Evaluating relative performances in disabled sports competitions

DAVID F. PERCY†

Centre for Operational Research and Applied Statistics, Salford Business School,
University of Salford, Greater Manchester M5 4WT, UK

AND

D. BRUCE WARNER

ABB Sécheron, 4–6 rue des Sablières, Meyrin, CH 1211, Geneva, Switzerland

Handicapping systems play important roles in many sports. This paper focuses on disabled sports competitions, which primarily aim to encourage participation. Nevertheless, prizes are awarded for relative performances and should be allocated fairly. We review handicapping procedures specifically for disabled Alpine skiing and identify imperfections with this system, particularly how historical and new results are combined to determine scaling factors. Significant problems also arise due to different class sizes and variations in courses and conditions. We investigate the difficulties in establishing a suitable handicapping system and propose modifications to the existing procedures that improve their fairness. We develop simple mathematical models and statistical analyses that might be suitable, based on results data relating to several years of competition, and we evaluate the performance of our proposed method in comparison with that of the existing approach.

Keywords: factor system; handicap; Paralympics; shrinkage; sport.

1. Introduction

Disabled Alpine skiing has become very popular in recent years and international competitions are keenly contested. Its governing body is the International Paralympic Committee (IPC), which maintains an informative website (www.paralympic.org) that provided much material for this paper. The sport's main events are the Winter Paralympic Games held every 4 years since 1976, the World Championships held about every 4 years since 1982 and now biennial and the annual World Cup that started in 1999. Although such meetings partly aim to encourage inclusiveness and participation, they are highly competitive and it is important that awards for relative performances should be allocated fairly. This paper reviews how this is currently achieved and proposes simplifications and improvements to these procedures.

Prior to 1991, there were several (eventually 13) classes of disability for each sex and each Alpine skiing discipline. Each of these classes had its own winners, resulting in many victory celebrations and medal allocations. Furthermore, many of these classes contained few competitors and so provided insufficient competition. These features tended to devalue the status of winning and the scheme became difficult to manage. Consequently, a handicapping system was introduced and subsequently refined, resulting in what is now known as the Realistic Handicap Competition and Kreative Renn Ergebnis

---

†Email: d.f.percy@salford.ac.uk

Kontrolle (RHC–KREK) factor system, as described on the IPC website. Other sports that adopt handicapping systems include horse racing (Bolton & Chapman, 1986; Edelman, 2003), golf (Lewis, 2005), weightlifting (Cleather, 2006), bar quizzes (Percy & Scarf, 2008) and general North American sports (Szymanski, 2003). Some of the models and analyses that these papers consider are relevant here, though we introduce new concepts and specific modifications that are relevant to all disabled sports competitions and Alpine skiing in particular.

2. RHC–KREK factor system

Distinct competitions are held for disabled Alpine skiers in all combinations of three factors:

- two sexes (male, female);

- three categories (visually impaired, standing, sitting);

- four disciplines (slalom, giant slalom, super-G, downhill).

At the beginning of August 2007, there were 281 registered male skiers (29 visually impaired, 125 standing, 127 sitting) and 69 registered female skiers (14 visually impaired, 34 standing, 21 sitting). There is also a team event and a super-combined event, which combines two of the four disciplines noted above.

2.1   *Deciding the winners*

Each sex/category/discipline combination decides its own winners based on the race positions of its competitors, after making some adjustments to allow for different disability classes within the relevant category. For the 2007/08 season, there are 3, 12 and 5 classes for visually impaired, standing and sitting skiers, respectively.

Separately for each discipline but irrespective of sex, a 'base factor' $x_i \in (0, 1]$ is allocated to class i for $i = 1, 2, \ldots, m$, where m is the total number of disability classes across all categories (currently $3 + 12 + 5 = 20$), subject to $\max_i(x_i) = 1$. At any particular event and for any particular discipline, the 'racing time' r_{ij} of competitor j within class i is then multiplied by the corresponding base factor to produce a 'scoring time'

$$s_{ij} = x_i r_{ij} \tag{1}$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n_i$, where n_i is the total number of skiers (male and female) in class i. All the scoring times within each sex/category/discipline combination are then compared to decide the winners for that combination.

For any specified event and discipline, we calculate the 'current race factor' for disability class i as

$$y_i = \min_{ij}(r_{ij}) / \min_j(r_{ij}) \tag{2}$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n_i$, corresponding to the fastest racing time divided by the fastest racing time of class i. The averages of these current race factors based on all international competition results over the period 1976–1991 were used as 'basis factors'. These basis factors have subsequently been updated sequentially to provide the base factors $x_1, x_2, \ldots, x_m$ in (1), which are used to adjust the racing times.
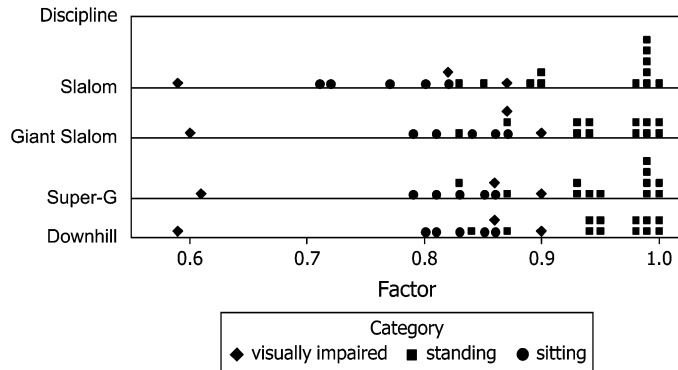
FIG. 1. Alpine skiing disability class factors for 2007/08 season.

