Current Reviews in Musculoskeletal Medicine (2022) 15:283–290
https://doi.org/10.1007/s12178-022-09763-6

Check for updates

Current State of Data and Analytics Research in Baseball

Joshua Mizels[1] · Brandon Erickson[2] · Peter Chalmers[1]

Accepted: 1 April 2022 / Published online: 29 April 2022

Abstract

Purpose of Review Baseball has become one of the largest data-driven sports. In this review, we highlight the historical context of how big data and sabermetrics began to transform baseball, the current methods for data collection and analysis in baseball, and a look to the future including emerging technologies.

Recent Findings Machine learning (ML), artificial intelligence (AI), and modern motion-analysis techniques have shown promise in predicting player performance and preventing injury. With the advent of the Health Injury Tracking System (HITS), numerous studies have been published which highlight the epidemiology and performance implications for specific injuries. Wearable technologies allow for the prospective collection of kinematic data to improve pitching mechanics and prevent injury.

Summary Data and analytics research has transcended baseball over time, and the future of this field remains bright.

Keywords Baseball · Sabermetrics · Motion analysis · Machine learning · Artificial intelligence

Introduction

While deeply rooted in tradition and considered America's National Pastime [1], the baseball landscape is everchanging in today's age of big data, artificial intelligence, and machine learning [2–4]. At the professional level, Major League Baseball (MLB) has created new sources of data such that up to seven terabytes of data are gathered at each game [4]. In this review, we will provide a historical perspective on the evolution of baseball analytics, where we stand today, and emerging technologies that are on the horizon.

This article is part of the Topical Collection on *Injuries in Overhead Athletes*

✉ Peter Chalmers
p.n.chalmers@gmail.com

Joshua Mizels
josh.mizels@hsc.utah.edu

Brandon Erickson
bericksо.24@gmail.com

1 Department of Orthopaedic Surgery, University of Utah, 590 Wakara Way, Salt Lake City, UT 84108, USA

2 Rothman Orthopaedic Institute, New York, NY, USA

Sabermetrics: a Historical Perspective on Data and Analytics in Baseball

Since the 1920s, people have been trying to utilize baseball data to their advantage to predict outcomes and develop winning teams. Ferdinand Cole Lane, a biologist turned editor-in-chief of *Baseball Magazine* in 1912, developed one of the first run expectancy models in the 1920s [3, 5]. Based on how many runners are on base and how many outs there are, this model provided the probabilities that the batting team would score [3]. Afterwards, there had been many notable attempts at strategically utilizing baseball data, including the Dodgers hiring of a statistician to their front office in the 1940s [3].

The utilization of baseball statistics as we know it today did not really develop until 1974 when Dick Cramer, Bill James, and Pete Palmer co-founded the Society of American Baseball Research's (SABR) Statistical Analysis Committee [3]. The term sabermetrics, coined from the SABR acronym, refers to "the search for objective knowledge about baseball" [6]—how best to succeed in an in-game situation, determine a player's value, etc. Initially, this movement was met with resistance as a disturbance to baseball tradition. In particular, many teams rely upon professional baseball scouts, who watch hundreds of players and write reports containing their opinions as to the potential of these players. These scouts have become deeply ingrained into the fabric of baseball culture and have long argued that their assessments must be considered in concert with objective data. As a result, sabermetrics did not truly take hold until the early 2000s when the Oakland "Moneyball" A's made their fourth straight postseason appearance under the guidance of their general manager, Billy Bean [3]. Bean favored statistics like on-base percentage and slugging percentage to build a winning team with a limited budget.

Today, sabermetrics has led baseball to become one of the biggest data-driven sports worldwide. At the fan level, websites like FanGraphs, Baseballsavant, Baseball Prospectus, and Pybaseball [7] have become increasingly popular where writers analyze baseball using sabermetrics. In fact, many have been hired to work for various MLB organizations [3]. At the organizational level, teams continue to use sabermetrics to build efficient rosters and predict player performance based on a seemingly endless list of performance metrics [3, 8]. Even in orthopedic surgery and sports medicine, we use many of these same performance metrics to predict injury, outcome, and return to play [8].

Performance Metrics

There are seemingly endless sabermetrics that attempt quantify player performance [3, 9–12]. While no single metric reliably quantifies an individual player's value, these are taken into context with each other and the team [3, 12]. Many statistics are imperfect and limited by measurement error and sample size [11]. For example, defense performance has been historically measured by errors and fielding percentage [11]. An error, by definition, is failing to get an out on a routinely batted ball that is expected to result in an out [13]. However, this requires a scorer's judgement and brings in human error, as well as it does not account for "bad defense" if a player got a slow jump on a ball and did not make the play—this is technically not an error [11].

