

Stableford: Equitable golf-scoring system or quality-of-life measure?

David Massel, David McCormack, James Lewis, Jeffrey Mahon

Golf is a game that requires a person to strike a small ball with a club from the teeing ground into a distant hole while following The Rules. It is supposed to be fun. Keeping score is part of the joy. Most people understand the simplest scoring method: one swing at the ball — usually resulting in a more-or-less successful hit — equals one stroke. The lowest score wins, which is fair as long as the competitors are of similar ability. The use of a handicap factor — an allowance of strokes given to a player based on past and current performance — permits golfers of all levels to compete together on an equitable basis.

Even though the handicap system works rather well, some still believe that golf is simply not fair. One advocate for a more satisfying scoring method was Dr. Frank Barney Gordon Stableford (1870–1959), a first-rate golfer and physician who served with distinction and decoration in the Royal Army Medical Corps during the Mad Mullah of Somaliland uprising, the Boer War and World War I. He devised an alternative scoring system born “out of frustration of being unable to reach some of the long par fours in regulation figures when harsh westerly winds made a nonsense of the traditional bogey [par] system of scoring.”

The objective of the Stableford scoring system is to accumulate the most points over 18 holes of golf. Good scores on individual holes are rewarded with points that reflect the difference between the net score of the golfer against par: 1 point for bogey, 2 for par, and so on.

When and where the system was developed is unclear. One claim is that Stableford devised a prototype system for use in September 1898 while a member at the Glamorganshire Golf Club, in Penarth, Wales. It was reported¹ that: “Each competitor plays against bogey level. If the hole is lost by one stroke only, the player scores one; if it is halved, the player scores two; if it is won by one stroke, the player scores three; and if by two strokes, the player scores four. To the score thus made, one third of the player’s medal handicap is added.”

However, the Wallasey Golf Club, in Cheshire, claims that Stableford invented the system while a member there



in 1931. The club’s first competition under Stableford Rules was played on May 16, 1932.

Modifications followed. The handicap adjustment was increased to $\frac{7}{8}$ from $\frac{1}{3}$ so that the scoring system would not grant any golfer more than 1 stroke per hole, or 18 per round (at the time, the maximum golfer’s handicap was 21). Later, however, players added their full handicap to the

Stableford points scored to get their total points. Clearly this gave high handicappers a distinct advantage. If, for instance, the weather was so bad that no one scored any points, the highest handicapper in the field would win. Moreover, when the handicap was simply added to the total score, the player with the higher handicap had the psychological advantage of starting off with a large lead on the first tee. The system was subsequently modified so that strokes were taken at the relevant holes.

One might argue that the Stableford scoring system does not offer enough reward for good play. After all, a birdie is worth only a little more than a par, and an eagle a little more than a birdie. Surely, these golf achievements are associated with far greater joy and satisfaction than the point differential would suggest. Similarly, the negative feeling generated by a very large score (with 0 Stableford points) is undoubtedly much greater than the acceptable bogey (with 1 Stableford point). It is not entirely clear what Stableford had in mind when he devised this scoring system — equity or quality of life (QOL). In this study, we explore the Stableford scoring system as it pertains to golf-related quality of life (GRQOL).

Methods

Actual scores from golf matches played May 15, 2002, at The Oaks Golf and Country Club, in Komoka, Ont., were used in this analysis. Eight golfers playing in 2 matches participated. Handicaps ranged from 3 to 18, with a mean of 10. Matches were played according to the Stableford scoring system, with adjustment for handicap of points on relevant holes. The player with the highest number of points won.

Two distinct types of measures exist for the assessment of QOL outcomes: health status measures and preference- or utility-based measures.² Health status measures are exemplified by the generic health-related QOL instruments. They can be used to estimate the burden of different chronic diseases at a single point in time, monitor a population's health and predict future health outcomes. Preference- or utility-based measures assess the value or preference for a given state against a standard metric. Essentially, they provide a subjective global assessment of the person's current state or hypo-

thetical states. We used both types of measures in our study.