For illustration, the dot plot in Fig. 1 shows the disabled Alpine skiing factors corresponding to the 2007/08 season. In all disciplines except the slalom, the maximum factor of 1 relates to standing skiers with one upper limb disability. For the slalom, it relates to standing skiers with a single above-knee amputation. For all four disciplines, the smallest factor is about 0.6 and relates to totally blind skiers who race with a guide. To demonstrate the RHC–KREK system in action, consider the 2008 IPC World Cup finals in Japan. In particular, the female visually impaired giant slalom competition had two competitors who finished the two runs fairly, with total racing times (minutes:seconds) of 2:07.21 and 2:22.85. Both skiers are in class B3 with base factor 0.9007885, leading to corresponding scoring times of 1:54.59 and 2:08.68 for these two skiers.

2.2 *General remarks*

Firstly, racing times and scoring times are usually rounded to the nearest hundredth of a second, whereas base factors are usually calculated to seven decimal places. In order that scoring times (seconds) are accurate to two decimal places, consider the propagation of errors in specifying base factors from (1): $(s_{ij} + \varepsilon_{ij}) = (x_i + \delta_i)r_{ij} \Rightarrow \varepsilon_{ij} = \delta_i r_{ij}$. To achieve the required accuracy, we set $|\varepsilon_{ij}| < 0.005 \Rightarrow |\delta_i| < 1/(200 r_{ij})$. As the racing times of finishing competitors in all sex/category/discipline combinations rarely exceed 5 min (300 s), a reasonable upper bound for $|\delta_i|$ is $1/(200 \times 300) = 0.000016$. Consequently, five decimal places are ample for base factors in order that scoring times are calculated correctly to two decimal places.

Secondly, the process of division into sexes and categories as described at the beginning of Section 2 does not treat all skiers fairly. Currently, e.g. a visually impaired female skier has to beat 13 others to win a competition, whereas a sitting male skier has to beat 126 others to win. These anomalies could be resolved by ensuring that all competitions contain similar numbers of skiers. Presuming that male and female athletes require separate competitions, one solution would be to allocate them different numbers of categories to reflect their relative participation rates. For this season, this would suggest having $281 \div 69 \approx 4$ times as many categories for men as for women. To implement this change partially, one could pool the visually impaired female and sitting female categories and split the standing male and sitting male categories into two new categories each for competition purposes. This would result in seven competitions comprising five male and two female categories, rather than six competitions comprising three male and three female categories.

2.3 *Updating the base factors*

The IPC Alpine Skiing Committee adjusts the base factors each spring, based on the racing results of World Cups, World Championships and Paralympic Games. It also reserves the right to make corrections and adjustments during the running season in case of extraordinary circumstances. The scatter plots in Fig. 2 illustrate how the factors changed from the 2006/07 season to the 2007/08 season. In all four disciplines, the fastest class is unchanged. Interestingly, almost all factors for the slalom and Super-G disciplines have increased slightly, suggesting that there has been a general reduction in differences among the classes for these events.

For any particular event that is being used to update the base factors of a specified discipline, the current race factors y_i are calculated from (2). A complicated flow chart that avoids overreacting to extreme race times decides the annual factor adjustments for each pair (x_i, y_i) of base factor and current race factor, for class $i = 1, 2, \ldots, m$. This algorithm can be simplified and expressed more concisely in the algebraic formulation

$$
x_i' = \begin{cases} x_i + (y_i - x_i) \min\left\{\frac{1}{10}, 10(y_i - x_i)\right\}, & 0 < y_i - x_i < 0.05, \\ x_i - (x_i - y_i) \min\left\{\frac{7}{200}, 10(x_i - y_i)\right\}, & 0 < x_i - y_i < 0.10, \\ x_i, & \text{otherwise}, \end{cases} \tag{3}
$$

