

Working in Sports Analytics

Tim B. Swartz *

Abstract

With the increasing fascination of sport in society and the increasing availability of sport related data, there are great opportunities to carry out sports analytics research. In this paper, we discuss some of the issues that are relevant to publishing in the field of sports analytics. Potential publication outlets are identified, some summary statistics are given, and some experiences and opinions are provided.

Keywords: journals, publishing, sports analytics, statistics in sport.

*T. Swartz is Professor, Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6. Swartz has been partially supported by the Natural Sciences and Engineering Research Council of Canada.

1 Introduction

One of the questions that I often hear is “Where should I attempt to publish my sports statistics research?” At the outset, I wish to stress that this is not a rigorous academic paper. Rather, it is a highly opinionated commentary based largely on my personal experiences where I attempt to provide some insight with respect to the above question, and more generally, working in sports analytics. At a minimum, I hope that the compiled list of potential outlets may be useful to new researchers in sports analytics.

A difficulty with the entire exercise is that the definition of sports analytics is not so clear. I will not even attempt to define sports analytics as there is bound to be disagreement in any definition. For example, the topic of player evaluation is surely a sports analytics problem. However, it could be viewed as a financial problem if the primary interest is the management of cap space. As another example, playing a game optimally with respect to tactics and formation is surely a sports analytics problem. However, to play optimally, players ought to consider their diet. Is this a sport analytics problem or a nutrition problem?

Another problematic definition is “sport” itself. For example, I am sure that disagreements exist on the sport/non-sport classification of hiking, horse racing and darts. For this investigation, I take a broad view of sport but insist that a legitimate sport needs to have a competitive component.

In the following sections, I give rough categorical rankings for publication outlets beginning with the most favourable. Within each category, I provide some biased discussion and anecdotes. However, before proceeding to the assessments, I discuss some general issues related to sports research and publishing. I conclude with a short discussion.

1.1 Criteria of a good paper

I think that the best sports statistics contributions have two components: (1) they address a real sporting problem and (2) they contain statistical novelty. Strangely, these two criteria can sometimes conflict. When you solve a real sporting problem, you would like the work to attract mainstream attention. And for your work to be embraced by the world, it cannot be excessively technical. If it is technical, then you better have nice visualizations that can be appreciated by coaches, team management and fans (e.g. Goldsberry and Weiss 2013).

The only black-box statistical procedure that I know of that has gained widespread sporting acceptance is the Duckworth-Lewis-Stern (DLS) method used in resetting targets in limited overs cricket (Duckworth and Lewis 1998, 2004, Stern 2016). And in my opinion, the success of DLS was opportunistic - the previous method for resetting targets (i.e. run rates) was widely known to be grossly unfair. Such opportunities in sport are far and few between.

On the other hand, criteria (2) involving statistical novelty is the bread and butter of statistical journals. In many cases, the more technical your paper, the better. The sporting application sometimes seems to be an afterthought to some statistical reviewers. Therefore, for a long time, the sports statistics researcher was trying to serve two masters - the sporting community who shun technical work and the statistics community who love technical work and don't care much about sport. Fortunately, with the acceptance of money-ball thinking (Lewis 2003) the landscape has improved.

I think a researcher needs to carefully appraise the novelty of one's work in order to choose an appropriate outlet. And "novelty" is a grey area measured on a continuum. For example, applying a standard statistical technique in a new setting may be useful, but is not likely to be considered novel by a mainstream statistics journal. However, in some other applied journals where data analyses are the norm, such work may be considered perfectly acceptable.

1.2 Dissemination

When I speak to my colleagues in actuarial science, they tell me that there are only a handful of journals where they consider submitting their manuscripts. The situation is starkly different with sports analytics research.

One of the difficulties in getting a handle on sports analytics research is that it is widely dispersed. More than once, I have submitted manuscripts for publication only to find that I have missed key papers in the literature. Given that sporting problems can have many flavours (e.g. economic, strategic, health related) sports statistics papers are spread across various disciplines, and often exist in journals as one-off papers. And this situation leads to one of the main obstacles facing researchers in sports analytics - they would like their papers to be read and recognized. Having your sports paper buried in an academic journal with

little precedent for sports analytics research means that sports practitioners are unlikely to see your work.