In today's age of analytics, new metrics have been created to try to solve this issue. For example, Revised Zone Rating (RZR) tries to quantify how often a fielder turn batted balls into outs—similar in theory to errors and fielding percentage [11]. This should better quantify player skill as it accounts for every play, not just ones where fielder encountered the ball [11]. However, it is limited by the difficulty or importance of the play.

Defensive Runs Saved (DRS) and ultimate zone rating (URS) build upon RZR and quantify how many runs a player prevents while playing defense [11]. This controls for difficulty of the play relative to how frequently a play is made by the entire league [11]. If a play is made 40% of the time, and the player makes the play, he gets credit for 0.6 times the run value of the play [11]. The run value of the batted ball reflects how many runs should be scored on a single play—a difficult ground ball, which may only be fielded 40% of time, may only result in the batter reaching first base without any runs scored [11]. There is limited consequence to this missed play. However, these statistics are limited by number of truly difficult batted balls hit to every fielder each year. These examples demonstrate the complexity in understanding defensive play, which has long been considered the most poorly captured with traditional sabermetrics.

Using advanced technologies, like Statcast (see discussion below), newer defensive sabermetrics have been developed to improve defensive metrics [3, 14]. For example, catch probability attempts to quantify outfield defense by measuring the difficulty of a batted ball based on how far the fielder had to go, how much time he had to get there, the direction he needed to go, and whether the proximity to the wall was a factor [14].

On the other hand, pitching and hitting have often been considered more well categorized. Generally, pitching metrics can be categorized as traditional (i.e., ERA, wins), advanced (i.e., FIP, WHIP, SIERA, BABIP), or value (i.e., WAR) [9]. They can also be categorized as workload measures (i.e., IP, G) or performance measures (i.e., ERA, WHIP, FIP). Finally, they can also be further subdivided as defense-dependent (i.e., hits, ERA) or defense-independent (i.e., strikeouts, walks, homeruns, FIP) [9].

Like defensive metrics, traditional pitching metrics are often imperfect and may not reflect the true performance of a pitcher. While pitchers may be solely responsible for an allowed homerun, the quality of their defense certainly influences the number of hits (versus making a play) or their earned run average (ERA) [15]. Field Independent Pithing (FIP), and advanced sabermetric, is an ERA estimator based on pitcher-controlled/defense-independent factors—strikeouts, walks, hit by pitches (HBP), and homeruns allowed—and adjusted for league average defense performance to balls in play and interpreted on the same scale [15]. Interestingly, FIP is a better predictor of future ERA than current ERA [15].

With the advent of Statcast, pitching metrics have also improved. Expected ERA (xERA) accounts for both type of contact (strikeout, walks, hit by pitch) and the quality of the contact (exit velocity and launch angle) [16]. This then credits either the pitcher or the hitter for the contact, which eliminates the confounding effects of ballpark, weather, or defense [16].

Finally, offensive or batting sabermetrics can be classified like those of pitching metrics. There are standard or traditional metrics (i.e., RBI, HBP), advanced (i.e., OPS, wRC), value (i.e., WAR), and Stacast (i.e., wRC+) [10, 17]. Interestingly, offensive metrics have changed with time, particularly regarding walks which have not always been valued the way they are today [18, 19]. Historically, batting average (BA) was used to quantify offensive production and is still reported today. Yet, this metric is imperfect and only counts official at-bats which excludes walks, hit by pitches, sacrifices, or catcher's interference. These can be valuable plate experiences for a team by avoiding an out and putting an additional runner on base. Sabermetrics like on base percentage (OBP), which factors in walks, and slugging percentage (SLG), which accounts for extra base hits, were developed to better quantify offensive performance than BA [20].

Prior to Statcast, weighted on-base average (wOBA) and weighted runs created plus (wRC+) were considered the best metrics to evaluate a player's offensive results [20]. wOBA, a scaled OBP, considers the weight of each offensive outcome (home runs are weighted more highly than walks) [20]. wRC+ adjusts for different ballparks and eras [20]. Now with Statcast, sabermetrics like expected weighted on-base average (xwOBA) remove defensive factors from the wOBA metric and account for the exit velocity, launch angle, and sprint speed of the batter on batted balls [21].

Table 1 provides a list of some of the most common sabermetrics and their definitions.

Statcast and Ballpark Sensors

By the 2015 season, all MLB stadiums had integrated the Statcast system to track the ball and every player on the field [2, 36]. For ball-tracking, the Statcast system utilizes Trackman phased-array Doppler radar technologies which sits behind home plate [36]. This system approximates the path, spin rate, and velocity of each pitch, as well as the initial speed, and vertical and horizontal launch angles of the batted ball [36]. While Doppler radar technology is well suited for ball-tracking, player's slower speeds make Doppler shifts too difficult to interpret. As such, Statcast uses a system of stereoscopic optical video from two arrays [36]. They are spaced 15 meters apart along the third base line, each with three high-resolution cameras that utilize optical video sensors and stereo vision techniques to allow for the precise tracking of player location on the field [2, 36]. These arrays are time synchronized with the Trackman radar data and allow for the quantification of performance metrics, such as defender reaction time, route efficiency, and speed [36].