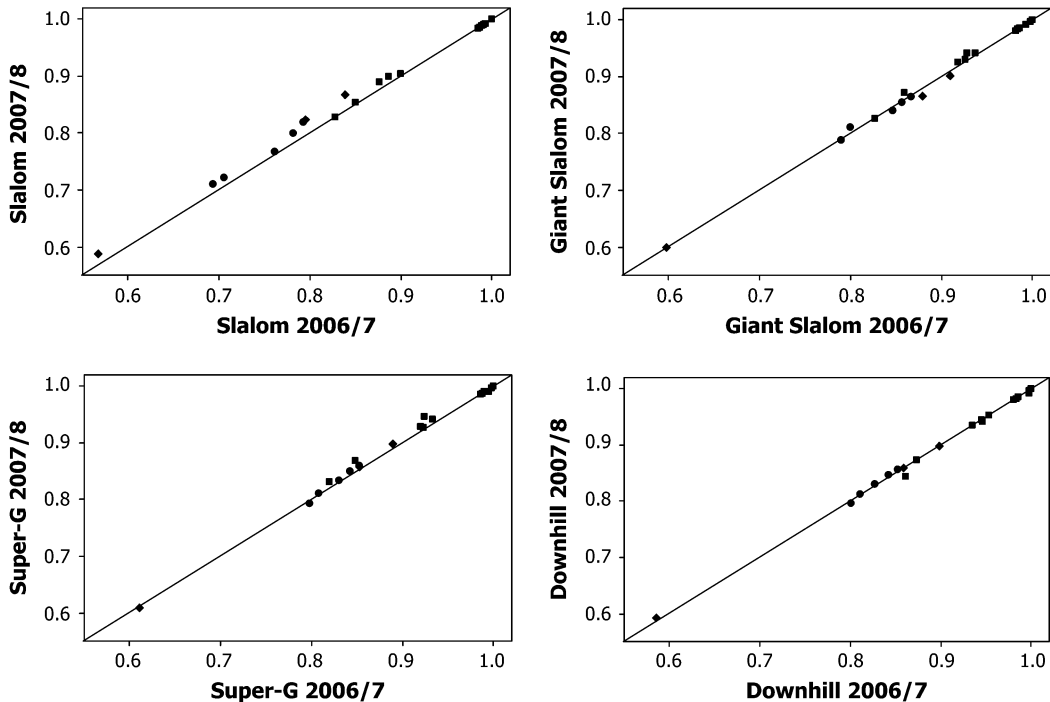where x_i' represents the updated base factor for class i.

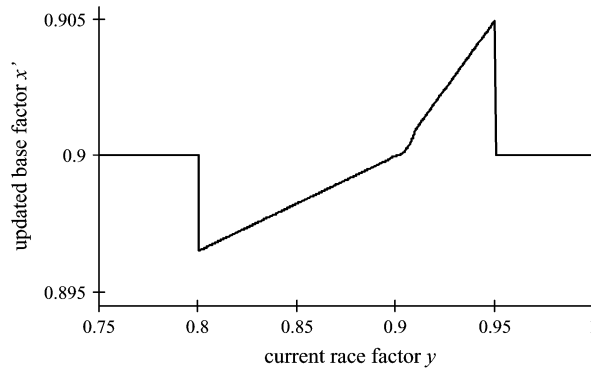FIG. 2. Alpine skiing class factors 2006/07 and 2007/08 for each of the four disciplines.

FIG. 3. Plot of the RHC–KREK updating function in (3) for a base factor of $x_i = 0.9$.

For example, suppose that in a particular race and for a particular class i, the base factor is $x_i = 0.9$. If the current race factor is $y_i = 0.94$, then $y_i - x_i = 0.04$ and the first condition in (3) is satisfied so we calculate an updated base factor as $x_i' = 0.9 + 0.04 \times (1/10) = 0.904$. If the current race factor is $y_i = 0.86$ instead, then $x_i - y_i = 0.04$ and the second condition in (3) is satisfied so we calculate an updated base factor as $x_i' = 0.9 - 0.04 \times (7/200) = 0.8986$. If the current race factor is $y_i \notin (0.8, 0.95)$ instead, then the third condition in (3) is satisfied and the base factor remains unchanged.

Figure 3 displays a graph of the updating algorithm described above for an arbitrary base factor of $x_i = 0.9$. The horizontal axis represents the current race factor y_i and the vertical axis represents the adjusted base factor x_i' calculated by adjusting the base factor x_i according to this algorithm. By necessity, the function passes through the point (x_i, x_i) corresponding to $(0.9, 0.9)$ here. It is immediately clear that adjustments to the base factor appear to be small, which might be a desirable property though these adjustments can lag behind technological advances. This implies that current race results do not have much impact upon the base factors. Moreover, the action interval is biased towards reduction of base factors and the general shape of this function within the action interval is convex.

If the class with base factor 1 enters a competition but does not achieve the fastest racing time, no class will have an updated base factor of 1. In this case, a constant additive adjustment known as the 'zero value' is applied to the updated base factors of all classes in the current competition in order to satisfy $\max_i (x_i') = 1$. Algebraically, the revised updated base factors become

$$x_i'' = x_i' + \left\{ 1 - \max_i (x_i') \right\} \tag{4}$$

if class i is in the current competition, and $x_i'' = x_i' = x_i$ otherwise, for $i = 1, 2, \ldots, m$. The RHC–KREK system has been used for 16 years with considerable success and we agree with the general principle that involves multiplicative scaling of racing times in order to compare athletes of varying physical capacities. However, there are weaknesses with this system, which we now address in order to propose modified handicapping procedures that resolve these issues.