At the end of each hole each participant underwent a modified Mini-Mental State Examination (total possible score of 30) and filled out the Medical Outcomes Study 36-item Short Form (SF-36).³ We then asked the golfers to express, in a single value, their strength of preference for particular golf scores anchored to par. They were provided with a 7-point Likert scale that included the Stableford scoring system and outcomes ranging from albatross (3 under par on a hole) on the far left and triple bogey (3 over par on a hole) on the far right, with par in the centre (Fig. 1). They were asked to mark an X corresponding to their handicap-adjusted and raw golf scores (if different) in relation to par. Underneath the scale they were asked to estimate the golf utility of that outcome (Fig. 1), assuming that par was equivalent to a GRQOL of 1 unit (perfect golf well-being) and that scores under par were associated with greater golf utility (> 1) and those over par with less utility (< 1). The scale could theoretically range from 0 (lower bound) to infinity (upper bound). Players scoring 4 over par or greater on a hole could extend the Likert diagram.

Finally, the golfers were asked to estimate the golf utility of other, even implausible, outcomes for that hole. Other variables recorded included handicaps and whether players were winning or losing the match at the time. Descriptive statistics and nonparametric methods were used for analysis. The relation between the Stableford scoring system and golf utility scores was assessed using regression analysis. A *p* value of less than 0.05 was assumed to be statistically significant.

Results

Health status

The results of the Mini-Mental State Examination did not change significantly through the 8-hour round. Orientation and attention tended to be worse among those losing matches, although this difference did not achieve statistical significance. In general, calculation skills were abysmal, although it was not entirely clear whether poor performance was feigned or not. The SF-36 assesses 8 health concepts: physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, general mental health, social functioning, energy/fatigue and gen-

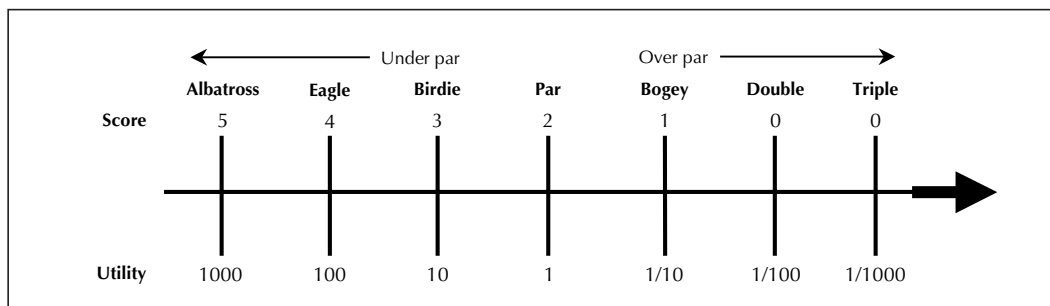


Fig. 1: Stableford golf-scoring system: handicap adjustment and quality-of-life measure in one scale. The upper portion of the scale indicates the points awarded for the handicap-adjusted score on each hole, ranging from 0 (for a double bogey or worse) to 5 (for an albatross). Underneath the scale is the estimated utility of the accomplishment (on a logarithmic scale) in relation to par, which is assumed to have a golf-related quality of life (GRQOL) value of 1 unit (perfect golf well-being). Note that, although the handicap-adjusted points do not go below 0, the (dis)utility scale continues indefinitely.

eral health perceptions. Scores for energy and fatigue were lower among those losing matches than among those who were winning ($p = 0.033$), but they did not correlate with scores on individual holes. Role limitations due to emotional problems were strongly correlated with poor outcomes on a hole ($p < 0.001$), particularly with double bogeys or worse.

Golf-related quality of life

Fig. 2 shows the relation between GRQOL (on a logarithmic scale) and the Stableford scoring system. The linear relation between the Stableford system and the logarithm of golf utility was strong and highly statistically significant ($r^2 = 0.85$, $p < 0.0001$). The utility for pars and birdies tended to be higher among the golfers who were losing their matches than among those who were winning and was most evident in the final 9 holes ($p = 0.09$). For pars, birdies and better, the results were similar whether the score on the hole was handicap-adjusted or the raw score, although the utility of handicap-adjusted scores was numerically slightly higher. Furthermore, this tendency was particularly strong among the golfers with higher handicaps. Such post-hoc subgroup analyses should be interpreted cautiously.