Using sensor data, many classic sabermetric analyses of player value are becoming increasingly obsolete. For example, batting average solely considers whether a batter got a hit. With sensor data, however, one can control for other

Table 1 Examples of advanced and Statcast performance metrics

Metric	Definition
Offense	
Batting average on balls in play (BABIP) [22]	Batting average exclusively on balls hit into the field of play, removing outcomes not affected by the opposing defense (homeruns, strikeouts)
Weighted on-base average (wOBA) [23]	On-base percentage weighted by how a player reached a base (i.e., a double is worth more than a single)
Weighted runs created plus (wRC+) [24]	Normalized runs created to ballparks and eras
Wins above replacement (WAR) [25]	Value of a player in terms of how many more wins he's worth than a replacement player
Exit velocity (EV) [26]	Speed of the ball immediately after the batter makes contact
Expected weighted on-base average (xwOBA) [21]	A predicted on-base average based on exit velocity, launch angle, type of batted ball, and sprint speeds
Pitching	
Fielding independent pitching (FIP) [27]	Similar to ERA, but removes results on balls hit into the field of play (strikeouts, walks, hit-by-pitches, and homeruns)
Adjusted earned run average (ERA+) [28]	A normalized ERA, adjusted for ballparks and opponents
Pitch movement [29]	Horizontal break and vertical drop of a pitch (inches), measured against average
Spin rate (SR) [30]	Spin rate in revolutions per minute after a pitch is released
Perceived velocity (PV) [31]	How fast a pitch is perceived by a hitter based on pitch velocity and release point
Defense	
Defensive runs saved (DRS) [32]	How many runs a defender saved based on errors, range, outfield arm, and double-play ability
Range factor (RF) [33]	Sum of fielder's putouts and assists divided by number of games played
Outs above average (OAA) [34]	Ranged-based metric based on how many outs a player as saved, used for both infielders and outfielders
Distance covered (DCOV) [35]	Total distance covered by a defender from the time of contact to the moment he fields it

confounding variables which are outside the scope of a simply reported batting average—opponent defender skill, ballpark dimensions, weather, pitch velocity, and spin rate [36].

Statcast has both enhanced the fan experience, displaying real-time data on telecasts and various media outlets, as well as provided baseball with a new source of data to challenge classic sabermetric models and revolutionize baseball strategy. The wealth of data has forced many clubs to hire teams of statisticians to better understand the data and to recommend how to act upon it.

Health and Injury Tracking System and Return to Play

Thomas et al. (2020) recently published a systematic review which described the RTP rates and performance of pitchers after both primary and revision ulnar collateral ligament (UCL) reconstruction (UCLR) [37]. This review included 29 studies published from 1980 to 2019 [37]. After primary UCLR, MLB pitchers had RTP rates from 80 to 97% at 12 months [37]. Return to the same level of play (RTSP) rates, however, were only 67 to 87% at 15 months [37]. Following revision UCLR, RTP ranged from 77 to 85% and RTPS ranged from 55 to 78% [37].

While this is a robust study about a specific injury, there are few similar studies in the literature. As such, to better understand player injury, the MLB, with its minor league affiliates, players' unions, and healthcare experts, established the MLB Health and Injury Tracking System (HITS) in 2010 [8, 38]. Previously, the understanding of the epidemiology of baseball injury was limited by disabled list (DL) reports. The DL only reported injury to a specific body region (rather than a specific injury) and was used as a roster management tool in addition to an actual accounting of injuries [38]. Since the establishment of HITS, numerous studies have been published which report the epidemiology of specific baseball injuries, their return to play characteristics, and related performance outcomes [8, 38–46].

Notably, Camp et al. in 2018 published a comprehensive report of the 50 most common injuries in the MLB and Minor League Baseball (MiLB) from 2011 to 2016, with specialized fact sheets for each injury outlining specific characteristics and return-to-play times [38]. They reported nearly 50,000 injuries over this six-year period, with about 8000–8500 injuries per year. Of these, roughly 5000 were season-ending injuries, and among the non-season-ending injuries, there was a mean 16 days missed per injury [38]. The upper extremity accounted for 39% of all injuries and the lower extremity 35%. Interestingly, pitchers were the most injured position, 3.6 times more frequently than catchers, 5.1 times more than outfielders, and 5.8 times more than infielders [38]. This incidence of upper extremity injury and pitcher injury correlates with other epidemiologic studies of baseball injury [47, 48]. These types of studies have greatly aided trainers in better understanding the prognosis and expected time course for return to play. Studies using the HITS database have also provided numerous insights into risk factors for injury, diagnosis, and best treatment practices.