3. Simplifications to the RHC–KREK factor system

As is clear from the preceding formulae, the algorithm for updating the base factors is fairly complicated, so we propose simpler alternatives in this section.

3.1 *Procedural matters*

Firstly, consider the adjustment of (4), applied when a new best class arises in competition. As the whole handicapping system is based on multiplicative factors rather than additive differences, we prefer to replace this adjustment with a normalization scaling

$$x_i'' = x_i' / \left\{ \max_i(x_i') \right\} \tag{5}$$

instead, for $i = 1, 2, \ldots, m$. For improved consistency, we also prefer to apply this revision to all classes rather than just those in the current competition.

There is a need for greater transparency regarding three other issues. One of these is the frequency of updating: it appears that adjustments are made annually based on each of the major races during the year. In this case, it might be more appropriate to update the base factors immediately after each of the major races instead. Another issue relates to a lack of clarity about the extraordinary circumstances when the IPC Alpine Skiing Committee might correct or adjust factors during a season, specifically the frequency and size of such amendments. Finally, it is unclear how the highly influential basis factors are calculated for new disability classes, for which historical performance data are unavailable.

The updating functions in (3) comprise piecewise linear and quadratic polynomials and are neither continuous nor smooth. Although this formulation is quite acceptable, continuous and smooth updating functions would involve less arbitrariness and ambiguity. One clear feature that Fig. 3 demonstrates is how the quadratic elements of adjustment that apply when $y_i \approx x_i$ are barely noticeable. These quadratic elements remain negligible for base factors other than $x_i = 0.9$, which was used to generate Fig. 3. Consequently, the algorithm could be simplified substantially by removing this unnecessary complexity and relying only on linear functions, which correspond to weighted averages of the base factor and current race factor.

3.2 *Continuous updating functions*

Strangely, the handicap updating procedures have discontinuities at the action limits in (3), outside of which no adjustment is made. On studying Fig. 3, a natural solution might level the updated base factor at the corresponding minimum or maximum value for current race factors outside the action interval. Such continuous handicapping procedures would involve less arbitrariness and ambiguity. However, the existing structure avoids the situation whereby truly outstanding performances penalize a whole class for the next season. It also means that exceptionally bad times are not used for adjusting the base factor, perhaps if the winning skier in a small class falls and registers a large time.

Very poor times might also arise if athletes in a small class deliberately perform badly in order to achieve a favourable base factor for the next season; such practice occurs in horse and greyhound racing. Nevertheless, extreme performances might be genuine and we deem it more appropriate to limit their impact upon the adjustments rather than to ignore them completely. A good example of a popular existing system that uses such a linear-plus-saturation algorithm for adjustment is that used by the International Rugby Board to rate national rugby teams; see the website www.irb.com/EN/World+Rankings for details.

As a result of the comments in the previous two paragraphs, we are already able to propose a considerable simplification and possible improvement to the updating algorithm, whereby we replace (3)
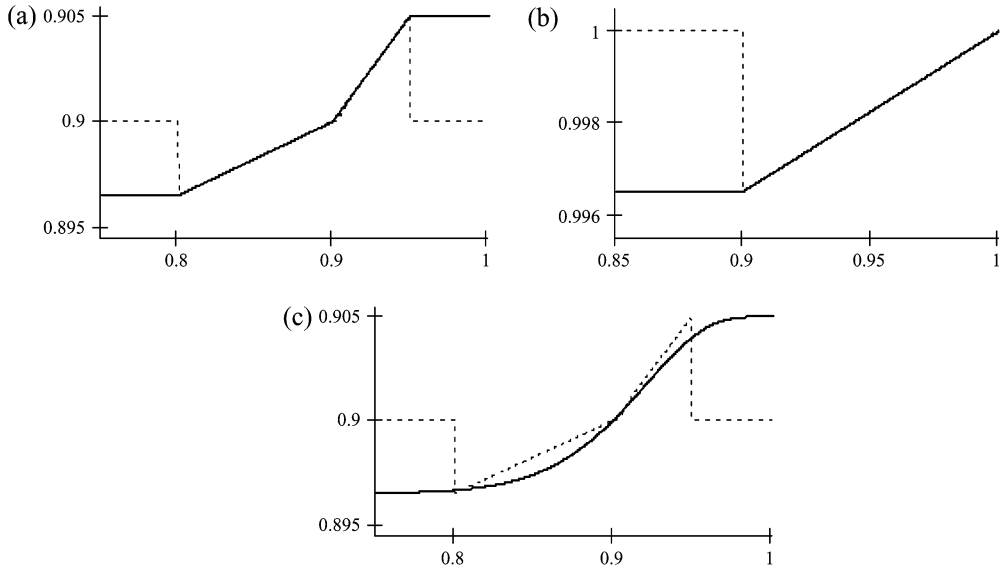
FIG. 4. Plots of the continuous updating function in (6) for a base factor of (a) $x_i = 0.9$ and (b) $x_i = 1$ and (c) the smooth updating function from (7) for a base factor of $x_i = 0.9$. Horizontal axes represent the current race factor y_i and vertical axes represent the updated base factor x_i'.

