Proceeding

7th INSHS International Christmas Sport Scientific Conference, 9-12 December 2012. International Network of Sport and Health Science. Szombathely, Hungary

The hot hand phenomenon: measurement issues using golf as an exemplar

LUKE HEATH ✉ , NIC JAMES, LEON SMART

Health and Social Sciences: London Sport Institute, Middlesex University, Hendon, London, UK

ABSTRACT

Heath L, James N, Smart L. The hot hand phenomenon: measurement issues using golf as an exemplar. *J. Hum. Sport Exerc.* Vol. 8, No. Proc2, pp. S141-S151, 2013. The so called "hot hand" phenomenon (phrase coined by American basketball fans) refers to the commonly held belief amongst sports fans, coaches and players that sports players can, on occasion, produce sequences of consistently superior (compared to their normal) performance. Gilovich, Vallone and Tversky (1985) first alerted the scientific community to this phenomenon but concluded that the belief in the hot hand was nothing more than a misperception of random sequences. Researchers have argued that the statistical methods used to test the hypothesis of randomness and its antithesis streakiness (a more inclusive term for hot hand) were flawed. Golf was used as an exemplar for a sport where there is no clear success and fail outcome, unlike basketball shooting. Success in golf has usually been compared against "par or better" or "birdie or better" although these measures are the same for all players irrespective of their standard and don't account for the hole difficulty on the day the hole is played. A momentum chart was used to identify potentially streaky performance and this data was then compared with the non-streaky data using the chi square test of independence. This approach would seem to satisfy a perceived limitation in previous approaches to evaluating streakiness where the whole sequence of performances is assessed for streakiness as opposed to just the streaky portions. Furthermore this approach allows streaky performance to contain a few non-successful performances. However future research needs to test this methodology out more rigorously using an appropriate criterion for success that is relevant to the skill level of the performer and the hole difficulty. **Key words**: STREAKINESS, MOMENTUM, GOLF, SUCCESS CRITERIA, SKILL LEVEL.

INTRODUCTION

The so called "hot hand" phenomenon refers to the commonly held belief amongst sports fans, coaches and players that sports players can, on occasion, produce sequences of consistently superior (compared to their normal) performance, sometimes referred to as streaky performance. This prompted American basketball fans to call a player, perceived to be in such a state, as having "hot hands". Gilovich, Vallone and Tversky (1985) first alerted the scientific community to this phenomenon but concluded that the belief in the hot hand was nothing more than a misperception of random sequences. The reason for this finding was perhaps that the complexity of the analysis did not match the complexity of the sporting situation examined. For example, a simple binomial model (commonly referred to as the *coin-tossing model*) assumes that the probability of success is constant with a constant hit rate (stationarity). However, such a model may not translate well to sport, such as in the case with basketball, where players' performances are known to fluctuate over time (the non-stationarity argument) (Bar-Eli, Avugos and Raab, 2006). Alternatively, it could be argued that the authors' views were somewhat biased despite whatever data they might obtain, since fans' perceptions were selected as the 'gold standard' over expert opinion (Wardrop, 1999). Nevertheless, whilst their paper was meticulous and often thought provoking Gilovich et al., have subsequently encouraged hundreds of similar studies in search of the elusive hot hand, since almost all sports fans believe that they have witnessed periods of play where performance is elevated beyond the normal. Larkey, Smith and Kadane (1989) were one of the first to challenge Gilovich and colleagues on their research after evaluating a new data set of players from the 1987 – 1988 NBA season. However, the results were somewhat similar. In their defense the authors' suggested that basketball tournaments were perhaps too rich an environment for extracting individual player sequences and that data analysis used for detecting hot hand streaks should be restricted to shorter periods within a game. Wardrop (1995) reexamined Gilovich et al.'s (1985) free throw data (ie. free throw data for the Boston Celtics from the 1980 – 1981 and 1981 – 1982 season) interpreting things from a slightly different perspective. He showed that when the data was aggregated over players then fan's perceptions of the hot hand were deemed to be true: Players made a basket 79% of the time after a hit compared to 74% of the time after a miss. His findings also showed that the Boston Celtics, as a group, were more likely to score a hit on their second attempt than on their first attempt, contrary to Gilovich and colleagues. Smith (2003) argued that Gilovich et al.'s (1985) basketball data did not take into account confounding variables, such as shot election, time taken between shots and game strategy. For example, a player's two successive shots might be taken before and after the halftime intermission or even in different games within tournaments. This he reasoned is not what fans perceived to be hot hands. Smith found horseshoe pitching data to be much cleaner and thus more effective at detecting hot and cold hand streaks, since every pitch is made up of the same distance at regular intervals and winning competitions possesses fewer strategies. Furthermore, findings from the analysis of the 2000 and 2001 World Championships revealed that top-class pitchers show moderate spells of hot and cold hands. Vergin (2002) extended our understanding of the hot hand by examining the momentum of 28 major league baseball teams and 29 NBA teams over the length of a season. However, despite the findings that winning and losing were independent of the results from previous games, he emphasized that winning in sports should not be considered a random event.