Motion Analysis Techniques

The pitching motion has been well described as involving six specific phases of muscle movement, and improper mechanics in each phase has been associated with an increased risk for injury [49]. These phases are as follows: (1) wind-up, (2) stride or early cocking, (3) late cocking, (4) acceleration, (5) deceleration, and (6) follow-through [49]. There are a number of modalities with which one can analyze a pitcher's mechanics including two-dimensional video, three-dimensional motion analysis in a laboratory, and emerging wearable technologies (see discussion below) [49]. Historically, three-dimensional motion analysis is considered the gold standard and provides detailed kinetic and kinematic data about each phase of the previously mentioned pitching motion [49].

Emerging Technologies

Wearable IMUs

Until recently, the kinetic and kinematic analysis of the pitching motion relied heavily on high-speed cameras to capture motion in three dimensions [50–53]. However, there are limitations associated with this technology including cost and the need for access to a controlled laboratory which has these capabilities [50, 53]. In particular, pitchers will not pitch with full velocity in the laboratory, raising questions as to whether this data really represents in-game kinetics and kinematics. This limits the widespread analysis for all pitchers across all levels of baseball [52].

However, there have been recent advancements in wearable technology that has made gathering of this kinetic and kinematic data easier and less cumbersome, more accessible by limiting the need for the laboratory, and even less expensive [50–54]. These technologies utilize inertial measurement units (IMUs) to allow for real-time motion analysis in three-dimensions. IMUs are lightweight and small systems with embedded three-dimensional accelerometers, gyroscopes, and magnometers which allow for the biomechanical assessment of motor functions by dynamically tracking anatomical segments [54]. IMUs have already demonstrated utility in gait analysis [53–55] and are currently being used to analyze pitching mechanics [50, 52, 53, 56–61].

The motusBASEBALL (Motus Global) IMU system was first approved for use in the MLB 2016 [51, 62], since there has been increasing interest in the sports medicine literature using this technology to study pitching biomechanics [52, 53, 56–61]. Table 2 lists all studies published to our knowledge using the motusBASEBALL technology and their conclusions.

Lizzio et al. (2020) published a protocol to optimize data collection and how to troubleshoot device malfunction [50].

While promising, the current evidence conflicts as to the accuracy of the system [63]. Four studies have assessed the validity of wearable technology to date [52, 53, 63, 64]. Camp et al. (2017) first validated wearable IMU technology against the "gold standard" of motion capture techniques to capture pitching kinetic and kinematic variables [53]. Wearing a compression sleeve with the IMU overlying the medial elbow, they recorded arm slot, arm speed, arm rotation, and elbow varus torque in 35 pitchers throwing fastballs [53]. These are well-studied variables and known risk factors for the development of pitcher injury [50–53]. The IMU data were subsequently transmitted via Bluetooth technology to a smartphone containing proprietary biomechanical algorithms [53]. They reported correlation coefficients (*r* values) between the IMU analysis and motion capture techniques of 0.93 (varus torque), 0.94 (arm rotation), 0.95 (arm slot), and 0.85 (arm speed) [53].

Makhni et al. (2018) used similar IMU technology (sensor placed over medial elbow) to measure elbow torque, arm speed, arm slot, and shoulder rotation across multiple pitch types (fastballs, curveballs, and change-ups) [52]. By measuring outlier rate, they reported precision rates for the IMU system of 96.9% for fastballs, 96.9% for curveballs, and 97.9% for change-ups [52]. Boddy et al. (2019) compared wearable IMU technology to motion capture technology [64]. They found statistically significant correlations (*r* values) between the two when measuring arm slot (0.975), shoulder rotation (0.749), and stress (0.667) while elbow extension velocity failed to reach statistical significance [64].

However, Camp et al. (2021) again attempted to validate the motusBASEBALL IMU technology with a dedicated study [63]. Comparing measurements obtained simultaneously by motion capture and the IMU system, they found this technology to be reliable for arm speed and not reliable for arm slot, arm stress, or shoulder rotation [63]. Overall, these results suggest that while this technology can be used to measure number of pitches and may be useful to compare one pitcher within themselves, it may not be accurate to compare between pitchers.

Even with these limitations, the technology may be useful for injury prevention and post-injury rehabilitation [49, 50]. Wearable technology allows for the prospective collection of biomechanical data which would allow for real-time pitch counts, which is more convenient and accurate than prior methods [50].