by the simpler form

$$x_i' = \begin{cases} \min\left(\frac{9x_i + y_i}{10}, x_i + \frac{1}{200}\right), & y_i > x_i, \\ \max\left(\frac{193x_i + 7y_i}{200}, x_i - \frac{7}{2,000}\right), & y_i < x_i, \end{cases}$$

(6)

which is easy to compute without the need for a flow chart. A graph of this modified updating algorithm is displayed in Fig. 4(a), superimposed over the original function for comparison.

The shapes of this continuous updating function are much the same for all values of the base factor away from the endpoints of $x_i = 0$ and $x_i = 1$. From Fig. 1, it is clear that the base factors are almost always greater than one half, so we only need to consider values of x_i close to 1 in order to assess how well our proposed continuous updating function performs for extreme base factors. Figure 4(b) displays a graph of this function corresponding to a base factor of $x_i = 1$, which identifies the best class in the category before the current race, superimposed over the original function for comparison as before. The quadratic effect of the original function is negligible again and the main difference is that our proposed continuous updating function also uses current race factors $y_i < 0.9$ to adjust the base factor $x_i = 1$.

3.3 *Smooth updating functions*

We next observe that the action limits in (3) and (6) result in updating functions that are not smooth, as there are discontinuities in their derivatives. Smoothness is a desirable property that would remove some arbitrariness and improve transparency of the factor system. There are infinity smooth functions that satisfy the condition of no change if $y_i = x_i$ with predetermined lower and upper bounds to be

achieved at $y_i = 0$ and $y_i = 1$, respectively. Perhaps, the simplest form is a linear transformation

$$x_i' = a\beta(y_i) + b \tag{7}$$

of the incomplete beta function

$$\beta(y) = \int_0^y z^{c-1}(1 - z)^{d-1}\,\mathrm{d}z, \tag{8}$$

which we evaluate using a simple numerical algorithm (see, e.g. Abramowitz & Stegun, 1965). Figure 4(c) displays such a smooth updating function for a base factor of $x_i = 0.9$, superimposed over the original updating function for comparison. Although the algorithm in (7) is easy to express algebraically and has nice mathematical properties, it is difficult to calculate in practice because of the need for a computer to evaluate the model parameters. There is undoubtedly wider acceptance of an understandable system rather than an unduly complicated one, as evidenced by FIFA's (Fédération Internationale de Football Association) recent decision to simplify the football world team rating system following widespread complaints that it was too complicated. For this reason, we do not pursue the idea of a smooth updating function here.

4. Improvements to the RHC–KREK factor system

Other problems arise with the factor system because of three particular shortcomings as follows, though there are significant interactions between them: (a) substantial variations in disability class sizes, which vary from 1 to about 25; (b) reliance upon only best racing times for deciding results and adjusting base factors and (c) disregard of racing conditions, including type of snow, course set, length of course and local weather. We now elaborate on the consequences of these shortcomings.

4.1 Consequences of shortcomings

A skier in a small class with a generous base factor has a good chance of winning the competition, whereas a skier in a large class must beat many others on equal terms for any chance of winning. Similarly, a skier who trains hard and improves might expect to win a lot of races during a season as the base factors are only updated annually, but such aspirations are far more likely to be realized in small classes than in large classes. Large classes are also disadvantaged as a consequence of using best racing times to set the base factors because extreme times arise more often in large classes and so average skiers have to perform well above average to win a competition.

A mathematical explanation is that the distance between a population mean and the expected value of the smallest order statistic based on a random sample from the population increases as the sample size increases. This situation is amplified in disabled swimming, whereby the best racing times are adjusted using all-time world records to evaluate the scoring times within each class. Similarly, any class that ever included a supreme individual athlete will always suffer through underestimation of the true handicap for that disability. Moreover, if the best skier in a large class has a bad run, the next best likely has a good run leading to no adjustment of the base factor, whereas the next best skier in a small class likely has a poor racing time that leads to adjustment of the base factor.

It would be interesting to analyse the correlation between disability class size and competition success, though this information is not readily available. The most obvious solution would be to pool classes so that the numbers of competitors are roughly equal, but this is impractical. We propose that base factors determined from means might be more robust than base factors determined from minima, though

we acknowledge that this tends to overcompensate and disadvantage small classes. Racing conditions can also have significantly different effects on different classes. If they favour a specific disability, perhaps due to the equipment used, then skiers in the relevant classes effectively have generous base factors and a higher chance of winning medals.