Hot hand in golf
Different sports have been analysed on the basis that some may be more prone to streaky performance due to the nature of the sport. For example, golf involves relatively long pauses between shots and if, as is commonly suggested, a mechanism that accounts for streakiness is psychological momentum, then golf may be prone to these effects more than some other sports. However, golf is different to basketball shooting and other sports where there is a clear success (score a basket) and fail (miss) outcome. In golf

each shot can be rated on a scale from excellent to poor but this classification is both player dependent and arbitrary. Of course some shots could go directly into the hole but for the vast majority of shots this is not likely or indeed possible. Clark (2003) investigated the occurrence of streakiness among professional golfers on the PGA Tour and Senior PGA Tour. The results showed that there was a significant tendency for players' "par or better" rounds to cluster together and for players' "above par" rounds to cluster together. However, this finding was attributed to course difficulty. To nullify this effect Clark (2005a) studied hole-to-hole scores within 747 tournaments using "par or better" as an indication of a successful hole and "above par" as an unsuccessful hole. He later re-investigated the data using "birdie or better" and "par or worse" (Clark, 2005b) but on both occasions found successive hole-to-hole scores were largely independent. Par is the number of strokes a professional golfer is expected to take to play any given hole based on its length and difficulty, but this is not a standard measure, as holes less than 475yrds in length can be a par 5 and those longer than 475yrds in length can be a par 4 (Hale & Hale, 1999). Add to this the fact that the top golfers do not play in the same tournaments as each other (typically 10 to 15 common events) thus any comparison of performances between players involves scores that have been accumulated under different conditions, with widely different prizes and field quality (Larkey, 1994). Rees and James (2006) argued that Clark's (2005a, b) criterion for judging success might have been too simplistic since the environmental factors coupled with the skill level of the golfer on each day would mean that a par on some holes could be classified as unsuccessful whereas on others it may be classified as successful. They argued that their analysis of hole average for the field on each day was considered a more suitable method for analysis, as it allowed for the strength of the players in that particular competition. However, the authors also emphasized that hole average tended to discriminate between players in terms of streakiness because it may only be the most appropriate measure for the average players. More recently, Livingston (2012) investigated hole-to-hole performances of four types of professional golfer (Nationwide Tour, PGA Tour, LGPA Tour, and Champions Tour) using "birdie or better" as an indication of a successful hole and "bogie or worse" as indication of an unsuccessful hole. The results showed glimpses of hot and cold hands when the effect was allowed to vary with a player's amount of experience as a professional that would otherwise have been masked. However, no consistent pattern of how hot or cold hand effects vary with experience as a professional could be found.