Machine Learning and AI to Predict Injury

Machine learning (ML) and artificial intelligence (AI) have already been transforming many facets of society, from self-driving vehicles, targeted advertisements, healthcare and orthopedic surgery, and sports [4, 8, 65–69]. While AI is a broad term

Table 2 Studies using motusBASEBALL sensor

Study	Conclusion
Camp et al. (2017) [53]	Shoulder flexibility, arm speed, and elbow varus torque (a proxy for injury risk) are interrelated.
Okoroha et al. (2018) [61]	Fatigue and injury risk are likely related. Pitch velocity decreased and medial elbow torque increased with innings pitched.
Okoroha et al. (2018) [59]	In youth and adolescent pitchers, fastballs generated the highest elbow torque and curveballs the highest arm speed. Increasing age and size of the pitchers' arm was protective against elbow torque. Increased ball velocity, BMI, and decreased arm slot predicted elbow torque.
Makhni et al. (2018) [52]	Medial elbow torque was highest in fastballs.
Okoroha et al. (2019) [60]	In youth pitchers, increased ball weight was associated with greater medial elbow torque, decreased pitch velocity, and decreased arm speed.
Melugin et al. (2019) [58]	Decreasing perceived pitching effort correlated with decreased elbow varus torque and velocity, but not proportionally.
Leafblad et al. (2019) [56]	Ball velocity and elbow torque do not necessarily correlate in long-toss; thus, some pitchers may benefit from long-toss programs for rehabilitation.

which refers to the analysis of large data sets using algorithms to gain useful inference, ML is a subset of AI which uses this data specifically to predict an outcome [67]. In ML, the machine studies and "learns" from "training sets" of real-world data using pattern recognition to determine relationships. Then, these machines are provided "testing sets" to create predictions and compare them with known outcomes. This process continuously repeats, helping to improve the accuracy of the models based on feedback from these "testing sets." In theory, ML mirrors the way in which humans learn with constant improvement in analyses as new data becomes available [67].

Recently, Makhni et al. published a review article (2021) regarding ML and AI in orthopedics, its current impact on the field, and future applications [67]. These technologies will continue to help improve outcomes, reduce costs and inefficiencies, and improve the overall value of the care we provide. Notably, there has been a tenfold increase in ML publications in over the past twenty years [69], which highlights the impact it has had on this field. Currently, ML and AI are being used to predict readmission risk in total hip and knee arthroplasty [70], to interpret preoperative radiographs in the setting of revision arthroplasty to determine the prior implant class and manufacturer [68], and predict post-operative outcomes after total joint arthroplasty [71], among many other applications.

In baseball, in conjunction with classic sabermetrics and performance data, modern ML algorithms are being used to try to predict injury risk and specific anatomic injury location. After compiling a database of 13,982 player-years, Karnuta et al. applied both logistical regression (LR) and ML techniques to develop an algorithm to predict MLB injuries before they occurred [8]. They based these models on age, performance data (sabermetrics for hitting, pitching, and overall), prior injury history, and DL data. While the AUC ranged from 0.71 to 0.80 for predicting position player injury which was fairly reliable, the AUC for pitchers ranged from 0.61 to 0.69 and was considered poorly reliable [8]. However, in almost all (13 of 14) cases, ML was superior to LR when predicting player injury and demonstrated improved accuracy with subsequent iterations of the injury-prediction model [8].

While these models have yet to demonstrate true clinical reliability, they certainly highlight the bright future of baseball analytics and the limitations of historical logistical regression. In this study, as the ML model continued to "learn," it continued to improve its reliability which is fundamental to ML. As ML models continue to improve, we may reach a point in baseball where we can reliably predict injury and intervene before injury actually happens.

Conclusion

Data and analytics research has transformed baseball into becoming one the largest data-driven sports worldwide. From its origins in sabermetrics, ML and AI are seemingly commonplace in the current analysis of baseball from a performance standpoint to injury prevention and rehabilitation. The future seems bright with wearable technologies, though more research is needed in this area to better validate this technology before it can be universally relied upon.

Declarations

Conflict of Interest The authors declare no competing interests.

Human and Animal Rights and Informed Consent All reported studies/experiments with human or animal subjects performed by the authors have been previously published and complied with all applicable ethical standards (including the Helsinki declaration and its amendments, institutional/national research committee standards, and international/national/institutional guidelines).

