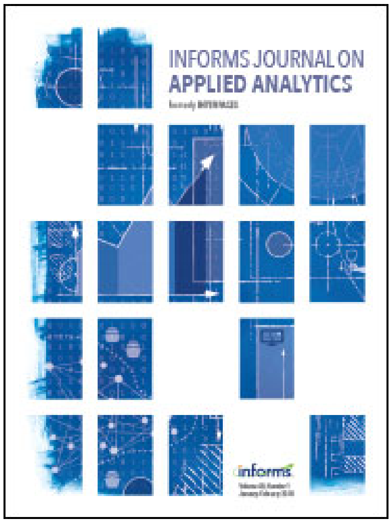

Interfaces

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Introduction to the Special Issue on Analytics in Sports, Part I: General Sports Applications

Michael J. Fry, Jeffrey W. Ohlmann,

Interfaces
Vol. 42, No. 2, March–April 2012, pp. 105–108
ISSN 0092-2102 (print) | ISSN 1526-551X (online)
http://dx.doi.org/10.1287/inte.1120.0633

Introduction to the Special Issue on Analytics in Sports, Part I: General Sports Applications

Michael J. Fry
Department of Operations, Business Analytics and Information Systems, University of Cincinnati, Cincinnati, Ohio 45221,
mike.fry@uc.edu

Jeffrey W. Ohlmann
Department of Management Sciences, University of Iowa, Iowa City, Iowa 52242,
jeffrey-ohlmann@uiowa.edu

Sports is a rapidly growing application area for analytics. The use of analytics is pervasive in the professional sports community as evidenced by the increased role for those practicing sports analytics in front-office management and coaching. Part I of this Special Issue on Analytics in Sports is devoted to the application of a variety of methodologies to a broad range of sports. The papers analyze golf, hockey, baseball, motorcycle racing, and college football. The methodologies employed draw from many areas of analytics including optimization, probabilistic modeling, and choice models.

Key words: sports; analytics.

Sports are big business. Most estimates place the total market value of spectator sports in the United States in the range of hundreds of billions of dollars. As one paper in this special issue notes, the market size of spectator sports in the United States has been estimated to be double that of the automotive industry and is easily one of the top-10 business markets globally. Given the size and impact of the sports industry, it should not be surprising that this is a fertile application area for operations research (OR) models.

Sports analytics has also been gaining widespread notoriety in the general populace. Michael Lewis' entertaining story about the use of data analysis in baseball in *Moneyball*: *The Art of Winning an Unfair Game* (Lewis 2004) is arguably the most visible account of sports analytics. Many of the strategies documented in *Moneyball*, which were then employed by the small-market Oakland Athletics team to help it compete with teams with much larger payrolls, have been adopted in some form by many other Major League Baseball (MLB) teams and teams in other sports. Although *Moneyball* is clearly not the earliest example of applying analytics to baseball—Bill James (James 2001), George Lindsey (Lindsey 1959, Lindsey 1961, Lindsey 1963), and many others preceded the work described in *Moneyball*—this book was the catalyst for introducing the broader sports community to the potential benefits of quantitative analysis. *Moneyball* became a best seller; in 2011, a movie of the same name (starring Brad Pitt) achieved considerable box office success.

Recently, a new breed of management has become more prominent in sports organizations. There has been an increase in the number of front office personnel with quantitative training and (or) the appreciation for the power of analytics to help improve the performance of their teams both on the field of play and from a business perspective. Paul DePodesta, one of the central figures in *Moneyball*, earned an economics degree from Harvard; he has worked in the front offices of several MLB teams and currently is the vice president of player development and scouting for the New York Mets. Daryl Morey, general manager of the Houston Rockets in the National Basketball Association (NBA), has degrees in computer science and statistics from Northwestern University and an MBA from MIT; yet, he has little actual basketball playing experience. Morey has been an active crusader for sports analytics through his involvement with the annual MIT-Sloan Sports Analytics Conference, which brings academics and practitioners together to discuss the latest advances in sports analytics. Another proponent of quantitative methods in the NBA is Mark Cuban, owner of the 2011 NBA Champion Dallas Mavericks. Cuban has worked directly with Wayne Winston, John and Esther Reese Professor of Decision Sciences at Indiana University's Kelley School of Business, to apply sports analytics concepts to on-court and off-court decisions made by the Dallas Mavericks. As these and other examples show, the sports world is teeming with opportunities for analytics to contribute to sports management.