Interpretation

The main finding was that the Stableford scoring system is strongly associated with the notion of utility. The association is logarithmic (akin to the decibel or Richter scales), with a 1-point difference on the Stableford scale actually marking a 10-fold change in utility (Fig. 1). For example, if

a par is worth 1 golf utility unit (perfect golf well-being), a birdie awarding 1 point is worth 10 on the utility scale (10 times more enjoyment than getting par) and an eagle awarding 2 points is now perceived as 100 times more enjoyment. Although the utility scale has no upper bound, the likelihood of making an albatross (3 under par on a hole) was felt to be excessively remote; moreover, the scale appears to be no longer linear at that extreme. We would estimate that particular golf utility to be about the same as the odds of winning Lotto 6/49, or about 14 million to 1. Similarly, a hole-in-one, most likely to occur on a par 3, is likely to be valued far greater than a simple eagle and, according to 2 participants who had actually performed this feat, appears to be valued about the same as an albatross. The GRQOL system also works in the opposite direction with, for example, a double bogey (2 over par) being associated with 1/100 the utility of a par. Although the Stableford scoring system does not penalize the player for a triple, quadruple or worse bogey, even our small sample made it clear that the (dis)utility scale continues on, and may do so indefinitely (Fig. 1). For example, scoring an 8 on a par 3 is at least 10 000 times worse than getting a bogey. Every golfer knows that intuitively.

Although data-derived, the handicap-adjusted utilities among high handicappers tended to be slightly higher than the utilities based on raw scores. We suspect this is due to a combination of factors. First, high handicappers may not anchor their GRQOL at par, but more likely at bogey or double bogey. This would result in a shift of the entire utility system by 1 logarithmic unit — a par will feel as exciting as a birdie, and so on. Moreover, as the scale may no longer

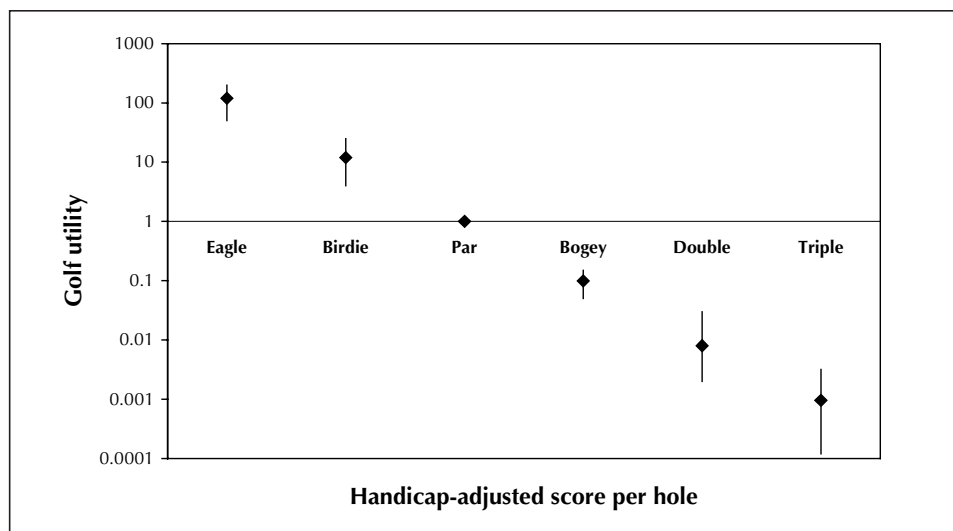


Fig. 2: Relation between golf utility (on a logarithmic scale) and handicap-adjusted score per hole. Par is assumed to have a golf utility of 1 unit. Results are shown as means (diamonds) and 95% confidence intervals (vertical lines). Not shown are the results for an albatross (3 under par) or quadruple bogey (4 over par) or worse. At these extremes the scale appears to become exponential. (Anyone who gets an albatross or quadruple bogey and is interested in calculating the corresponding golf utility can contact the corresponding author for the formulas [and to boast, in the case of an albatross]).

be linear at the extremes, a natural birdie will be associated with marked utility.