Psychological theories of momentum
Psychologists typically look for connections between past and present performance using theories of psychological "momentum". The term "momentum" we define here as being a state of emotional arousal marked by reductions or elevations in a player's performance (Adler, 1981). Adler (1981) defined psychological momentum as a bidirectional force, one which affects either the probability of losing or the probability of winning as a function of the outcome of the preceding event. Put simply, a shot will be influenced by the outcome of a previous shot. Early models of momentum by Silva, Hardy and Crace (1988) and Vallerand, Colavecchio and Pelletier (1988) provided evidence, although preliminary, of the cognitive process associated with performance and outcome. Momentum, as it appears, can be lost or gained from two alternative effects: positive inhibition and negative facilitation (Silva et al., 1988). Positive inhibition is the process whereby success results in a loss of momentum, thus increasing the possibility of subsequent failure, whereas negative facilitation is the situation whereby failure can lead to increases in the probability of success. Using these earlier models, Taylor and Demick (1994) constructed a "multidimensional model of momentum" in sport, one which could identify how past events can influence future outcomes through categorizing each step of the process (momentum chain). Their paper suggested that there would be a significantly greater number of changes in immediate outcome following a precipitating event than when no precipitating event occurred. Thus ruling out the assertion that the relationship between precipitating events and immediate outcome could be due to chance alone.

Criterion for judging success in golf

To date many studies have suggested reasons for streaky performance, such as incidences where several good shots may make it easier for players to recall past successes, expect more good shots and pay minimal attention to performance processes (Clark, 2003; Rees & James, 2006; Livingston, 2012) or the concept of momentum which implies early success leads to future success (Adler, 1981; Silva et al., 1988; Vallerand et al., 1988; Taylor & Demick, 1994; Landsberger & Beauchamp, 1999). Each of these ideas seem reasonable although from a golf perspective is it reasonable to expect the hole score to accurately reflect these processes? For example, the widely cited model score of "par" on a hole, which occurs roughly 60 percent of the time and is viewed as the norm, must take into account scores "above par" being perceived as failures and scores "below par" being perceived as successes (Livingston, 2012). However, a player may be playing particularly well, perhaps even streaky in nature, when a birdie putt hits the hole and lips out resulting in a par or worse. Current analysis techniques would deem a streaky run to have ceased at this point. Is this necessarily the case however? Since we hypothesize that streakiness is a relatively unusual phenomenon, analysis of a complete series of performances using previous methods is therefore unlikely to yield evidence of streakiness. Instead calculating thresholds for what could constitute streakiness in golf should reveal normal performance parameters and therefore determine whether short sequences fall outside those which would occur randomly.

Statistical analysis used for detecting hot hands

In their paper Gilovich et al. (1985) operationally defined "hot hand" using several within-game measures (e.g., analyses of conditional probabilities, run counts, and serial correlations) and one between-game measure (a comparison of the observed and expected variability in shooting percentages across games). This method was selected based on fans responses from an initial enquiry that shoot sequences were non-stationary and serially dependent. From this they performed four sets of calculations for each player's sequence of hits and misses:

- A long-run probability of a player's hit compared to the conditional probability of a hit following one, two and three hits, and following one, two or three misses.
- A serial correlation (first-order correlation coefficient).
- A Wald – Wolfowitz Runs Test.
- A chi-square test based on 4 shots in successive groups, with each block classified according to hit rate: "high", "medium" or "low".

Wardrop (1999) questioned whether these tests were suitable for analysing the outcome of successive shots in basketball. He argued that both the runs test and the test of fit used by Gilovich et al. (1985) had little power in detecting non-stationarity. Furthermore, a runs test and a serial correlation test give virtually the same results (Wardrop, 1999). Albright (1993) analysed streakiness in professional baseball players over a 3-year period using a logistic regression model. This allowed the data to include situational effects, such as pitcher's earned run average, pitcher's handedness and home and away status of the game. Moreover, because the author had access to several years' worth of data, it was possible to observe whether a player was particularly streaky or whether the length of streaks continued over a season. As it happens, Albright found that some batters exhibited streakiness in some seasons, although they did not do so consistently. Stern and Morris (1993) commented that Albright's (1993) lack of evidence was partly due to a bias in the logistic regression, due to a small effect size. However, Stern (1995) also admitted that evidence of streakiness could be found in Albright's data if players' performances were combined. Forthofer (1991) classified the most extreme players from newspaper box scores of all NBA players in all games during the 1980 – 1990 season (i.e., the game summary of shots taken and made by each player) in order

to produce several categories of what he deemed streak shooters (ie. those with ''hot and cold streaks'', those with ''hot hands'', and those with ''cold hands''). The results showed that of the 123 players examined, 17 players exhibited some form of streakiness.