4.2 *Possible solutions*

For exploratory purposes, it is reasonable to make the simplifying assumption that in any particular competition, the independent racing times r_{ij} of competitors within disability class i are normally distributed with unknown mean μ_i and common standard deviation σ. Linthorne (2007) made a similar assumption in relation to athletics performances. In this case, the minimum and mean of all n_i racing times are defined by

$$\check{r}_{i\bullet} = \min_j(r_{ij}) \tag{9}$$

and

$$\bar{r}_{i\bullet} = \frac{1}{n_i} \sum_{j=1}^{n_i} r_{ij}, \tag{10}$$

respectively. Elementary probability theory tells us that the standard deviation of the mean of all n_i racing times is given by

$$\mathrm{SD}(\bar{r}_{i\bullet}) = \frac{\sigma}{\sqrt{n_i}}. \tag{11}$$

However, the standard deviation of the smallest of the n_i racing times is not easily expressible in closed form and is given by

$$\mathrm{SD}(\check{r}_{i\bullet}) = \sqrt{E(\check{r}_{i\bullet}^2) - \{E(\check{r}_{i\bullet}^1)\}^2}, \tag{12}$$

where, using theoretical results for order statistics,

$$E(\check{r}_{i\bullet}^k) = n_i \int_{-\infty}^{\infty} r^k f(r)\{1 - F(r)\}^{n_i-1}\, \mathrm{d}r \tag{13}$$

for $k = 1, 2$ in terms of the probability density function $f(r)$ and cumulative distribution function $F(r)$ for a normal distribution with mean μ_i and standard deviation σ (see, e.g. Stuart & Ord, 1994).

Figure 5 illustrates how the standard deviation for the mean racing time, in (11), is less than that for the minimum racing time, in (12), for all class sizes n_i that might be of interest, taking $\sigma = 1$ without loss of generality. The graph also illustrates how the standard deviations are larger for small classes than for large classes, whether we use minimum or mean racing times. It strongly suggests that the use of means rather than minima would be more robust, so leading to fewer anomalies. However, we might wish to avoid including the times of skiers who fall but still finish, or new racers who are still some way off the pace, when using mean racing times to update the base factors for each class.

One approach would be to use trimmed means, according to a specified criterion for inclusion. For example, we might average only the fastest 50% of racing times. Alternatively, we might include only racing times that are within 10% of the best racing time. Unfortunately, this implies that the trimmed
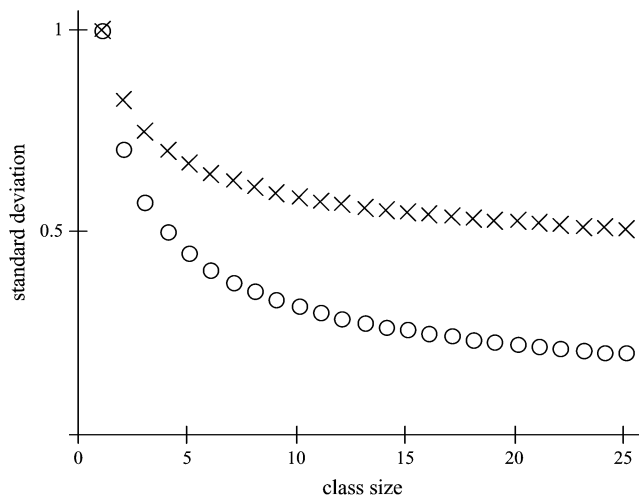
FIG. 5. Plots of standard deviations for minimum (x) and mean (o) racing times against class size.

mean is again effectively an order statistic of some kind, not a genuine mean, and the exact location of the cut-off for inclusion becomes crucial. A good competitor wants as many skiers in his or her class as possible, who are poorer than him or her but just good enough to be included in the trimmed mean. This could potentially generate opportunities for corruption and race fixing, though skiers race individually so relative performances are harder to predict during competition than for team sports (e.g. football, cricket) and simultaneous events (e.g. horse racing, cycling).

We now consider the problems that arise because the base factors take no account of race conditions, which can affect the various disability classes differently. One possible solution might be to collect historical data in order to develop a regression model that adjusts base factors according to the interactions between conditions and classes. However, such data are not currently available and this approach would require us to calculate factor adjustments after each race, which would remove some of the anticipation and excitement during racing, and might lead to substantial delays in compiling and verifying the results. To resolve this issue, we propose that adjustments should reflect how a skier does against the rest of his or her class, taking account of the number of skiers in that class. Although some of the anticipation and excitement of racing would still be lost, the *post hoc* computational effort would be minimal and racing conditions would automatically be taken into account.

