**The State Of Data Analytics In Sports**

**Prepared for**

Dr. Nicole Infanta Keller

ENGW 3302

Northeastern University

**Prepared by**

Payton McAlice

Northeastern University

**May 29, 2022**

**INTRODUCTION**

Like many different professional industries, the advancement of computer technology has allowed for the sports industry to embrace different data science practices in its day-to-day operations. Developments in both motion-tracking and information-storage technologies have given front offices and coaching staffs from all different sports the ability to work with large datasets pertaining to their respective industries. As a rapidly developing practice within an internationally visible industry, data analytics have been spotlighted within sports as a point of conversation. Fans, athletes, and other personnel are all asking how data science within the sports industry will continue to grow and develop. Also being called into question is how data analytics do and should affect decision-making both within play and during player-development.

This report discusses the recent growth of data analytics within the sports industry. The differences of development within different sports will be considered and given rationale. Additionally, this report will look at points of conflict within the industry regarding how analytics are used and how this is balanced with what are considered more traditional practices and methodologies. All of this information will serve to evaluate the future of sports analytics in the coming years.

**FINDINGS**

**Background and Context**

Since the early days of sports, players, coaches, and front office personnel have all looked for ways to use data and information to improve their on-field performance. These practices date back all the way to the 1940’s when coaches, such as American football’s Paul Brown, used game footage to scout opponents and observe trends in the types of plays they tended to run (Abee). As information has become more widely available, data analytics have become the center of attention. The sport of baseball has found itself at the forefront of analytics techniques, and these practices have been largely publicized because of the well-documented history of the Oakland Athletics use of data analysis in the early 2000’s.

Today, more decisions both on and off the field are now being made using data and statistical analysis. Data science practices continue to integrate within front offices in every sport. Coaches are relying on data to develop new strategies and approaches to the game, while front offices are relying on data to scout, evaluate, and develop potential players. In some instances, this has created tension with those who prefer more traditional methods of strategy and development. How data will be used in the future of sports is highly contentious, but it is generally agreed that it will continue to play a bigger role in all sports.

**The Growth of Analytics Teams Within Sports**

*Analytics Teams Within Professional Baseball*

The consensus amongst sports professionals is that baseball is at the forefront of using data analytics within their industry. The development of baseball analytics has largely been accredited to the Oakland Athletics organization of the early 2000’s, led by general manager Billy Beane. Beane’s strategy for roster development focused on finding available undervalued players who could be acquired at low costs. For example, instead of just looking at a batter’s ability to get base hits, Beane and his staff look at a batter’s holistic ability to reach base via base hit, walk, hit-by-pitch, and other means. His front-office shifted its focus from player scouting and more traditional methods of analyzing player value to statistic-driven evaluation. Oakland was generally successful in this era- making the playoffs every year between 2000 and 2003- despite being often considered an underdog. Their story has been immensely publicized because of the 2003 book *Moneyball: The Art of Winning an Unfair Game* and the 2011 film *Moneyball* and is often viewed as a turning point in how data is used in sports (Boylan).

Over the past few seasons in Major League Baseball (MLB), multiple teams have accredited surprise success to the growth of their internal analytics departments. One such team is the Tampa Bay Rays, who reached the World Series in 2020 despite having the third-lowest payroll amongst all teams in the league. Their shift in philosophy is accredited to Andrew Friedman, a former Wall Street analyst who led the team until 2014. Under Friedman, the franchise placed a focus on organizational depth and challenged the conventions of baseball strategy. For their efforts, Tampa Bay has been recognized as a leader in modern baseball analytics. In 2019, their general manager Erik Neander was recognized by MLB as Executive of the Year. Many within the Tampa Bay front office have gone on to take leadership positions with other teams, as Chaim Bloom did in 2019 with the Boston Red Sox (who then reached the playoffs in 2021). They continue to push analytics today, being the first MLB team to hire a process and analytics coach who sits in the dugout with players and other coaches during games (Lee).

Many teams are now shifting their personnel to value statistical analysis and data science. Mike Elias and Sig Mejdal, two front-office executives in the Baltimore Orioles organization, mentioned an increase in interest in statistical analysis. In a Forbes interview, Mejdal notes that proficiency in database engineering and languages such as R are skills of growing popularity and demand (Karpovich).

*Analytics Teams Within Other Sports*

Data analytics in sports other than baseball are further behind in terms of development. This is largely a result of having significantly less available in-game data in other sports. However, this does not mean that there is a lack of interest in other sports and in developing analytics in said sports.