Hooke (1989) emphasized the inherent difficulty of using statistical methods to study complicated situations, such as basketball and other sports. He reasoned that each sport has its own set of specific circumstances. Therefore the methods used to detecting hot hands in one sport may not translate well across other sports. According to his own intuition ''the alternative hypothesis is not that there is a 'hot hand' effect that is the same for everyone, but that the real situation is much more complex''. Since the work of Gilovich et al. (1985) most research has continued to examine actual sports data in an effort to determine if there are sports where hot hands can be detected. However, whilst most of the research has concentrated on either finding support or rejecting the hypothesis of the hot hand using archival and simulated data sets, theoretical improvements have been almost neglected (Bar-Eli et al., 2006). Crust and Nesti (2006) suggested that future research should adopt a qualitative approach, one that explores the mechanisms by which "psychological momentum" and its closely related phenomenon "hot hand" might occur. They argue that while many researchers have acknowledged the difficulties surrounding different modes of testing, the resultant research has probably led to more confusion than clarity.

New methodological approach to evaluating streakiness
An alternative approach is to consider theoretically what constitutes both normal and streaky performances for a particular individual or team. If two (three including unusually bad performances) distinct performance profiles could be identified then an alternative statistical procedure could be used to determine if they were indeed significantly different. How might these periods of unusual and usual performance be identified? One approach could be to produce momentum charts which record performance over time, where performance is rated as good or bad (although this does not have to be dichotomous) such that positive performances cause the plot to rise on the vertical axis and poor performance cause the plot to fall. An extension to this simple example would mean that varying degrees of good performance cause the plot to rise proportionately in relation to the extent to which the performance was judged good on a Likert type scale. Poor performances would have the opposite effect whilst neutral performances would not affect the height on the y axis. Using a graphical representation like this would facilitate the identification of sequences of performance that might be deemed streaky, as the plot would exhibit a steady incline or decline over a relatively long period of time in comparison to the rest of the plot.

This approach to trying to distinguish streaky from non-streaky performance might be criticized as simply trying to determine the difference between unusually good or bad performance from normal performance since no mention has been made of the requirement of a continuous run of successful performances. However the requirement for continuous runs seems to be determined by some statistical tests, in order for significance to be achieved, rather than a prerequisite for a logical definition of streakiness. Indeed, logically a run of say fifty successful performances interspersed with just two unsuccessful ones could be deemed streaky. For example if the two unsuccessful performances had been unlucky not to succeed due to external factors it would be rather punitive to suggest a run of successful performances had ended. However this is precisely what some statistical tests do.

Having suggested that streaky performance may exist even when some performances within a sequence were unsuccessful it may be asked what the difference is between streaky performance and excellent performance? Perhaps the answer lies in the duration of the performance. For example a player might hit an outstanding forehand volley in tennis. This could easily be deemed an excellent shot but could it

constitute streaky performance? No is the simple answer because there is no sequence of performances, it was just one event. However once two consecutive outstanding shots are considered then we must consider this to be potentially streaky since there is now a sequence. This is where there does not seem to be strong evidence in the literature as to the time frame under which streakiness does, or is likely, to occur. Theoretically this is likely to be related to the level of expertise of the performer(s). For example a novice basketball player who makes about 1 in every 50 free throws may produce a streaky performance by simply making two or three throws in a row. On the other hand an elite player with a free throw success rate of 80% may need to make 50 or 100 throws in a row to be deemed streaky. Critically, in statistical terms, the sequence thought to be streaky would need to be shown to be significantly different from any sequence expected by chance for the ability level of the player. The difficulty is determining where the sequence starts and ends and whether the sequence could include one or more unsuccessful performances within it. Then an appropriate statistical technique would need to be identified and applied.