1.3 Recognition

As an academic, you also want to receive recognition for your work. It is obviously not good to be perceived as the individual who works on wacky topics that are not considered serious and who publishes in suspect outlets. Therefore, to receive the traditional academic benefits (e.g. tenure, promotions, salary increments, grants, etc), it is in our best interest to publish in academically recognized journals.

And herein lies another issue associated with sports analytics research. My perception is that the field of sports analytics is exploding, more and more people are involved, and more and more people are blogging. I am increasingly finding myself referencing blog articles, and despite the general lack of credible reviewing involving blog papers, not all blog publications are bad. Some of them contain very good ideas. Therefore, a problem that the academic faces is turnaround time. The blogger can write something up, have it appear almost immediately and make an impact. Yet the academic may suffer from publication delay with traditional journals. And with the recently changing landscape (largely due to the big data phenomena), our work can be outdated by the time it appears in print. As a personal example, and this is no criticism of the journal, I first submitted a manuscript (Perera, Davis and Swartz 2018) to the *Journal of the Operational Research Society* on September 29, 2015. After several revisions, it has been accepted and will appear in the latter stages of 2018 (Issue 11).

1.4 What might you do upon graduation?

If you are a young researcher or perhaps a student, you might be nervous about putting too many eggs in the sports research basket. My response to this is don't be afraid - if you do good work, the good work will be recognized.

Many of the techniques that are used in sports research translate well to other domains. For example, the handling of big data is a skill that is widely sought after in both industry and academia. And big data is becoming an integral part of sports analytics. In the National Basketball Association (NBA), player tracking data records the (x, y) coordinates of every

player on the court and the ball at a rate of 24 times per second. You can do the arithmetic: 10 players on the court plus the ball, 24 observations/second, 48 minute games, 82 games per season and 30 teams. That is certainly big data. In other leagues, although player tracking is on the horizon, event data may be available. For example, in the National Hockey League (NHL), the company SPORTLOGiQ uses machine learning and optical recognition software to provide detailed event data occurring every 1.2 seconds on average.

In my own department, we have various success stories related to employment and sports analytics. For example, my colleague Luke Bornn is currently VP of Strategy and Analytics for the Sacramento Kings of the NBA, and has hired some of our students. My former MSc student Sarah Bailey is now working in analytics for the Los Angeles Rams of the National Football League (NFL). And three PhD graduates (Saman Muthukumarana, Harsha Perera and Rajitha Silva) are now working in academia as statisticians after completing theses that were entirely or mostly sports related.

One of the things that our “sports students” have experienced during job interviews is that the subject matter of their theses can be easily communicated and that there is often an interested audience.

1.5 Opportunities for sports research

It appears to me that problems in sports are plentiful. A reason for this may be that games are not played optimally. And a reason for lack of optimality may be tradition. Players are often coached by former players who were coached by former players, etc.

Another advantage of doing sports research is that you are often less reliant on others. For example, in biostatistics, it is typically difficult to do cutting edge research if you don't have a medical team supplying you with both data and problems. The beauty of sports research is that you are the subject matter expert. I have been playing and watching sports on television my entire life, and this has given me intuition for sports problems.

Our intuition regarding sport also facilitates modelling. We typically know what to measure in terms of performance (e.g. goals), we have at least a vague understanding of which covariates are important, games typically terminate in pre-defined amounts of time and sporting matches adhere to well defined rules. Most importantly, data are often available.

2 Sports Statistics Journals

There are not many pure sports statistics journals. However, at the top of this very short list is the

- Journal of Quantitative Analysis in Sports (JQAS)

JQAS is the best outlet for your sports statistics paper. It was founded in 2005 as a Berkeley Electronic Press (Bepress) publication where Ben Alamar served as the founding Editor from 2005-2011. In 2011, the journal became a publication of the *American Statistical Association* (ASA) where Jim Albert assumed the role of Editor and served from 2011-2014. The takeover was a monumental step for sports statistics researchers as the ASA seal of approval provides instant academic credibility for *JQAS* publications.