In October 2021, the Chicago Blackhawks of the National Hockey League (NHL) hired Kyle Davidson as their general manager. Coming from an analytics background, Davidson in turn hired Jeff Greenberg as his assistant general manager. Greenberg is a former executive of the Chicago Cubs baseball organization and helped oversee the development of their internal analytics department. Together, they have envisioned a heavy investment into breaking ground within the realm of hockey analytics in a vein similar to baseball; this includes the development of internal data warehouses and research of what statistical and data analysis can be applied in hockey player-development and front office operations (Pope).

Other sports such as American football have also seen a growing interest in data analytics. While these sports have had a history of analytics since the early 2000’s, they are seeing a sharp increase in public interest due the spotlight that baseball put on data-based decision-making. In 2016, the Cleveland Browns football organization hired Paul Depodesta, someone considered to be one of Billy Beane’s biggest front office assets during their time in Oakland together. As of January 2022, there are 86 reported data analytics jobs within the National Football League (NFL). This number is reportedly expected to grow by up to 50% over the next three years (Abee).

**The Effects of Analytics on In-Game Decision-Making**

*In-Game Decision-Making in Fluid Sports*

The biggest problem faced by my analysts of most sports is how to use data to inform decision-making. This problem is most applicable to more “fluid” sports, which is to say sports that do not operate in a series of plays. An example of a non-fluid sport is baseball. Baseball can be easily broken down into individual matchups between pitcher and batter, and fielders are positioned at the start of each play. This means there is a surplus of information that can be tracked between how different players react in different isolated situations. Conversely, sports like American football, soccer, hockey, and basketball are more fluid by nature; their more continuous and dynamic play makes for more complexities regarding how data can be applied to every small decision that goes into the outcome of one play (Tuyls).

In spite of the difficulty involved in decision-making within fluid sports, there is some extent to which this is done in different sports. For example, in certain situations within American football a coach can rely on analytics to determine the best play-call. Certain metrics such as Expected Points (EP) and Expected Points Added (EPA) are advanced calculated metrics based on the game situation (Abee). In some instances, they can be used to evaluate the value of the play run in context to the rest of the game. This informs decision-making at certain influential points within a game, such as when a team needs to decide between a field goal attempt, a punt, or a fourth-down running or passing attempt.

*In-Game Decision-Making in Baseball*

As previously mentioned, the baseball industry has not had the issues of game-state representation that other sports are facing. This means that in many ways, data analytics have integrated themselves into the in-game decision making of coaches and players to a far greater extent than in other sports. In turn, this has led to a number of new innovations in strategy. One of these innovations is the advent of the fielding shift, which is when fielding players change their placement on the field in accordance to where on the field a given hitter tends to hit the ball. Shifts have been hyper-tuned to individual players using advanced data collection and analysis.

A large proponent of experimental baseball strategy motivated by data analytics is Tampa Bay Rays manager Kevin Cash. Cash has relied on analytics to inform him on how he should manage his pitchers. A new strategy often used by Cash involves changing pitchers often, regardless of their performance up to the point in which they are pulled. The motivation is that statistically, pitchers often struggle immensely later in games due to fatigue and the fact that the hitters being faced have now seen the pitcher multiple times during the game to that point (Lee).

*Controversies Regarding Analytical Decision-Making in Baseball*

A big conversation around modern baseball has been how these “new-school” strategies have gone against conventional “old-school” baseball strategies. With each new practice introduced, criticisms are introduced about how the sport is affected either economically or in terms of entertainment. The fielding shift in particular has come under intense scrutiny within baseball discourse. This has culminated in the MLB experimenting with rules that regulate how much fielders can reposition themselves in minor-league affiliate clubs and leagues. Furthermore, during the renegotiation of the Collective Bargaining Agreement between the league and its players in the 2021-22 offseason, the implementation of a fielding shift ban was heavily discussed. Some feel as though it has created an unfair standard for hitters to live up to in order to produce offensively at the same rate they did before the shift became mainstream. Others feel that it has made for a negative fan experience that has less exciting action on balls put in play (Snyder).