Considerations for assessing streakiness in golf
Identifying a streaky performance is not straightforward if the parameters for such a performance are not known! However if we could identify a performance that was so unusual that it would certainly be streaky, if streaky performance exists of course e.g. the best round ever played in professional golf, and compared this performance against normal performance for that player, then provided the statistical test was appropriate we would have some evidence for streaky performance. Of course this doesn't help in the identification process but at least we would have a methodology for testing a sequence of performances. However there is always the problem of categorization. We may inadvertently identify a streak as normal performance and vice versa. For example, if we compared a streaky performance against a normal performance and the normal performance contained a streaky performance that we failed to identify then the statistical test would be compromised.

Identifying unusual performances also depends on how performances are measured. For some performances this is reasonably straightforward, a basketball shot either goes in or doesn't and is therefore good or bad. Many other performances are not as simple as is the case for golf performance. Professional golfers hit around 72 shots on any particular tournament day and if reasonably successful will play four consecutive days. They will then have a break of between a couple of days and a few weeks before repeating this pattern. Golf is characterized by the need to play different types of shots (any of which could potentially be streaky), on different holes (requiring different strategies) and over different rounds. Streaky performance may exist at any of these levels but how can performance be judged so that good and bad performances are discriminated? A number of golf studies have offered solutions using methods of varying complexity although none have provided a satisfactory solution that answers all of the problems identified above. However this can be said for all studies of streakiness and is probably the reason for its popularity as a research topic.

The methodology for assessing golf performance has mainly focused on hole scores of which there are 18 in a round, usually played on a single day (weather conditions being the reason for some divergence occasionally). A single round takes about 4 hours so a typical round of golf involves playing a shot, walking for a few minutes and playing the next shot. This inter-shot time has been proposed as a mechanism for promoting streakiness (using a cognitive explanation for streakiness such that this time allows the player to utilize psychological factors to promote and maintain good runs of performance) but also the opposite, perhaps the mechanism for enabling streakiness doesn't last for 4 hours or more.

From a statistical perspective 18 performance values (hole scores) is a convenient number of performances against which other rounds of 18 hole scores could be compared. However evaluating the performance of an individual hole score is not straight forward. Golf has a simple mechanism for this in that each hole is given a target number of shots in which the hole should be completed, called the par. The distance of the hole determines this number and consequently all holes are awarded pars of 3, 4 or 5. This simple solution allows a categorization for any individual performance against par e.g. scoring better than par may be deemed successful, par as neutral and worse than par unsuccessful. Other variations are possible and have been presented previously in this paper. However, as already mentioned, using par as the arbiter of success fails to recognize that on some holes par will be relatively simple to achieve whereas on others this may be very difficult. This is because holes with the same par may have quite different lengths, they may have markedly different topography and have different environmental conditions e.g. the wind may be following or against the direction of play. To compound this it is usual that each hole has different playing characteristics on a daily basis because the hole is usually changed slightly for each day of a competition and the environmental features are also changeable. These considerations have led some authors to suggest that par is not the perfect measure for deriving success. They have suggested comparing a player's hole score against the average hole score for all of the competitors on each particular day. Intuitively this seems sensible although this does not account for the individual's normal performance level and hence does not satisfy the criteria presented above. This is not insurmountable however as either the criterion for success could be manipulated to account for the player standard or the different scores for an individual could be compared between rounds.

