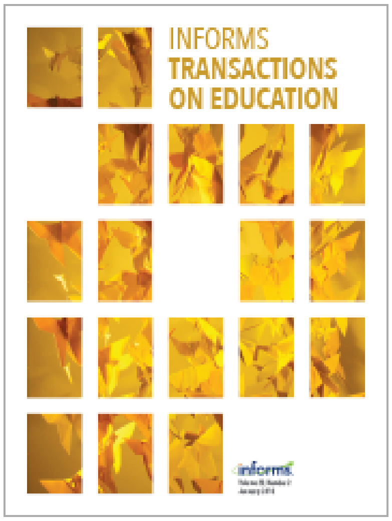


INFORMS Transactions on Education

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

Case—Moneyball for Murderball: Using Analytics to Construct Lineups in Wheelchair Rugby

Timothy C. Y. Chan, Craig Fernandes, Albert Loa, Nathan Sandholtz

INFORMS TRANSACTIONS ON EDUCATION
Vol. 24, No. 2, January 2024, pp. 182–183
ISSN 1532-0545 (online)
https://pubsonline.informs.org/journal/ited

Case

Moneyball for Murderball: Using Analytics to Construct Lineups in Wheelchair Rugby

Timothy C. Y. Chan,[a,*] Craig Fernandes,[a] Albert Loa,[a] Nathan Sandholtz[b]

[a] Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario M5S 3G8, Canada; [b] Department of Statistics, Brigham Young University, Provo, Utah 84602

*Corresponding author

Contact: tcychan@mie.utoronto.ca, https://orcid.org/0000-0001-6508-6755 (TCYC); craig.fernandes@mail.utoronto.ca (CF); albert.loa@mail.utoronto.ca (AL); nsandholtz@stat.byu.edu (NS)

Received: April 27, 2022
Revised: July 4, 2022
Accepted: August 31, 2022
Published Online in Articles in Advance: February 13, 2023

https://doi.org/10.1287/ited.2023.0282cs

Keywords: data analysis • regression • optimization • sports analytics • wheelchair rugby

1. Introduction

Ming was the new assistant coach of Canada's national wheelchair rugby team. The Paralympic Games were coming up in one year, and he needed to get up to speed quickly to help his team prepare. The head coach had noticed the impact that sports analytics was having in other sports and wondered if wheelchair rugby could be analyzed in a similar way. He knew that Ming had an analytics background, so he asked Ming to determine whether analytics could give their team a competitive advantage.

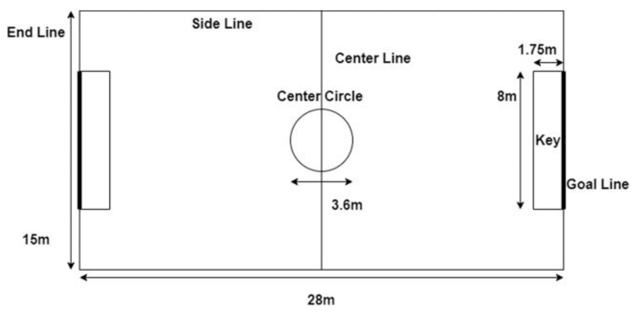
Unlike in professional sports where players can be acquired through drafting, trades, or free agency, on a national amateur team, Ming had to work with the players already on the roster. He thought that if he could understand which players contributed more to team performance, then he could improve the lineups that his team put on the court. A major focus of sports analytics was finding "hidden gems," that is, individuals that performed better than they appeared to coaches and management teams. If he could accurately *estimate* or *predict* the on-court value of the players on his team, then he could use those estimates to *optimize* the lineup decisions made by the coaches.

2. Wheelchair Rugby

Wheelchair rugby, also known as "murderball" because of the game's aggressive nature, is a mixed-gender sport that is played on a standard sized basketball court (28 m × 15 m) with an 8-m goal line and an 8 m × 1.75 m–key area at each end of the court (Figure 1). The objective of the game is to score more goals than your opponent. A goal is scored by crossing the opposing team's goal line while in possession of the volleyball-sized game ball. Teams consist of 12 players, with 4 players from a team allowed on the court at the same time. Substitutions are unlimited but can only be made during a stoppage in play. Stoppages in play occur when a goal is scored, when a foul is committed (e.g., holding another player's wheelchair), or when a time violation occurs (e.g., taking more than 10 seconds to dribble the ball). Games consist of four eight-minute quarters with three-minute overtime periods as necessary.

