Moneyball: Brad Pitt, the statistician and the movie

A major Hollywood release has an unlikely hero. In *Moneyball*, Brad Pitt, Hollywood heart-throb and twice voted sexiest man alive, plays not a glamorous spy, nor a hard-bitten detective but … a statistician. **Ray Stefani** and **Jim Albert** say the movie hits a home run …

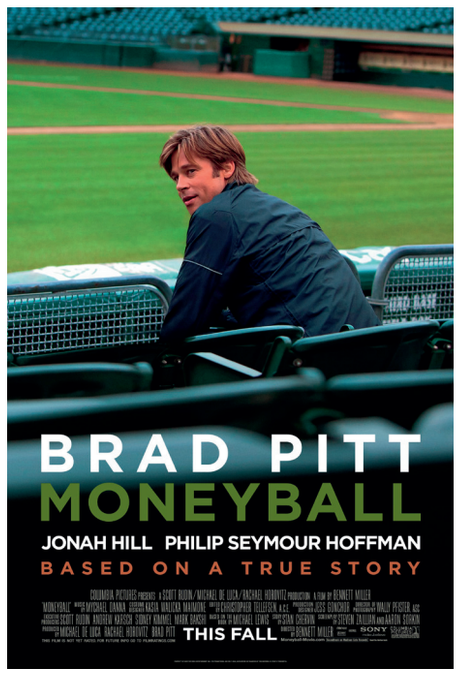
The release of the movie *Moneyball* has been an exciting event for statisticians. It is new for a statistician to be the hero of a major Hollywood film, and even more encouraging that the statistician is played by Brad Pitt. Perhaps a whole new image of our profession is developing. What next? Shall we see Angelina Jolie and Anne Hathaway competing for roles as statistician super-heroines, saving the world through Bayesian analyses? We can but hope.

Moneyball is based on the book of the same title by Michael Lewis and tells the true story of Billy Beane, who as general manager of the failing Oakland Athletics baseball team transformed the team and its fortunes by selecting players using sports statistics (or "sabermetrics" as they are known – SABR being the Society for American Baseball Research).

It is ironic that the basic storyline of the *Moneyball* book is about building a baseball team with impersonal statistics, whereas the *Moneyball* movie is intensely personal from start to finish. *Moneyball* immerses the viewer in the world of baseball, and Brad Pitt brings the complex and compelling personality of Billy Beane to life. Indeed the movie *is* Brad Pitt. So do the statistics come through?

When history is rewritten for entertainment there can be some casualties. Paul DePodesta, the person who actually convinced Beane that statistics could revolutionise the way baseball teams are constructed, chose not to be associated with the movie. Hence the movie created the fictional statistical protagonist Peter Brand, played wonderfully by Jonah Hill. Statisticians are seekers after, and sticklers for, truth; but they must perhaps forgive such departures from it in the interests of entertainment, or art – or of conveying statistics through entertainment and art.

As Peter Brand explains in the movie, one should value a player by how many wins he creates for his team, and baseball games are won when teams score more runs than their opponents. The traditional batting average (hits per at bat) is at best an imperfect measure of runs production. If one looks at team hitting data, say for the 2008–2010 seasons, then the batting average has a modest 0.74 correlation with runs scored. In the movie, when Oakland loses the prestigious hitter Jeremy Giambi, Billy Beane says that Oakland cannot replace Giambi but they can replace Giambi's on-base percentage (OBP). Indeed, the OBP (the number of times on base per plate appearance) is a superior offensive measure -- it has a 0.88 correlation with runs scored. Bill James, pioneer of sabermetrics and the man who coined the word, in one of his *Baseball Abstracts* explained that a



Baseball and cricket

Baseball and cricket share much common history. Baseball is mentioned in Jane Austen's *Northanger Abbey*, written in 1798. The first international cricket match was played between the USA and Canada in 1844 (Canada won by 23 runs).

Henry Chadwick (1824–1908), often called the founder of baseball, was an English-born American brought up on cricket.

Not everyone in the world understands the rules, or the statistics, of baseball. Cricket fans may be equally surprised that some of the arcana of that game are not universally known. Fans of each, however, share a common interest: the two games are probably the most statistically analysed of all sports. In the interests of multiculturalism we therefore provide the following information on the fundamentals of baseball statistics for those who know only cricket.