An exemplar for assessing streakiness using golf
Eleven simulated round scores were created where each hole score was converted to the score relative to par such that a 3 on a par 4 (birdie) was given a score of 1, a par was awarded a 0 and a 6 on a par 5 (bogey) awarded -1. One round was exceptionally good and thought to be streaky (12 under par and equivalent to one of the best rounds produced in professional golf) whilst the other 10 rounds were fairly average (mean = 0.4 shots over par, SD = 2.63 shots from par) and considered not streaky. One method for presenting categorical data relating to performance levels is to use a momentum chart, so called because the plot presents how successful and unsuccessful performances occur over time. Whilst the terms momentum and streakiness could be used interchangeably we suggest that momentum simply describes a trend such that a player with the momentum would have an inclining momentum chart whereas a declining momentum chart would suggest the opponent had the momentum or the player has negative momentum. Thus having momentum may result in a streak (significantly better performance than normal) but this is not a necessary requisite. We believe this is what commentators are describing when they suggest a team or player has the momentum. If a momentum chart is plotting performance of two competing opponents then when one team or player has momentum it is due to the differential in performance between the two opponents i.e. one may be performing well, the other may be performing poorly or both. Thus a momentum chart simply presents the trend of performance and does not confirm or deny statistical significance. On this basis a momentum chart for the scores relative to par for the successive rounds (Figure 1) was plotted and clearly showed a consistent elevation of the plot for the exceptional (potentially streaky) round.
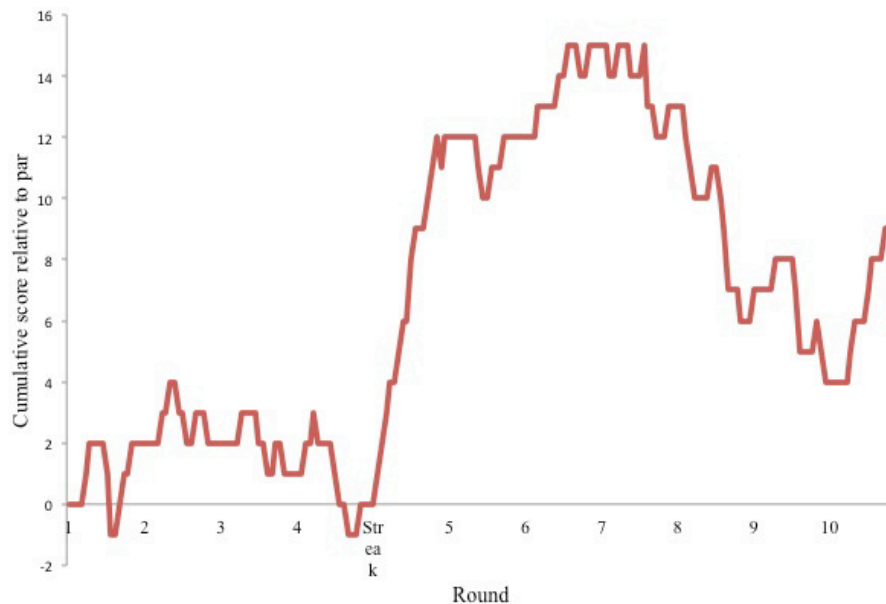
Figure 1. Momentum chart of scores relative to par.

Whilst the momentum chart appeared to show the difference between the so called streaky round (relatively linear rise in the plot) in comparison to the other rounds (varying rises and falls), this did not, as explained, provide statistical evidence of a difference. However it did enable the categorization of the performance data into streaky and non-streaky performance for subsequent analysis.

Analysis technique
Using this technique of splitting the performance data into two i.e. streaky and non-streaky enables new statistical techniques to be used that were not possible otherwise. The data could be split into rounds and a repeated measures analysis of variance used where post hoc pairwise comparisons used to determine if the streaky round is significantly different to the other. An analysis of variance for categorical data (Light & Margolin, 1972) would need to be used although the weakness of this approach is that the 11 rounds would be considered as different levels for one independent variable rather than being able to specify two levels, streaky and non-streaky.

A simpler method that satisfies this data well is the chi square test for independence. A two-way contingency table was constructed (Table 1) and scoring relative to par was found to be dependent on performance type (chi-square = 29.62, df = 4, $p < 0.001$, phi = 0.39).