While Cash has been often praised for his influence on modern baseball (he was recognized as American League Manager of the Year in both 2020 and 2021), many of his strategies have come under criticism. In one instance during the 2020 World Series, Cash decided to remove starting pitcher and 2018 Cy Young Award winner Blake Snell from the game in the sixth inning in spite of a low pitch count and dominant performance to that point. Cash’s rationale was his rationale from the entire season to that point; Snell was statistically more likely to struggle as he began to face the same hitters for the third time that night. His pitchers for the rest of the game struggled, and Tampa Bay ultimately lost the game and series (Heyen).

*Decision-Making In Other Sports*

Other sports are also seeing an increase in advanced metrics. Many sports are currently developing or have developed a metric similar to the NFL’s EP and EPA statistics, which aim to indicate how a play-call can affect a team’s chance to score. Game theory techniques have become a recent point of interest within sports like soccer and are being combined with artificial intelligence models such as neural networks to draw up optimal strategies (Tuyls). While this task has proven more difficult in more fluid sports, the interest and work going towards this is undeniable.

**The Effects of Analytics On Player Development**

*Advanced Metrics And Player Evaluation*

Alongside its effects on strategy, data analytics have had an immense impact on player development in various sports. How players are developed and how their mechanics are adjusted is often a result of the data available about them.

In baseball, shifts in player development can be seen through the advent of sabermetrics. Sabermetrics are advanced metrics that try to quantify a player’s performance using objective measures that work independently of ballpark and location, game situation, etc. (Wulff). Additionally, with the advent of tracking tools such as Trackman and Statcast, there is now information about every pitch and how it spins and moves and information about every hit and how hard and how high it was hit (Meoli).

The effects of having this data are already observable in how players are developed and valued today. In modern baseball, an emphasis on “three true-outcome” hitters is apparent. These are hitters who tend to frequently either strikeout, walk, or hit a homerun. Evidence of a three true-outcome hitter is one who has high exit-velocities and launch angles on their balls hit in play. How defense is measured has shifted as well, with a newly increased interest in a fielder’s range and ability to cover large areas of the field. Catchers can now also be evaluated in terms of how effectively they can manage pitch-calling and framing pitches as they are thrown.

*Kinetics, Biological Analysis, and Video Analysis in Player Development*

Conversely to advanced metrics, kinetics and video analysis are readily available in almost all professional sports. This has allowed for a lot of advancement in player development using previous footage and studying the motion of the body.

One area in which this has become extremely important is injury prevention. With the ability to track movements of the body, professional trainers can observe trends that lead to stress and straining and can help shift player mechanics to be safe and replicable. This has been highlighted as a point of interest in several sports including basketball and soccer. According to National Basketball Association (NBA) commissioner Adam Silver, teams can monitor players during practice and games to observe levels of fatigue. Even samples of player saliva can be used to monitor the condition of a player’s body. Biometrics are used to influence player diets and workout regimens. Per artificial intelligence professor and DeepMind employee Karl Tuyls, kinetics and video analysis are already being used in the front offices of soccer teams to monitor player mechanics and injury prevention (Tuyls).

**The Trajectory of Analytics in Sports**

*Analytics In Professional Baseball*

The question of how analytics will evolve in baseball is not a matter of what new metrics will be found. There has already been a noticeable effect on the sport at a professional level. Strategies and development techniques have shifted towards favoring what is considered statistically optimal. New strategies introduced pertaining to how pitchers are used, fielders are positioned, and how hitters approach at-bats are all against the grain of what was considered conventional strategy for much of the 20th century.

The biggest question regarding baseball analytics is how it will integrate with old-school ideologies and coaching. Scouts now have access to data they have never had access to before that can complement much of the work they’ve previously done via intuition and human observation. The marriage of conventional knowledge and modern analytics will greatly determine the future of baseball as a whole and could very well set a precedent for analytics in other sports.

*Analytics In Other Sports*

While it is difficult to predict how sports analytics will shape over the coming years in other sports, it is possible to anticipate where attention will be directed. As kinesiological data and video are already so readily available, there will be an emphasis on developing models that can use this data to further advance player development and evaluation. Where many of the questions lie is what new metrics will be made available in the coming years to help bring about a statistical-revolution akin to what has already been observed in baseball. When the issue of how to evaluate fluid game-states is resolved, this will prove groundbreaking within the sports analytics realm and likely introduce a number of new strategies and gameplay that will be subject to public scrutiny and pushback against those who prefer old-school methodologies.