References

1. Baseball History, American history and you [Internet]. [cited 2021 Oct 27]. Available from: https://baseballhall.org/baseball-history-american-history-and-you
2. Healey G. The new moneyball: how ballpark sensors are changing baseball. Proceedings of the IEEE. Institute of Electrical and Electronics Engineers Inc.; 2017. p. 1999–2002.
3. Matt Kelly. What is sabermetrics? Modern analytics impact nearly every part of today's game [Internet]. MLB.com. 2019 [cited 2021 Oct 27]. Available from: https://www.mlb.com/news/sabermetrics-in-baseball-a-casual-fans-guide
4. Hall B. Artificial intelligence, machine learning, and the bright future of baseball [Internet]. The National Pastime: The Future According to Baseball. 2021 [cited 2021 Oct 27]. Available from: https://sabr.org/journal/article/artificial-intelligence-machine-learning-and-the-bright-future-of-baseball/
5. Fred Ivor-Campbell. F.C. Lane. 2000 [cited 2021 Oct 29]; Available from: https://sabr.org/bioproj/person/f-c-lane/
6. A guide to sabermetric research [Internet]. [cited 2021 Oct 31]. Available from: https://sabr.org/sabermetrics
7. James LeDoux. Introducing pybaseball: an open source package for baseball data analysis. 2017.
8. Karnuta JM, Luu BC, Haeberle HS, Saluan PM, Frangiamore SJ, Stearns KL, et al. Machine learning outperforms regression analysis to predict next-season major league baseball player injuries: epidemiology and validation of 13,982 player-years from performance and injury profile trends, 2000-2017. Orthopaedic Journal of Sports Medicine. SAGE Publications Ltd; 2020;8.
9. Neil Weinberg. Complete list (pitching) [Internet]. FanGraphs. 2014 [cited 2021 Nov 5]. Available from: https://library.fangraphs.com/pitching/complete-list-pitching/
10. Neil Weinberg. Complete list (offense) [Internet]. FanGraphs. 2014 [cited 2021 Nov 5]. Available from: https://library.fangraphs.com/offense/offensive-statistics-list/
11. Neil Weinberg. Overview [Internet]. FanGraphs. 2015 [cited 2021 Nov 5]. Available from: https://library.fangraphs.com/defense/overview/
12. ben Harris. A sabermetric primer understanding advanced baseball metrics [Internet]. The Athletic. 2018 [cited 2021 Nov 5]. Available from: https://theathletic.com/255898/2018/02/28/a-sabermetric-primer-understanding-advanced-baseball-metrics/