When you submit a manuscript to *JQAS*, you are guaranteed to have your paper handled by people who actually appreciate sport and who have statistical expertise. It is not easy to publish in *JQAS*; from July 2014 through June 2017, they had only a 10.8% acceptance rate (Glickman 2017). However, *JQAS* tends to review work quickly, and in my opinion, it is the flagship publication for sports analytics. It is also my impression that the *JQAS* editorial board is willing to work with authors; if they see a glimmer of a good idea in your manuscript, they may help you see it through to publication. Your *JQAS* paper will be visible to the sports analytics community and it will be recognized as a solid statistical contribution. What you do need to keep in mind is that *JQAS* is an ASA publication. Therefore, your paper ought to have significant novelty and it ought to involve data. In the near future, *JQAS* will likely have an impact factor, and this will open the door further for authors who require such accreditation.

A second quantitative sports journal that is gaining recognition is the

- Journal of Sports Analytics (JSA)

which was founded in 2015. The focus of *JSA* is a little different than *JQAS*. While *JQAS* is statistical, *JSA* entertains papers from other quantitative disciplines. *JSA* is less concerned with technical novelty but wants to see practical sporting contributions. *JSA* has a non-academic advisory board consisting mostly of analytics insiders from professional sports teams. For example, the celebrity Mark Cuban of the Dallas Mavericks is a member of

the Advisory Board. Currently, JSA produces two issues a year; I believe that keeping the number of publications low helps maintain standards.

3 Operations Research Journals

As a statistician, you may find it strange that I am providing a strong endorsement for journals from the field of Operations Research (OR) or Operational Research as it is known in the UK. The focus of OR may be described as decision making using quantitative tools. In my experience, OR journals are often open to sporting contributions. More generally, OR journals appear to be open to most application areas where problem solving techniques can be brought to the table. I believe that the field of OR puts considerable value on the problem solving component of research. Table 1 provides a list of OR journals and related Management journals which are sympathetic to articles on sports analytics. I have listed the number of sports papers that have appeared in these journals over the five year window 2013-2017. There may be other good OR/Management journals for sports analytics research; the journals listed in Table 1 are those that have come to my attention over years of research and those which have a high SCImago ranking (i.e. in the second quartile or above amongst Operations Research and Management Science journals). Manana-Rodriguez (2015) provides a critical review of SCImago rankings. Perhaps it is fair to say that the leading sport-friendly journals *EJOR* and *JORS* contain more papers on sports such as cricket and soccer, and have less of an American focus. The publication numbers for *IMAJMM* was aided by 8 papers that appeared in a special issue on “Mathematics in Sport” in 2013 (Volume 24, Issue 3).

One of the more popular OR-management topics is scheduling. Obviously, there are scheduling applications in sport. OR and management journals are also keen on resource allocation and efficiency assessment; this topics can also involve sport. For me, these topics are on the periphery of sports analytics since they are less concerned about the way sport is played.

Journal	SJR	Sports Pubs
European Journal of Operational Research (EJOR)	2.505	17
IMA Journal of Management Mathematics (IMAJMM)	0.857	10
Journal of the Operational Research Society (JORS)	1.004	9
Management Science	3.885	6
Annals of the Operations Research	1.009	6
Interfaces	0.837	4
Operations Research	2.940	2
Computers and Operations Research	2.326	2

Table 1: The 2016 SCImago Journal Rank (SJR) and the number of sports articles published in selected OR-type journals over the five-year window 2013-2017.

4 Mainstream Statistics Journals

I will begin by sharing two short and possibly amusing stories related to publishing in mainstream statistics journals.

Story #1: Back in 2008, I had a cricket paper that I thought was a nice contribution; it contained statistical modelling, statistical computing and utilized detailed ball-by-ball data that had not previously been exploited in cricket analytics. I was feeling good and was contemplating which lucky journal should receive my outstanding paper. I was thinking maybe an Indian journal where they truly appreciate the game of cricket. I came across the following description of *Statistical Methodology*, the Official Journal of the *International Indian Statistical Association*. It read, “Statistical Methodology provides a forum for original, high-quality articles on contemporary statistical theory as well as significant applications.” I thought to myself, perfect, that’s the journal for me. Within 20 minutes of my submission to *Statistical Methodology* (now discontinued after 33 volumes), I received the following email:

I am sorry to inform you that your manuscript is not suitable for Statistical Methodology.