This special issue of *Interfaces* seeks to add to the academic literature examining sports applications. Historically, *Interfaces* has been a leading outlet for sports analytics research, but this is the first special issue of *Interfaces* dedicated to this topic. Sticking with the overall spirit of the *Interfaces* journal, we present practice-oriented sports analytics papers. As evidenced by the overwhelming response to the call for papers for this special issue (we received over 40 submissions), many researchers in OR-related fields are working on sports-related applications.

Papers in Part I of This Special Issue

Because of the large number of high-quality submissions to this special issue, we publish it as a double issue. The first issue focuses on general (nonscheduling) sports applications; the second issue focuses solely on sports scheduling, which has emerged as a highly successful application area for OR. As seen by the variety of sports examined in the first issue, the application of analytics is not limited to any particular sport. Three of the papers examine golf applications. Sports analytics applied to golf has experienced increased interest recently because of the wealth of data made available through the PGA TOUR's ShotLink database (Deason 2006). The other papers in this special issue discuss hockey, baseball, motorcycle racing, and college football.

Although several dedicated sports analytics journals have recently been established, *Interfaces* has traditionally been a leader in publishing papers on this topic. This, and other findings, are highlighted in "Identifying the 'Players' in Sports Analytics Research" by B. Jay Coleman. This paper provides an overview of the sports analytics research landscape: which universities are contributing research on this topic, where the papers are appearing, and which papers are most commonly cited. This paper is similar in spirit to the Rothkopf rankings (Fricker 2011) published by *Interfaces*, which identifies the universities and organizations that are the leading contributors to OR practice. Coleman shows that sports analytics is a fast-growing research area, outpacing other application areas in OR. The author finds that a relatively small group of researchers works on sport analytics, and that many of these researchers are based outside the United States.

Most team-based professional sports leagues assign newly eligible players to teams through some form of sports draft where each team selects players for its team from the pool of all draft-eligible players. In MLB, this is a particularly difficult task because 1,500 players are selected each year over 50 draft rounds. All these players must be evaluated by the baseball teams, but player evaluations are inherently imprecise. Multiple scouts evaluating the same player often have varying opinions on the player's projected performance in the major leagues. In "A Major League Baseball Team Uses Operations Research to Improve Draft Preparation," Streib et al. present an application of OR to improve an MLB team's process of assessing baseball players. The authors develop an optimization tool that helps a team come to a consensus on rankings for draft-eligible players. The algorithm is based on voter-preference research, including the classic Kemeny-Young method; however, the size of the problem faced by MLB teams requires new methods and solution approaches. The authors' tool has been used by one MLB team for several years, and other MLB teams have expressed interest in developing a similar tool. This paper presents an excellent example of using analytics to better utilize the invaluable subject matter expertise of personnel (scouts in this case).

The goal in team sports is to create an optimal lineup of players at different positions to defeat the opposing team. However, this problem is difficult because different player positions require different skills and a collection of the best individual players does not necessarily produce the best lineup because of interaction effects. In "Quantifying the Contribution of NHL Player Types to Team Performance," Chan et al. examine the question of how to determine the optimal player lineups in ice hockey by using a clustering technique to define distinct player types for offense, defense, and goalies. These clusters are used in regression models to determine the relationship between team performance and the clustered player types. The authors show how the results from their models can be used to analyze trades in the National Hockey League, and they provide an Excel-based tool that team managers can use to evaluate proposed trades and new player signings.