Baseball for cricketers

- A *hit*, also called a *base hit*, is credited to a batter when he safely reaches first base after hitting the ball. A hit involves reaching only first base; a *run* is a circuit of all four bases on the diamond.
- The *plate* is the base at which the batter stands to receive the ball.
- A player's *plate appearance* is the number of times he comes to bat.
- The number of times *at bat* is the same as plate appearance – but with deductions for: receiving four pitches that the umpire calls "balls" (the equivalent of "no balls" in cricket); being hit by a pitch (if the ball impacts his body rather than the bat); making a "sacrifice hit", intended to help a team-mate complete a run while sacrificing his own chance; or being awarded a first base due to interference or obstruction.
- A batter is *on base* when safely on any one of the four corners of the diamond-shaped infield.
- A batter's *on-base percentage* is the number of times he gets on base, expressed as a percentage of his total number of plate appearances.
- A player's *slugging percentage* is calculated as the total number of bases reached by the player divided by the number of times at bat.

batting measure should include the ability to get on base and the ability to advance runners to home. A simple good "sabermetric" measure of batting performance is the "on base percentage plus slugging" or OPS; for team data, the OPS has a 0.96 correlation with runs scored. In cricket, a similar debate rages as to batting average (runs per dismissal) being inferior to strike rate (runs per 100 balls faced).

Beane, as general manager of the Oakland Athletics, was faced with severe financial restrictions. He said the team must "adapt or die", a remarkably liberal philosophy for a sport that does things "by the book". He embraced Peter Brand's (read Paul DePodesta's) innovative approach and through the power of Beane's personality insisted that the scouts and team on-field manager comply, with much resistance from the traditional baseball people. Beane had been a failed player himself, which shaped his desire to work at a higher level in management. Yet Beane had such a intense dislike of failure that he couldn't bring himself to attend a game; he spent many games in the exercise room of the team locker area.

Ironically, Oakland reached the 2002 playoffs with a major league record 20-game winning streak – which was relying on the improbable to prove the probable. The team did not win the World Series, which discouraged Beane who felt that his quest had failed. The "failure" was in his own mind only. The most telling lines in the movie come from the owner of the Boston Red Sox who tried unsuccessfully to hire Beane. He tells Beane that Oakland spent $250,000 per win while the New York Yankees spent $1.4 million per win: Beane's method had worked. Boston (with enough money invested) went on to win a World Series – and they did it by playing moneyball.

Whether the viewer likes sports or not, or statistics or not, this movie will provide an intensely compelling story that hits a home run (or hits for six, as would be said in cricket). For those who want more, the *Moneyball* book provides a fuller insight into the sabermetrics movement in baseball. Chapter 6 describes in detail Beane's dealings in what he calls the "science of winning an unfair game".

Ray Stefani is Professor at California State University, Long Beach, and a specialist in sports statistics. Jim Albert is Professor of Maths and Statistics at Bowling Green State University, Ohio

Moneyball is on general release in the USA; it opens in the UK on 25th November 2011.

Moneyball – The Art of Winning an Unfair Game, by Michael Lewis is published by W.W. Norton (2003), ISBN 0-393-05765-8.

Writing with *Significance*!

A joint writing competition hosted by the Young Statisticians Section of the Royal Statistical Society and *Significance* magazine

Illustration: Tom Boulton

Calling all budding writers within the first 10 years of their statistical career!

- Do you have an idea for an interesting article which would be suitable for *Significance*?
- Have you always wanted to try your hand at writing in a more journalistic style? Are you good at interpreting data?

Then why not write an article and submit it to our competition? This is the first writing competition hosted jointly between *Significance* and the Young Statisticians Section of the Royal Statistical Society. *Significance* is published by the Royal Statistical Society and the American Statistical Association, and is for anyone interested in statistics and the analysis and interpretation of data. It is primarily a general interest magazine for statisticians, users of statistics and all those interested in statistics. It is not a research journal, and articles are not peer-reviewed (see `www.significancemagazine.org/`). Articles should be accessible to a wide and non-specialist audience and should be about an area or application of statistics that is of broad relevance or has an important and topical application, in a way that lives up to both meanings of the tagline "Statistics making sense". Articles should be between 1800 and 3000 words long and can include tables, figures and photographs. Their style should be clear and easy to read – avoid the formal layout of an academic report – and technical terms and mathematics should be used sparingly if at all, and suitably explained. End references are optional, but should be limited to three or four at most.

Anyone is welcome to enter, regardless of membership or affiliation. The only stipulation is that you should be "young" (in career terms, not necessarily in age) – that is, you must be a student or within the first 10 years of your career. The article could be on work that you have done, or it could explain the work of others. Only submissions in English will be considered. Manuscripts must be original and not under consideration for publication elsewhere, though we welcome magazine articles based on work in theses or in papers that have been submitted to or accepted by academic journals, provided the two are sufficiently different. The closing date for entry is *1st March 2012*. All articles will be assessed by a review committee made up of representatives from both the Young Statisticians Section and *Significance*. The winning article will be published in *Significance*. Runner-up articles will be published on the *Significance* website, or in *Significance* at the editor's discretion.

Please email your submissions, in Word or as a PDF, to `significance@rss.org.uk`. We hope you enjoy writing your piece.

Laura Gray, RSS Young Statisticians Section
Julian Champkin, Significance