In any particular event and for any specific discipline, suppose that disability class i has base factor x_i and racing times r_{ij} for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n_i$, where m is the number of disability classes across all categories and n_i is the total number of skiers in class i irrespective of sex. Suppose also that we calculate the base factor and adjustments using the mean racing times of (10) rather than the minima of (9). Better still, we use geometric means rather than arithmetic means because of the multiplicative nature of the scaling factors. To do this, we redefine

$$\bar{r}_{i\bullet} = \sqrt[n_i]{\prod_{i=1}^{n_i} r_{ij}} = \left(\prod_{i=1}^{n_i} r_{ij}\right)^{\frac{1}{n_i}} \tag{14}$$

for the following analysis.

Naively, we might use the mean racing time of the class to calculate the scoring times, rather than the base factor, so that we replace (1) by

$$s_{ij} = u_i r_{ij}, \tag{15}$$

where

$$u_i = \min_i(\bar{r}_{i\bullet})/\bar{r}_{i\bullet}. \tag{16}$$

This approach has the effect of scaling the racing times of all classes within each discipline so that their average scoring times are identical, thereby automatically allowing for differing disabilities and race conditions. It also has the amazing simplicity and advantage of not requiring basis factors, base factors and factor updating algorithms. If preferred, we could modify (16) to use trimmed means instead, as described earlier.

The principal disadvantage of this approach is its inability to distinguish between genuine race times and class disability effects. Suppose the skier in a class of size 1 records a particularly fast or slow time. Equation (15) cannot distinguish whether this is because of an exceptionally good or bad performance or because the disability for that class has a minor or major effect on performance: it merely attributes an average rank to this skier. If the skier recorded this time and the class included several other skiers, (15) would be able to compare their time with the class average, thereby increasing or decreasing this skier's chance of winning, as appropriate. This effect is analogous to a student's examination in which all members of a small class achieve high marks. Confounding exists, so we cannot distinguish between a bright class and an easy examination without external moderation. Fay & Herriot (1979) discuss the use of shrinkage estimates for the similar scenario of estimating regional income from census data. In order to apply shrinkage estimation to the handicapping formulae in (15) and (16), we modify the mean racing times of small classes by regressing them to the grand mean but do not modify the mean racing times of large classes.

Defining the grand (geometric) mean to be

$$\bar{r}_{\bullet\bullet} = \left(\prod_{i=1}^{m}\prod_{j=1}^{n_i} r_{ij}\right)^{1/\sum_{i=1}^{m} n_i} = \left(\prod_{i=1}^{m} \bar{r}_{i\bullet}^{n_i}\right)^{1/\sum_{i=1}^{m} n_i}, \tag{17}$$

we calculate a revised class mean $\bar{r}'_{i\bullet}$ such that

$$n_i \to \infty \Rightarrow \bar{r}'_{i\bullet} \to \bar{r}_{i\bullet}$$

and

$$n_i \to 0 \Rightarrow \bar{r}'_{i\bullet} \to \bar{r}_{\bullet\bullet}$$

for $i = 1, 2, \ldots, n$. Perhaps, the simplest function that achieves this is a weighted geometric average of the class mean and the grand mean, given by

$$\bar{r}'_{i\bullet} = (\bar{r}_{i\bullet}^{n_i} \bar{r}_{\bullet\bullet}^{k})^{1/(n_i+k)} \tag{18}$$

for a shrinkage parameter $k > 0$ to be specified. This leads to the revised formula for calculating scoring times

$$s_{ij} = v_i r_{ij}, \tag{19}$$

where

$$v_i = \min_i(\bar{r}'_{i\bullet})/\bar{r}'_{i\bullet}, \tag{20}$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n_i$. Although actual values are double counted because they are included in the class mean and the grand mean, this does not invalidate the analysis.

In passing, it is interesting to note from (18) that the revised class mean is simply the geometric average of the class mean and the grand mean, if the class contains $n_i = k$ racers for $k \in N$. Otherwise, the only obvious parameter value to use is $k = 1$ and we use this below, though empirical validation should be sought before implementing this in practice. There is a close analogy and justification for our proposed approach above with the James–Stein shrinkage methods that arise with empirical Bayes' procedures. These also have the effect of shrinking individual (arithmetic) means towards the grand mean but take account of differences between the variances, corresponding to our varying class sizes. Due to the extra complexity involved, we do not pursue this idea here, though we acknowledge that it might offer beneficial refinements to our analysis. Everson (2007) presented an excellent sporting review of this approach.

An alternative suggestion for overcoming the difficulty in distinguishing between performance and disability is to use an intermediate timing gate, perhaps 10 m from the start. This is because equipment and hence disability can have a big impact on the time to get up to speed: poles and the ability to skate with two skis are a big advantage. However, it is not clear how this information might be used to modify the class factors in (16) or simply to adjust the actual racing times to allow for the different disabilities.