13. Neil Weinberg. Getting started [Internet]. FanGraphs. 2015 [cited 2021 Nov 5]. Available from: https://library.fangraphs.com/getting-started/
14. Catch probability [Internet]. [cited 2021 Nov 7]. Available from: https://mlb.com/glossary/statcast/catch-probability
15. Neil Weinberg. How to evaluate a pitcher, sabermetrically [Internet]. 2014 [cited 2021 Nov 7]. Available from: https://www.beyondtheboxscore.com/2014/6/2/5758898/sabermetrics-stats-pitching-stats-learn-sabermetrics
16. Expected ERA (xERA) [Internet]. [cited 2021 Nov 7]. Available from: https://www.mlb.com/glossary/statcast/expected-era
17. GLOSSARY [Internet]. [cited 2021 Nov 7]. Available from: https://www.mlb.com/glossary
18. John Foley. Twinkie town analytics fundamentals: the flaws of batting average [Internet]. 2020 [cited 2021 Nov 7]. Available from: https://www.twinkietown.com/2020/5/11/21253031/twinkietown-analytics-fundamentals-come-learn-baseball-with-john-sabermetrics-batting-average-flaws
19. Shaun Payne. Walks are important: why batting average is less than enough in today's game [Internet]. 2011 [cited 2021 Nov 7]. Available from: https://www.twinkietown.com/2020/5/11/21253031/twinkietown-analytics-fundamentals-come-learn-baseball-with-john-sabermetrics-batting-average-flaws
20. Neil Weinberg. How to evaluate a hitter, sabermetrically [Internet]. 2014 [cited 2021 Nov 7]. Available from: https://www.beyondtheboxscore.com/2014/5/26/5743956/sabermetrics-stats-offense-learn-sabermetrics
21. Expected WEIGHTED ON-BASE AVERAGE (xwOBA) [Internet]. [cited 2021 Nov 7]. Available from: https://www.mlb.com/glossary/statcast/expected-woba
22. Batting average on balls in play (BABIP) [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/advanced-stats/babip
23. Weighted on-base average (wOBA) [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/advanced-stats/weighted-on-base-average
24. Weighted runs created plus (wRC+) [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/advanced-stats/weighted-runs-created-plus
25. Wins above replacement (WAR) [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/advanced-stats/wins-above-replacement
26. Exit velocity (EV) [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/statcast/exit-velocity
27. Fielding independent pitching (FIP) [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/advanced-stats/fielding-independent-pitching
28. Adjusted earned run average (ERA+) [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/advanced-stats/earned-run-average-plus
29. Pitch movement [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/statcast/pitch-movement
30. Spin rate (SR) [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/statcast/spin-rate
31. Perceived velocity (PV) [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/statcast/perceived-velocity
32. Defensive runs saved (DRS) [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/advanced-stats/defensive-runs-saved
33. Range factor (RF) [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/advanced-stats/range-factor
34. Outs above average (OAA) [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/statcast/outs-above-average
35. Distance covered (DCOV) [Internet]. [cited 2022 Jan 22]. Available from: https://www.mlb.com/glossary/statcast/distance-covered
36. Healey G. Combining radar and optical sensor data to measure player value in baseball. Sensors (Switzerland). MDPI AG. 2021;21:1–14.
37. Thomas SJ, Paul RW, Rosen AB, Wilkins SJ, Scheidt J, Kelly JD, et al. Return-to-play and competitive outcomes after ulnar collateral ligament reconstruction among baseball players: a systematic review. Orthopaedic Journal of Sports Medicine. SAGE Publications Ltd; 2020;8:232596712096631.
38. Camp CL, Dines JS, van der List JP, Conte S, Conway J, Altchek DW, et al. Summative report on time out of play for major and minor league baseball: an analysis of 49,955 injuries from 2011 through 2016. The American journal of sports medicine. SAGE Publications Inc.; 2018;46:1727–32.
39. Makhni MC, Curriero FC, Yeung CM, Leung E, Kvit A, Mroz T, et al. Epidemiology of spine-related neurologic injuries in professional baseball players. 2021
40. Rubenstein WJ, Allahabadi S, Curriero F, Feeley BT, Lansdown DA. Fracture epidemiology in professional baseball from 2011 to 2017. Orthopaedic Journal of Sports Medicine. SAGE Publications Ltd; 2020;8.
41. Hultman K, Szukics PF, Grzenda A, Curriero FC, Cohen SB. Gastrocnemius injuries in professional baseball players: an epidemiological study. American Journal of Sports Medicine. SAGE Publications Inc.; 2020;48:2489–98.
42. Lucasti CJ, Dworkin M, Warrender WJ, Winters B, Cohen S, Ciccotti M, et al. Ankle and lower leg injuries in professional baseball players. American Journal of Sports Medicine. SAGE Publications Inc.; 2020;48:908–15.
43. Erickson BJ, Chalmers PN, D'Angelo J, Ma K, Romeo AA. Performance and return to sport after latissimus dorsi and teres major tears among professional baseball pitchers. American Journal of Sports Medicine. SAGE Publications Inc.; 2019;47:1090–5.
44. Camp CL, Desai V, Conte S, Ahmad CS, Ciccotti M, Dines JS, Altchek DW, D'Angelo J, Griffith TB Revision ulnar collateral ligament reconstruction in professional baseball: current trends, surgical techniques, and outcomes. Orthopaedic Journal of Sports Medicine. SAGE Publications Ltd; 2019;7
45. Hodgins JL, Trofa DP, Donohue S, Littlefield M, Schuk M, Ahmad CS. Forearm flexor injuries among major league baseball players: epidemiology, performance, and associated injuries. American Journal of Sports Medicine. SAGE Publications Inc.; 2018;46:2154–60.
46. Esquivel A, Freehill MT, Curriero FC, Rand KL, Conte S, Tedeschi T, Lemos SE Analysis of non-game injuries in major league baseball. orthopaedic journal of sports medicine. SAGE Publications Ltd; 2019;7
47. Li X, Zhou H, Williams P, Steele JJ, Nguyen J, Jäger M, et al. The epidemiology of single season musculoskeletal injuries in professional baseball. Orthopedic Reviews. Open Medical Publishing; 2013;5:3.
48. Posner M, Cameron KL, Wolf JM, Belmont PJ, Owens BD. Epidemiology of major league baseball injuries. American Journal of Sports Medicine. 2011;39:1676–80.
49. Christoffer DJ, Melugin HP, Cherny CE. A clinician's guide to analysis of the pitching motion. Current reviews in musculoskeletal medicine. Humana Press Inc.; 2019. p. 98–104.
50. Lizzio VA, Cross AG, Guo EW, Makhni EC. Using wearable technology to evaluate the kinetics and kinematics of the overhead throwing motion in baseball players. Arthroscopy Techniques. Elsevier B.V.; 2020;9:e1429–31.
51. Fleisig GS. Editorial commentary: changing times in sports biomechanics: baseball pitching injuries and emerging wearable technology. Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North

America and the International Arthroscopy Association. W.B. Saunders; 2018;34:823–824.