Table 1. *Contingency table of frequency of scores relative to par for streaky and non-streaky performance.*

			Performance	
			Non-Streaky	Streaky
Score relative to par	Double bogey	Count	4	0
		Expected count	3.6	0.4
		% within group	2.2%	0.0%
	Bogey	Count	27	1
		Expected count	25.4	2.6
		% within group	15.2%	5.6%
	Par	Count	116	5
		Expected count	109.9	11.1
		% within group	65.2%	27.8%
	Birdie	Count	31	11
		Expected count	38.1	3.9
		% within group	17.4%	61.1%
	Eagle	Count	0	1
		Expected count	0.9	0.1
		% within group	0.0%	5.6%

Whilst this technique has determined that this player's performance relative to par was different on the streaky round compared to the non-streaky rounds, which is indicative of streaky performance, the fundamental problem that using par as a criterion for success fails to recognize that some pars are likely to be easier to achieve than others. In this example it may have been the case that the round in which the player scored 12 under par may have been on a significantly easier course than the other rounds were played on. This would invalidate the chi-square result but difficult to determine using just the data here.

This paper has argued that a player's hole score against the average hole score for all of the competitors on each particular day is a fairer comparison. However the hole average may only be a fair criterion for success for the average player. We have also argued that streakiness is related to the level of expertise of the performer and thus what might constitute successful performance for one player may not be successful for another. Consequently the hole average, whilst a good measure for the difficulty of the hole, does not constitute a fair criterion for success for all players. It is necessary, therefore, to adjust the hole average to reflect the skill level of the player. At this point in time there is no research in this area and thus work needs to undertaken to determine how this might be achieved.

CONCLUSIONS

The phenomenon known as the hot hand refers to a series of performances that are, in their entirety, above those normally associated with the player or team in question. This sequence of performances has widely been assumed to consist entirely of successful performances in order that the sequence may be deemed hot, or streaky. However this paper has argued that one or two non-successful performances within a relatively long run of successful performances may not necessarily invalidate the sequence as being streaky. Traditional statistical techniques do not allow for this assertion and so an alternative method for evaluating streakiness was presented using a momentum chart. These charts plot performance over time where successful performances produce inclines in the plot over time and unsuccessful performance

causes declines. A relatively consistent steep incline in the momentum plot in comparison to the rest of the plot signals the potential presence of a streaky performance profile. If the data associated with this steep incline is then compared with the rest of the data then a statistically significant difference would suggest the presence of streaky performance.

For some sports the criteria for success is straightforward as performance can only be successful or unsuccessful. However in other sports such as the example given in this paper, golf, performance does not fit into this dichotomous category set. For these performances some criteria for judging successful and unsuccessful performances must be used. We argue that this must be both comparable between performances and appropriate for the individual. Thus the par for a hole is not as appropriate as the average score by all players in a tournament on a particular day as the par for a hole does not distinguish between difficult and easy holes. However the hole average is only an appropriate measure of success for the average player and so some manipulation of this value needs to be undertaken for each player to account for their skill level.

Future research needs to evaluate the strength of these convictions using data from different sports. To date there does not appear to be a satisfactory method for creating an appropriate criterion for success in sports where performance is not easily categorized as successful or unsuccessful and this is something that is urgently required. Similarly the statistical techniques used to evaluate streakiness have been generally considered as inadequate. This paper has presented a novel approach to identifying potential streakiness and a statistical method but future work should continue to seek out the most appropriate statistical methods for determining streakiness.