Determining the winning golfer in stroke-play competition is easy: the golfer who requires the fewest strokes to complete all holes over the course of the tournament is the winner. However, as Mark Broadie explains in "Assessing Golfer Performance on the PGA TOUR," identifying the facets of a golfer's game that contributed to winning play is more difficult. Did the golfer win because of his ability to hit the ball farther than other golfers? Or did his superior putting lead to victory? In this paper, Broadie uses the PGA TOUR ShotLink database to perform an analysis of eight years of strokes hit by PGA TOUR golfers. This allows him to evaluate golfers in different phases of their game: long, short, and putting. The author uses a dynamic-programming formulation to measure the "strokes gained" of golfers in the different phases of their game. As Broadie explains, "strokes gained represents the decrease in the average number of strokes to finish the hole from the beginning of the shot to the end of the shot." The author's analysis allows him to rank all PGA TOUR golfers in the various phases of their game. Broadie finds that Tiger Woods (examined between 2002 and 2010) was the top-ranked golfer in terms of strokes gained and that Woods' long game was the largest contributor to his dominance. The author generalizes this finding and also concludes that the long game is the most important differentiating factor of golfers on the PGA TOUR. This conclusion is contrary to many previous academic papers and contradicts the commonly held belief characterized by the oft-repeated quote from golfer Bobby Locke that golfers "drive for show, but putt for dough" (Riach 2007).

In "A Proposal for Redesign of the FedEx Cup Playoff Series on the PGA TOUR," Hall and Potts propose a new design for this season-ending event. The FedEx Cup playoff series was established in 2007 to bring the professional golf season to an exciting end, similar to the Super Bowl in the National Football League or the World Series in MLB. The FedEx Cup system is a season-long points competition that ends in a playoff series of four events where the top 125 point-scoring players are reduced each week until 30 players compete in a final 72-hole stroke-play tournament. However, recent FedEx Cup playoffs have been anticlimactic because the competition was essentially decided before the final round was played. Under the authors' redesign, the FedEx Cup playoff would be a match-play event where the highest-ranked players (based on season-long performance) would be seeded in such a way that the top seeds would have byes during the first rounds of the playoff. By using a match-play format, the authors assure an exciting end to the event and also guarantee that the top-seeded players will participate in the final rounds, which are televised. The authors show that their redesign meets the requirements of the PGA TOUR Commissioner who has called for a redesigned format.

The Ryder Cup is a biennial golf competition where a team of golfers from the United States competes against a team of golfers from Europe in a match-play format. The captain of each team must determine the sequence in which the golfers compete during the 12 head-to-head matches without knowing the sequence selected by the opposing team's captain. Several possible strategies are available to the team captains, including putting the team's best golfers in the beginning matches or saving the best golfers for the final matches. The latter strategy has the benefit of using the best golfers in matches that are likely to determine the competition's winner; however, its risk is that these golfers will not participate if the opposing team has won a sufficient number of previous matches to clinch the competition. How to best sequence the golfers has been examined previously in *Interfaces* in Hurley (2002). Hurley uses a game-theoretic argument to state that randomly determining the sequence of golfers is the best strategy. In "Optimizing the Sunday Singles Lineup for a Ryder Cup Captain," McClure et al. use an optimization model to determine the best lineup of golfers, assuming that the opposing captain will assign golfers to matches based on known probability distributions. Using data from the World Golf Ranking, the authors fit a logistic regression model to predict the outcome of each potential pairing of opposing golfers. The optimization model then maximizes the expected value of points earned by the decision maker's team, which is used as a proxy for maximizing the probability of winning the Ryder Cup. The results of the optimization model generally agree with the findings of Hurley (2002), albeit for very different reasons. They conclude that it is difficult for the team captain to choose a sequence that will provide a significant advantage; however, this is unlikely to end the debate on this subject.