52. Makhni EC, Lizzio VA, Meta F, Stephens JP, Okoroha KR, Moutzouros V. Assessment of elbow torque and other parameters during the pitching motion: comparison of fastball, curveball, and change-up. Arthroscopy : the Journal of Arthroscopic & Related Surgery : Official Publication of the Arthroscopy Association of North America and the International Arthroscopy Association. W.B. Saunders; 2018;34:816–822.
53. Camp CL, Tubbs TG, Fleisig GS, Dines JS, Dines DM, Altchek DW, et al. The relationship of throwing arm mechanics and elbow varus torque: within-subject variation for professional baseball pitchers across 82,000 throws. The American Journal of Sports Medicine. SAGE Publications Inc.; 2017;45:3030–5.
54. Roggio F, Bianco A, Palma A, Ravalli S, Maugeri G, di Rosa M, et al. Technological advancements in the analysis of human motion and posture management through digital devices. World Journal of Orthopedics. Baishideng Publishing Group Co. 2021;12:467–84.
55. Prasanth H, Caban M, Keller U, Courtine G, Ijspeert A, Vallery H, et al. Wearable sensor-based real-time gait detection: a systematic review. MDPI AG: Sensors; 2021.
56. Leafblad ND, Larson DR, Fleisig GS, Conte S, Fealy SA, Dines JS, et al. Variability in baseball throwing metrics during a structured long-toss program: does one size fit all or should programs be individualized? Sports Health. SAGE Publications Inc.; 2019;11: 535–42.
57. Lizzio VA, Smith DG, Jildeh TR, Gulledge CM, Swantek AJ, Stephens JP, et al. Importance of radar gun inclusion during return-to-throwing rehabilitation following ulnar collateral ligament reconstruction in baseball pitchers: a simulation study. Journal of Shoulder and Elbow Surgery. Mosby Inc.; 2020;29:587–92.
58. Melugin HP, Larson DR, Fleisig GS, Conte S, Fealy SA, Dines JS, et al. Baseball pitchers' perceived effort does not match actual measured effort during a structured long-toss throwing program. American Journal of Sports Medicine. SAGE Publications Inc.; 2019;47:1949–54.
59. Okoroha KR, Lizzio VA, Meta F, Ahmad CS, Moutzouros V, Makhni EC. Predictors of elbow torque among youth and adolescent baseball pitchers. American Journal of Sports Medicine. SAGE Publications Inc.; 2018;46:2148–53.
60. Okoroha KR, Meldau JE, Jildeh TR, Stephens JP, Moutzouros V, Makhni EC. Impact of ball weight on medial elbow torque in youth baseball pitchers. Journal of Shoulder and Elbow Surgery. Mosby Inc.; 2019;28:1484–9.
61. Okoroha KR, Meldau JE, Lizzio VA, Meta F, Stephens JP, Moutzouros V, et al. Effect of fatigue on medial elbow torque in baseball pitchers: a simulated game analysis. American Journal of Sports Medicine. SAGE Publications Inc.; 2018;46:2509–13.
62. Maury Brown. How wearable technology got quietly into major league baseball. 2016.
63. Camp CL, Loushin S, Nezlek S, Fiegen AP, Christoffer D, Kaufman K. Are wearable sensors valid and reliable for studying the baseball pitching motion? An independent comparison with marker-based motion capture. The American Journal of Sports Medicine. SAGE Publications Inc.; 2021;49:3094–101.
64. Boddy KJ, Marsh JA, Caravan A, Lindley KE, Scheffey JO, O'Connell ME. Exploring wearable sensors as an alternative to marker-based motion capture in the pitching delivery. PeerJ. PeerJ Inc.; 2019;2019.
65. Kakavas G, Malliaropoulos N, Pruna R, Maffulli N. Artificial intelligence: a tool for sports trauma prediction. Injury. Elsevier Ltd. 2020;51:S63–5.
66. Luu BC, Wright AL, Haeberle HS, Karnuta JM, Schickendantz MS, Makhni EC, Nwachukwu BU, Williams III RJ, Ramkumar PN Machine learning outperforms logistic regression analysis to predict next-season NHL player injury: an analysis of 2322 players from 2007 to 2017. Orthopaedic Journal of Sports Medicine. SAGE Publications Ltd; 2020;8.
67. Makhni EC, Makhni S, Ramkumar PN. Artificial intelligence for the orthopaedic surgeon: an overview of potential benefits, limitations, and clinical applications. The Journal of the American Academy of Orthopaedic Surgeons. NLM (Medline); 2021. p. 235–43.
68. Helm JM, Swiergosz AM, Haeberle HS, Karnuta JM, Schaffer JL, Krebs VE, Spitzer AI, Ramkumar PN machine learning and artificial intelligence: definitions, applications, and future directions. Current Reviews in Musculoskeletal Medicine. Springer; 2020. p. 69–76
69. Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review.
Frontiers in Bioengineering and Biotechnology. Frontiers Media S.A.; 2018
70. Goltz DE, Ryan SP, Hopkins TJ, Howell CB, Attarian DE, Bolognesi MP, et al. A novel risk calculator predicts 90-day readmission following total joint arthroplasty. Journal of Bone and Joint Surgery - American Volume. Lippincott Williams and Wilkins. 2019;101:547–56.
71. Jayakumar P, Bozic KJ. Advanced decision-making using patient-reported outcome measures in total joint replacement. Journal of Orthopaedic Research. John Wiley and Sons Inc.; 2020;38:1414–22.