the minimal lengths of time that different statistics should be followed for specific injuries to capture the true changes over time. Finally, additional work is needed to determine the minimal clinically important differences for these statistics. Although they hold significant promise moving forward, further study is certainly warranted.

REFERENCES


For reprints and permission queries, please visit SAGE’s Web site at http://www.sagepub.com/journalsPermissions.nav.