In "TEAM ASPAR Uses Binary Optimization to Obtain Optimal Gearbox Ratios in Motorcycle Racing," Amorós et al. present a novel application of a binary optimization model to determine the optimal combination of gears to include in a motorcycle's gearbox for a Grand Prix competition. Motorcycle racing teams have a relatively short time to prepare for each race. An important decision in this preparation is the choice of gearbox configuration. Several hundred thousand gear combinations are possible, and the optimal choice is based on track configuration and rider preferences. Prior to this project, racing teams made these decisions manually using simple calculations and rider-technician experience. In this paper, the authors formulate a relatively simple binary program that can be solved quickly to determine the optimal gear selections. TEAM ASPAR has been using the authors' model and has observed significant improvements in its racing times.

Large firms in industries such as consumer packaged goods, electronics, and entertainment typically spend vast amounts of money and time characterizing customers to better understand what entices these customers to consume products. Yet, although the sports industry is one of the world's largest revenue generators, little work has been done to understand fans' relationships to their favorite sports, teams, and athletes. "Exploring the Demand Aspects of Sports Consumption and Fan Avidity" seeks to fill this void. In this paper, DeSarbo and Madrigal propose a new multidimensional scaling procedure to describe sports fans' choices of forming avidity for a particular team, league, or athlete. The authors collect a large amount of empirical data from university students regarding their interest in college football to construct the model; they then segment fans based on the manner in which they express their avidity (e.g., purchasing apparel, investing time or effort, attending games). The authors discuss several ways that an organization can use such market segmentation to increase revenues. This paper, which provides one of the first analysis of sports-related fan avidity, should spark continued interest in expanding this field of study.

References

Amorós J, Escudero LF, Monge JF, Segura JV, Reinoso O (2012) TEAM ASPAR uses binary optimization to obtain optimal gearbox ratios in motorcycle racing. *Interfaces* 42(2):191–198.

Broadie M (2012) Assessing golfer performance on the PGA TOUR. *Interfaces* 42(2):146–165.

Chan TYC, Cho JA, Novati DC (2012) Quantifying the contribution of NHL player types to team performance. *Interfaces* 42(2):131–145.

Coleman B (2012) Identifying the "players" in sports analytics research. *Interfaces* 42(2):109–118.

Deason L (2006) ShotLink a statistical superstar. Accessed January 3, 2012, http://www.pgatour.com/story/9596346/.

DeSarbo W, Madrigal R (2012) Exploring the demand aspects of sports consumption and fan avidity. *Interfaces* 42(2):199–212.

Fricker RD (2011) Editorial: The ninth Rothkopf rankings of universities' contributions to the INFORMS practice literature. *Interfaces* 41(6):548–563.

Hall NG, Potts CN (2012) A proposal for redesign of the FedEx Cup playoff series on the PGA TOUR. *Interfaces* 42(2):166–179.

Hurley WJ (2002) How should team captains order golfers on the final day of the Ryder Cup matches? *Interfaces* 32(2):74–77.

James B (2001) *The New Bill James Historical Baseball Abstract* (Free Press, New York).

Lewis M (2004) *Moneyball*: *The Art of Winning an Unfair Game* (W. W. Norton and Company, New York).

Lindsey GR (1959) Statistical data useful for the operation of a baseball team. *Oper. Res.* 7(2):197–207.

Lindsey GR (1961) The progress of the score during a baseball game. *J. Amer. Statist. Assoc.* 56(295):703–728.

Lindsey GR (1963) An investigation of strategies in baseball. *Oper. Res.* 11(4):477–501.

McClure B, Cassady R, Rainwater C, Chimka JR (2012) Optimizing the Sunday singles lineup for a Ryder Cup captain. *Interfaces* 42(2):180–190.

Riach S (2007) *Par for the Course*: *Golf Tips and Quips*, *Stats and Stories* (Hallmark Books, Kansas City, MO).

Streib N, Young S, Sokol J (2012) A Major League Baseball team uses operations research to improve draft preparation. *Interfaces* 42(2):119–130.