Second, handicap adjusting gives the high handicapper a greater chance of converting a bogey to something positive and a par to something really impressive. Not only is there a chance of halving or winning the hole, there is the psychological advantage of matching points with an inferior score.

Finally, high handicappers know in advance that there is a good chance that they will experience an automatic 10-fold improvement in utility for every hole on which they receive a handicap adjustment. Over the course of a round, this could be a deciding factor. After all, golf is 90% mental and 10% mental.

In summary, we believe from our findings that Stableford devised his scoring system with GRQOL in mind. It is interesting to note that the first health-related QOL measures date back to the 1930s and 1940s with the New York Heart Association classification⁴ and Karnofsky's performance scale.⁵ Stableford may have been aware of their development and seen potential applications in medicine and elsewhere.

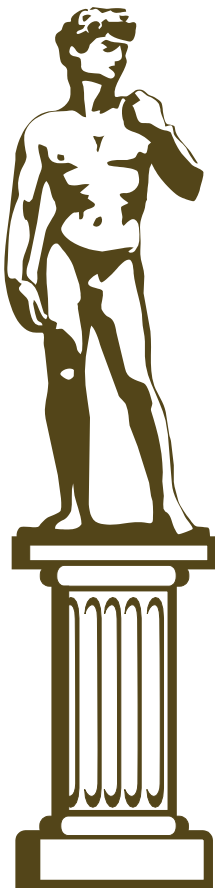
The authors are with the Departments of Medicine and of Epidemiology and Biostatistics, University of Western Ontario, London, Ont., and The Oaks Golf and Country Club, Komoka, Ont.

Competing interests: None of the authors is a professional golfer (or even a good golfer for that matter), and none owns shares in a golf club or company that manufactures golf equipment.

References

1. *South Wales Daily News* 1898 Sep 30.
2. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993;118:622-9.
3. Ware JJ, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
4. Criteria Committee, New York Heart Association. *Nomenclature and criteria for diagnosis of diseases of the heart and great vessels*. Boston: Little Brown and Company; 1939.
5. Karnofsky DA, Abelmann WH, Craver LF, Burchenal JH. The use of nitrogen mustards in the palliative treatment of carcinoma. *Cancer* 1948;1:634-56.

Correspondence to: Dr. David Massel, London Health Sciences Centre, Victoria Campus, Rm. 205 Colborne Building, 375 South St., London ON N6A 4G5; fax 519 667-6687; dmassel@lhsc.on.ca



Ignobel (1): dressing for success

The IgNobel prizes have honoured a diverse range of ground-breaking scientific studies. The awards, presented annually for scientific research that "cannot or should not be reproduced," are sponsored by the science humour magazine *Annals of Improbable Research* (see www.improbable.com). They are handed out by genuine Nobel laureates during a tongue-in-cheek ceremony held every October.

That "ground-breaking research" trend continued again this year, when one of the prizes went to a study of the asymmetry of the testicles found in Greek sculpture. The ancient Greeks showed a keen interest in realistic depictions of the nude male form, but it turns out that their sculptures show left testicles to be lower and larger than right ones. Evidence shows that the right testicle is usually higher and also larger than the left one, at least in right-handers.

Chris McManus of University College, London, received the 2002 IgNobel Prize for Medicine for a letter he had published

in *Nature* in 1976. McManus, a medical student at the time, cited research showing that the right testicle tends to be higher in right-handed subjects, whereas the converse is true in left-handers.

A holiday in Italy allowed McManus to examine 107 antique sculptures. "I like going to art galleries, and it is the work of a moment to look at each [sculpture]. And in some ways, it made each gallery more interesting — would they have any more sculptures to add to the collection?" explains McManus.

He discovered that in most sculptures the right testicle was higher, but the left (lower) testicle was larger. McManus doesn't know why ancient sculptors were dressing to the left, as it were, but speculates that it may be symbolic. The Greeks believed that the right side was associated with the male and the left side with the female — male children were thought to come from the right (higher) testis and female children from the left. — *Carolyn Brown, Ottawa, Ont.*