I think the lesson from this story (which I have experienced more than once) is that some statisticians don’t want anything to do with a sports paper. It is beneath their dignity.

Story #2: Later, that same paper that was rejected by *Statistical Methodology* was sent to

The Canadian Journal of Statistics (CJS). This was a bold move on my part as I had never seen a sports contribution in *CJS*. I think *CJS* was nervous of our paper. We received seven (yes, 7!) reports; reports from five referees, an Associate Editor and the Editor. The reports totalled 15 pages. Our paper which was eventually published in *CJS* consisted of 18 pages (Swartz, Gill and Muthukumarana 2009).

Years later, the Editor (Paul Gustafson) told me that it was his practice to send out papers to multiple reviewers knowing that not all of them always respond. In the case of our paper, everybody responded, and they had experts from everywhere - computation, Bayes, cricket, etc.

I think that one of the take-home lessons from this experience (and other publication adventures) is that people have strong opinions on sport. I have found that reviewers often treat the assumptions in sports modelling in a very critical way. And I suppose this is a good thing. Reviewers typically have more to say about my sports applications than when I write about other applied topics. It seems that most everyone is an expert when it comes to sports analytics, including my barber and my bus driver.

Overall, I am suggesting that you may need a thick skin when submitting to mainstream statistics journals. Your sports contribution may simply be a non-starter for some editors and some reviewers. However, I think there are more reasonable people out there than unreasonable people. So, you might take a shot and submit your paper to a statistics journal. And I think that some of the very best statistics journals will take your sports research seriously. Keep in mind that your work should have a significant statistical contribution for publication in a top statistics journal.

Table 2 lists mainstream statistics journals that have come to my attention as having a track record of being sport friendly. Some of these journals are highly regarded. The non-technical magazine *Chance* is without doubt the leader in statistical articles on sport, and the tally in Table 2 does not include the on-again, off-again column “A Statistician Reads the Sports Pages”. Although *JASA* has only one sports publication in the window, they had a special issue on “Statistics in Sports” in 1994 where 8 papers were published. I suggest that the statistics for *TAS* in Table 2 are misleading; *TAS* has a long history of publishing sports papers. In addition to the two papers that appeared in the window, *TAS* has three more sports manuscripts have been accepted and will likely appear in 2018. I was surprised

to find only one sports article in the *A&NZJS*; they have published a couple of my sport papers over the years, and I know that they are open to sporting contributions.

Journal	SJR	Sports Pubs
Chance Magazine	NA	12
Journal of the Royal Statistical Society, Series A	1.368	8
Journal of Applied Statistics	0.402	8
Journal of Data Science	NA	4
The Annals of Applied Statistics	1.793	3
The American Statistician (TAS)	0.517	2
South African Statistical Journal	0.122	2
Journal of the American Statistical Association (JASA)	2.869	1
The Australian and New Zealand Journal of Statistics (A&NZJS)	0.365	1
Statistica Neerlandica	0.402	1
Brazilian Journal of Probability and Statistics	0.470	1

Table 2: The 2016 SCImago Journal Rank (SJR) and the number of sports articles published in selected mainstream statistics journals over the five-year window 2013-2017.

5 Special Issues

I love special issues on sports analytics. You know that your work is going to be reviewed in a fair context. You can also expect some exposure for your work since a collection of sport papers is more likely to attract attention than an isolated sports paper. Because sports analytics is now a hot topic, more and more special issues on sport are being considered. Here is a sampling of some recent special issues in sport:

- Statistical Analysis and Data Mining: The ASA Data Science Journal - Sports Analytics Special Issue (volume 9, issue 5 in 2016)
- Data Mining and Knowledge Discovery - Special Issue on Sports Analytics (volume 31, issue 6 in 2017)
- Electronic Journal of Applied Statistical Analysis - Special Issue on Statistics in Sports (volume 10, issue 3 in 2017)

- Machine Learning - Special Issue on Machine Learning for Soccer (publication likely in 2018)
- International Journal of Forecasting - Special Issue on Forecasting in Sports (publication likely in 2018)

One of the “tricks” these days is to have special issues tied to conferences (see Section 8). The publication opportunity provides an incentive for researchers to attend the conference. For example, the Athens Institute for Education and Research organizes an annual conference on sport. This conference is sponsored by the *Athens Journal of Sports* where many of the conference papers eventually appear.