REFERENCES

1. ADLER P. *Momentum: A theory of social action*. Beverly Hills, CA: Sage. 1981.
2. ALBRIGHT SC. A statistical analysis of hitting streaks in baseball. *Journal of the American Statistical Association*. 1993; 88:1175-1183.
3. BAR-ELI M, AVUGOS S, RAAB M. Twenty years of "hot hand" research: Review and critique. *Psychology of Sport and Exercise.* 2006; 7:525-553.
4. CLARK III RD. Streakiness among professional golfers: Fact or fiction. *International Journal of Sport Psychology.* 2003; 34:63-79.
5. CLARK III RD. Examination of hole-to-hole streakiness on the PGA Tour. *Perceptual and Motor Skills.* 2005a; 100:806-814.
6. CLARK III RD. An examination of the "hot hand" in professional golfers. *Perceptual and Motor Skills.* 2005b; 101:935-942.
7. CRUST L, NESTI M. A review of psychological momentum in sports: Why qualitative research is needed. *Athletic Insight*. 2006; 1:1-15.
8. DORSEY-PALMATEER R, SMITH G. Bowlers' hot hands. *The American Statistician*. 2004; 58:38-45.
9. FORTHOFER R. Streak shooter – The sequel. *Chance*. 1991; 4:46-48.
10. GILOVICH T, TVERSKY A, VALLONE R. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*. 1985; 17:295-314.
11. HALE T, HALE GT. Analysis of performance in the open championship 1892-1997. In *Science & Golf III: Proceedings of the World Scientific Congress of Golf.* (edited by A.J. Cochran and M.R. Farrally). Champaign, Il.: Human Kinetics. 1999. Pp. 395-403.
12. HOOKE R. Basketball, baseball, and the null hypothesis. *Chance*. 1989; 2:35-37.

13. LANDSBERGER LM, BEAUCHAMP PH. Indicators of performance momentum in competitive golf: An exploratory study. In *Science & Golf III: Proceedings of the World Scientific Congress of Golf.* (edited by A.J. Cochran and M.R. Farrally). Champaign, Il.: Human Kinetics. 1999. Pp. 353-362.

14. LARKEY PD. Comparing players in professional golf. In *Science & Golf II: Proceedings of the World Scientific Congress of Golf.* (edited by A.J. Cochran and M.R. Farrally). London: E & FN Spon. 1994. Pp. 193-198.

15. LARKEY PD, SMITH RA, KADANE JB. It's okay to believe in the ''hot hand''. *Chance*. 1989; 2:22-30.

16. LIGHT RJ, MARGOLIN BH. An analysis of variance for categorical data. *Journal of the American Statistical Association*. 1972; 66:534-544.

17. LIVINGSTON JA. The hot hand and the cold hand in professional golf. *Journal of Economic Behavior & Organization*. 2012; 81:172-184.

18. REES C, JAMES N. A new approach to evaluating streakiness in golf. In: Dancs H, Hughes MD, O'Donoghue PG. (2008). Performance analysis for sport 7. Cardiff: UWIC CPA Press. 2006. Pp. 352-360.

19. SILVA JM, HARDY CJ, CRACE RK. Analysis of psychological momentum in intercollegiate tennis. *Journal of Sport and Exercise Psychology*. 1988; 10:346-354.

20. SMITH G. Horseshoe pitchers' hot hands. *Psychonomic Bulletin & Review*. 2003; 10:753-758.

21. STERN HS. Who's hot and who's not: Runs of success and failure in sports. In 1995 Proceedings of the section on statistics in sports. American Statistical Association. 1995. Pp. 26-35.

22. STERN HS, MORRIS CN. A statistical analysis of hitting streaks in baseball: Comment. *Journal of the American Statistical Association*. 1993; 88:1189-1194.

23. TAYLOR J, DEMICK A. A multidimensional model of momentum in sports. *Journal of Applied Sport Psychology*. 1994; 6:51-70.

24. VALLERAND RJ, COLAVECCHIO PG, PELLETIER LG. Psychological momentum and performance inferences: a preliminary test of the antecedents-consequences psychological momentum model. *Journal of Sport and Exercise Psychology*. 1988; 10:92-108.

25. VERGIN RC. Winning streaks in sports and the misperception of momentum. *Journal of Sport Behavior*. 2000; 23:181-197.

26. WARDROP RL. Simpson's paradox and the hot hand in basketball. *The American Statistician*. 1995; 49:24-28.

27. WARDROP RL. Statistical tests for the hot-hand in basketball in a controlled setting. http://www.stat.wisc.edu/wardrop/papers/tr1007.pdf. 1999.