A key feature of wheelchair rugby that distinguishes it from able-bodied team sports is the existence of a quantitative player classification system that is based on physical ability. Players are given a physical rating between 0.0 and 3.5 in increments of 0.5. Men are rated between 0.5 and 3.5, whereas women are rated between 0.0 and 3.0. Players with a smaller range of movement are given a lower physical rating. For example, an athlete rated 0.5 would have extensive shoulder, triceps, and wrist weakness, whereas an athlete rated 3.5 might have only some impairment in their trunk function. These physical ratings are used to determine permissible lineups. In particular, the sum of the player physical ratings on the court for one team cannot exceed eight at any time. That means

Figure 1. Wheelchair Rugby Court



that lineups such as 2.0-2.0-2.0-2.0 or 3.5-3.0-1.0-0.5 are allowed, but 3.0-3.0-2.0-1.0 is not.

3. Analysis

Ming started by examining historical game data to see what insights he could derive around player and team performance. He had access to 660 games worth of data involving 12 top international teams. The data were stored in two files: one that contained a list of all players and their physical rating and one that contained detailed information in each game, organized by "stint." A stint is a period of time in which there are no substitutions and no end-of-quarter breaks. In other words, the home and away lineups are held constant during a stint. For each stint, the file contained information on which game it came from, the home and away teams, the duration, the number of goals each team scored, and the players on the court.

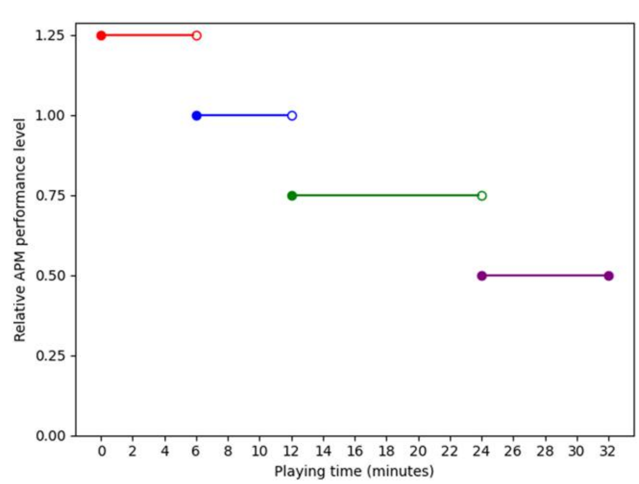
Ming first wanted to assign a value to each player based on plus-minus, a primitive performance metric that is measured on stints. To do so, he calculated the plus-minus per minute for each stint (i.e., the number of home goals minus the number of away goals in that stint divided by the number of minutes of that stint), and then he assigned that value to all the players on the court for that stint. To get a player's overall plus-minus, Ming would need to somehow combine the plus-minus values from the stints that player played. Although in principle this could be accomplished using a simple average, Ming was aware that one of the main principles of sports analytics is that raw statistics can be misleading. Therefore, after exploring the data thoroughly, he decided to develop a statistical model to better estimate each player's on-court value. In particular, he decided to construct an "adjusted plus-minus" model (APM). The main idea of this approach is to use regression techniques to "adjust" a player's raw plus-minus value based on knowing who they played with during their stints. Other variables could also be in cluded in the model, such as which team was the home team.

Upon fitting the model, armed with estimates of the value of each player on the team, Ming was ready to use the estimates to help the coach optimize lineup decisions. The coach had observed that the quality of the starting lineup had a significant impact on the early part of the game. A strong starting lineup leads to more scoring opportunities and generally built confidence that the team could use going forward. The coach wanted to optimize the way the starting lineup was chosen and tasked Ming to come up with a method to do so. A follow-up question to Ming was whether he could optimize lineups throughout the game, not just the starting lineup, if the coach provided recommended playing times for each player. These lineups would need to account for the maximum physical rating allowed on the court at any time.

In addition to the two questions posed by the coach, Ming started thinking about optimizing playing time and lineups in the context of player fatigue. To incorporate fatigue into the analysis, Ming spoke to the strength and conditioning coach who had data on how on-court performance declined as a function of playing time. Figure 2 summarizes this relationship. The values represent performance relative to a player's average APM, which is assumed to be positive. That is, up to the sixth minute, a player is estimated to perform at roughly 1.25 times their average APM. Between their 12th and 24th minute of play, they play at 0.75 times their average APM, and so on.

With this knowledge of fatigue and each player's APM estimate, Ming wanted to determine the optimal number of minutes each player should play to maximize team APM over the course of a game. Finally, he wondered whether he could extend this idea to determine optimal lineups for every minute of the game that maximized total APM and accounted for fatigue in a single optimization model. He believed that this model would be more challenging to build, but ultimately would provide better solutions.

Figure 2. Relationship Between Player Performance Level and Playing Time



Note. Values represent a performance multiplier over average performance.