Along the same lines as special issues in sport, I believe that edited books on sport provide an excellent opportunity for disseminating your work. The drawback here is that you typically need an invitation to write a chapter. There have been a number of sports analytics books that have been written over the years. At the moment, the *Handbook of Statistical Analyses in Sports* edited by Albert et al. (2017) provides the most current snapshot of research in sports analytics.

6 Sports Science Journals

There are many sports science journals. For example, a quick internet search on “sports science” (www.scimagojr.com/journalrank.php?category=3699) reveals 128 journals.

Although sports science journals may publish articles that are interesting to a statistician, the focus is different. The scope of these journals tends to be on topics such as physiology, training, exercise, psychology, biochemistry, biomechanics, health and medicine. The papers are also typically experimental in nature.

Here are some sports science journals that may entertain sporting contributions with a greater mathematical/statistical perspective:

- Journal of Sports Science and Medicine
- Journal of Sports Sciences
- European Journal of Sport Science

- Journal of Human Sport and Exercise

7 Other Journals

I admit that I do not have familiarity with all journals that publish sports papers. And because we are now in section 7 of the categorical assessment, one might assume that the following journals are lesser journals in some respect. This is not the case. Rather, I am merely listing journals that you may consider; I simply don't have much experience with these journals. I do not believe that any of the following journals are particularly statistical. The first list contains journals that contain "sport" in their title. You would think that this is a positive sign.

- Sports Engineering
- International Journal of Computer Science in Sport
- International Journal of Sports Science and Engineering
- Sport and Society
- Sport Management Review
- Sport Business and Management
- Athens Journal of Sports
- International Journal of Performance Analysis in Sport
- International Journal of Sport Management and Marketing
- International Journal of Sport Science and Coaching
- South African Journal for Research in Sport, Physical Education and Recreation

A journal which does not fit into any of the aforementioned categories is the *International Journal of Forecasting (IJF)*. *IJF* is neither OR, nor statistical, nor has "sport" in its title.

Yet *IJF* has a 2016 SCImago rank of 1.685 and has published three sporting papers over the five-year window 2013-2017 in addition to the special issue mentioned in Section 5.

You may have observed that I did not include any economics journals in the lists. I do know that there are some economics journals that publish sports papers such as the *Journal of Sports Economics*. Personally, I have had bad experiences with economics journals. My impression is that you need to speak their language, and I have not figured out how to do that.

I have also not discussed any journals that are sport specific (e.g. *International Journal of Golf Science*). Whereas these journals may accept your analytics paper, they tend to be everything for everyone in the sport, covering topics from the social sciences to the hard sciences.

8 Conference Proceedings

Conference proceedings are perceived differently across different academic disciplines. For example, I know that in Computing Science, there are conferences where presenting a paper is considered a major achievement. At the other end of the spectrum, I know of meetings where anybody can present on any topic as long as they pay their registration fee.

That being said, there are three longstanding meetings in sports analytics that produce proceedings. They are:

- Mathematics and Computers in Sport Conference (a biennial Australian meeting that began in 1992)
- MathSport International (a biennial meeting that began in 2007 which is held in either the UK or in Europe)
- MIT Sloan Analytics Conference (now an annual meeting that began in 2000)

The above meetings are all enjoyable but are different. For example, in the early days, I think it is fair to say that the Sloan meeting was largely attended by industry insiders and those seeking analytics positions. In those days, you might not have heard much groundbreaking research since teams are reluctant to give away their “secrets”. In the last few years,

Sloan has grown to the point where the meeting is sold out, it attracts a media frenzy, there is a fierce competition for speaking slots and the presentations are high-level. The Sloan meeting still has an aura of a business meeting; perhaps this is not surprising given that Sloan is the MIT School of Management. To their credit, the Sloan conference maintains a webpage (www.sloansportsconference.com) where video presentations and papers can be accessed. The other two sport conferences have more of an academic flavour. MathSport International now makes their proceedings available (www.mathsportinternational.com/previous.html). With the Mathematics and Computers in Sport Conference, it is difficult to get your hands on the proceedings unless you attend the conference.