**CONCLUSION**

In summary, sports analytics is a fast-growing field that continues to draw more interest in all major sports. Baseball is currently at the forefront of sports analytics due to the ease with which its game-states can be represented as independent, rigid matchups. This has led to advancements in player metrics and evaluation as well as new statistically backed strategies. The future of the sport and baseball analytics is largely dependent on how these changes in baseball culture will mesh with conventional baseball school of thought. Other sports that are further behind are still struggling with the issues of how to quantify on-field performance in a more fluid game environment. However, there have been advancements in game theory, artificial intelligence, and kinesiology. It can be expected that these sports will continue to develop in the direction that baseball has developed over the previous decades. The uncertainty of their futures rests in what data will become available in the coming years.

The focus on big data, data management, analytics, and artificial intelligence within the sports industry are reflective of many other different industries across the board. As the world continues to accumulate more data, there will always be an increased interest in making that data accessible and manipulable to be used in the decision-making processes. Sports can be seen as a microcosm for other industries and serves as an example of what a data revolution can do to shift industry standards and practices.

**REFERENCES**

**MLA9**

Abee, Jerrod. “The Slow Growth of Data Analytics in Football.” Samford University, Samford University, 6 Jan. 2022, <https://www.samford.edu/sports-analytics/fans/2022/The-Slow-Growth-of-Data-Analytics-in-the-Football?image>.

Boylan, Mike. “‘Moneyball’ and the Oakland A's: How Has It Been so Misunderstood?” Bleacher Report, Bleacher Report, 3 Oct. 2017, <https://bleacherreport.com/articles/679950-revisiting-moneyball-and-the-oakland-as-how-has-it-been-so-misunderstood>.

Heyen, Billy. “Kevin Cash's Decision to Pull Blake Snell, Explained: How Analytics Overruled World Series Context Clues and Cost the Rays.” Sporting News, Sporting News, 29 Oct. 2020, <https://www.sportingnews.com/us/mlb/news/blake-snell-kevin-cash-analytics-explained-world-series/15ja52nunza2b1b37untxyltk3>.

Karpovich, Todd. “Baltimore Orioles Look to Be at MLB Forefront with Analytics.” Forbes, Forbes Magazine, 17 Dec. 2019, <https://www.forbes.com/sites/toddkarpovich/2019/12/17/orioles-look-to-be-at-forefront-with-analytics/?sh=2c0ba8881bc2>.

Lee, Joon, and David Schoenfield. “World Series 2020: How the Rays Became the Rays. inside the Brain Trust of MLB's Most Innovative Team.” ESPN, ESPN, 21 Oct. 2020, <https://www.espn.com/mlb/story/_/id/30146884/world-series-2020-how-rays-became-rays-brain-trust-mlb-most-innovative-team>.

Meoli, Jon. “Omar Comin': Orioles' Uniquely Named Analytics Database Is a Rare Beneficiary of Baseball's Shutdown.” Baltimore Sun, Baltimore Sun, 14 Apr. 2020, <https://www.baltimoresun.com/sports/orioles/bs-sp-orioles-database-omar-the-wire-20200414-mywqyq46hjbpfpogknpxrf2mo4-story.html>.

Pope, Ben. “Blackhawks Seeking to 'Close the Gap' between Baseball, Hockey Data Systems.” Chicago Sun-Times, Chicago Sun-Times, 19 May 2022, <https://chicago.suntimes.com/blackhawks/2022/5/21/23103605/blackhawks-cubs-hockey-baseball-systems-analytics-jeff-greenberg-kyle-davidson>.

Rajan, Rahul. “Why the Proliferation of AI Technology Is Important to the Future of Sports.” Sports Business Journal, Sports Business Journal, 19 May 2022, <https://www.sportsbusinessjournal.com/SB-Blogs/COVID19-OpEds/2022/05/19-Rajan.aspx?hl=analytics&amp;sc=0>.

Snyder, Matt. “Why It's Time for Major League Baseball to Ban the Shift.” CBS Sports, CBS Sports, 3 Sept. 2020, <https://www.cbssports.com/mlb/news/why-its-time-for-major-league-baseball-to-ban-the-shift/>.

Tuyls, Karl, et al. “Game Plan: What Ai Can Do for Football, and What Football Can Do for Ai.” Journal of Artificial Intelligence Research, vol. 71, 2021, pp. 41–88., <https://doi.org/10.1613/jair.1.12505>.

Wulff, Shaun S., and Priyantha G. De Silva. “A Multi-Criteria Approach for Evaluating Major League Baseball Batting Performance.” Journal of Sports Analytics, 2022, pp. 1–14., <https://doi.org/10.3233/jsa-200298>.