Sports Science groups also have their own meetings with proceedings. For example, the European College of Sports Science hosts a large annual Congress from which they maintain a Book of Abstracts. However, my impression is that they don't care greatly about technical work and statistics in particular. You may wander around the meeting for days and not find anybody of your type.

9 Blog Sites

The danger with blog sites is that there are limited controls on quality. But clearly, blog sites provide a great avenue for sharing ideas. Sometimes the demarcation between a blog site and the site of a professional organization which enforces standards is blurred. Most of the popular sports blog sites are sport specific and some are backed by news organizations. There are too many sites to list. Here is a sampling of some sites that contain more thoughtful analytics discourse than ranting:

- General Sports - FiveThirtyEight (fivethirtyeight.com)
- Basketball - HoopsHype (hoopshype.com) and 82games (82games.com)
- Baseball - FanGraphs (fangraphs.com)
- Hockey - Hockey-Graphs (hockey-graphs.com)
- Football - Football Outsiders (footballoutsiders.com)

- Soccer - StatsBomb (statsbomb.com) and 5 Added Minutes (5addedminutes.com)

There have been some good sports analytics blog sites (e.g. WAR On Ice, Advanced Football Analytics) that have been discontinued. In both of these cases, the bloggers' skills were recognized and more lucrative positions were offered in the sports world.

10 Discussion

I hope that this opinionated discussion may be helpful to those considering research in the expanding field of sports analytics. In recent years, sports datasets have become larger, are more widely available and have increasingly complex structures. With such developments, the opportunity for sports analytics research is great. In discussing quantitative research in golf, Larkey (1998) provided the following quote which also seems relevant to the general field of sports analytics:

Some of these theoretical and and methodological issues may turn out to be at least as important as the n th marginal contribution to heavily studied, more traditional areas.

Whereas the major sports have received the most attention in sports analytics, the opportunity to do work in smaller sports and women's sport is vast. Do good work, and get your work out there for people to read. Good luck.

11 References

- Albert, J.A., Glickman, M.E., Swartz, T.B. and Konig, R.H. Editors (2017). *Handbook of Statistical Methods and Analyses in Sports*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods.
- Duckworth, F.C. and Lewis, A.J. (1998). A fair method for resetting targets in one-day cricket matches. *Journal of the Operational Research Society*, 49, 220-227.
- Duckworth, F.C. and Lewis, A.J. (2004). A successful operational research intervention in one-day cricket. *Journal of the Operational Research Society*, 55, 749-759.
- Glickman, M.E. (2017). Discussion of practical problems in Sports Analytics. *JQAS Invited Session 490 at the Joint Statistical Meetings*, Baltimore.

- Goldsberry, K. and Weiss, E. (2013). The Dwight effect: A new ensemble of interior defense analytics for the NBA. *MIT Sloan Sports Analytics Conference*. Retrieved online December 5/17 at <http://www.sloansportsconference.com/content/the-dwight-effect-a-new-ensemble-of-interior-defense-analytics-for-the-nba/>
- Larkey, P.D. (1998). Statistics in golf. In *Statistics in Sports*, edited by J. Bennett, Arnold Applications of Statistics Series, Arnold Publishers, New York, 121-140.
- Lewis, M. (2003). *Moneyball: The Art of Winning an Unfair Game*. WW Norton, New York.
- Manana-Rodriguez, J. (2015). A critical review of SCImago journal and country rank. *Research Evaluation*, 24, 343-354.
- Perera, H., Davis, J. and Swartz (2018). Assessing the impact of fielding in Twenty20 cricket. *Journal of the Operational Research Society*, 69, pp not yet determined.
- Stern, S.E. (2016). The Duckworth-Lewis-Stern method: Extending the Duckworth-Lewis methodology to deal with modern scoring rates. *Journal of the Operational Research Society*, 67, 1469-1480.
- Swartz, T.B., Gill, P.S. and Muthukumarana, S. (2009). Modelling and simulation for one-day cricket. *The Canadian Journal of Statistics*, 37, 143-160.