A further alternative offers a compromise to avoid this problem, by using a standardized geometric mean of the base factor and mean racing time factor for calculating the scoring times. The adjustment formula then becomes

$$s_{ij} = w_i r_{ij}, \tag{21}$$

where

$$w_i = \sqrt{x_i u_i} / \max_i \sqrt{x_i u_i} = \sqrt{x_i u_i / \max_i(x_i u_i)}, \tag{22}$$

for $i = 1, 2, \ldots, m$, the number of classes in the discipline concerned. In (22), the base factor x_i is updated using (6) and the mean racing time factor u_i is given by (16). This compromise partly resolves the problem with base factors, related to the use of best times only, and the disregard of class size and racing conditions. It also partly resolves the problem with mean racing times factors, related to the confounding of effects. However, it does not completely resolve either issue and involves more calculations than either of the previous approaches.

Table 1 illustrates how the above methods would apply, using actual data from the male sitting Giant Slalom 2008 World Cup finals in Japan, with racing times reported in seconds. The first skier in class LW 11 attained the fastest actual racing time, but came second in the competition after the RHC–KREK base factor handicap adjustment. With only the third fastest actual racing time, the first skier in class 10-2 attained the fastest scoring time and so won the competition. Pleasingly, the same top three finishing positions are also identified using the mean racing time factors, shrinkage factors and combined factors, all without knowledge of the base factors, and there is a general consistency between our new results and the RHC–KREK results. Our preferred scaling adjustments are the shrinkage factors v_i of (20), as these allow for racing conditions and class sizes while avoiding the need for setting and updating historical base factors. Figure 6 illustrates the general overall agreement between the final rankings for the above example, when using the RHC–KREK base factors and our proposed shrinkage factors. Some variation is expected, as the former do not allow for specific racing conditions whereas the latter do.

TABLE 1 *Racing times r_{ij} and scoring times s_{ij} for the male sitting Giant Slalom* 2008 *World Cup final using base factors x_i (actual results), mean racing time factors u_i, shrinkage factors v_i and combined factors w_i*

| Class	r_{ij} (rank)	$s_{ij}(x_i)$ (rank)	$s_{ij}(u_i)$ (rank)	$s_{ij}(v_i)$ (rank)	$s_{ij}(w_i)$ (rank)
LW 10-1	152.40 (17)	120.20 (12)	114.77 (13)	127.04 (17)	117.53 (13)
LW 10-2	134.86 (3)	109.40 (1)	106.26 (1)	111.66 (1)	107.89 (1)
	149.75 (15)	121.48 (15)	118.00 (16)	123.98 (15)	119.80 (15)
LW 11	133.22 (1)	112.09 (2)	109.05 (2)	112.01 (2)	110.63 (2)
	133.37 (2)	112.22 (3)	109.18 (3)	112.14 (3)	110.76 (3)
	135.94 (5)	114.38 (4)	111.28 (6)	114.30 (6)	112.89 (5)
	136.87 (7)	115.16 (5)	112.04 (7)	115.08 (7)	113.66 (7)
	138.41 (8)	116.46 (7)	113.30 (8)	116.37 (8)	114.94 (8)
	139.55 (9)	117.42 (9)	114.24 (10)	117.33 (9)	115.89 (9)
	139.62 (10)	117.48 (10)	114.30 (11)	117.39 (10)	115.95 (10)
	140.59 (11)	118.29 (11)	115.08 (14)	118.20 (12)	116.75 (11)
	143.45 (13)	120.70 (13)	117.43 (15)	120.61 (13)	119.13 (14)
	153.63 (18)	129.26 (18)	125.76 (18)	129.16 (18)	127.58 (18)
LW 12-1	135.15 (4)	115.65 (6)	109.26 (4)	112.95 (4)	112.48 (4)
	136.12 (6)	116.48 (8)	110.04 (5)	113.76 (5)	113.28 (6)
	141.21 (12)	120.84 (14)	114.16 (9)	118.02 (11)	117.52 (12)
	150.90 (16)	129.13 (17)	121.99 (17)	126.12 (16)	125.59 (17)
LW 12-2	146.39 (14)	126.75 (16)	114.65 (12)	121.90 (14)	120.62 (16)
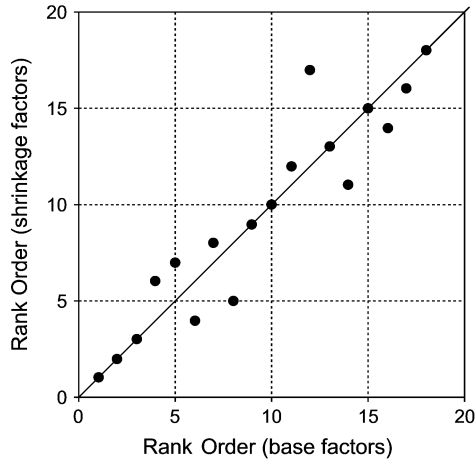
FIG. 6. Rank orders for male sitting Giant Slalom in 2008 World Cup finals, using base factors (actual results) and shrinkage factors of (20).

5. Conclusions

We began by briefly reviewing general handicapping systems in sport, before focussing on the RHC–KREK factor system for allocating handicaps to different classes in the highly competitive sport of disabled Alpine skiing. While acknowledging that this system has operated fairly well for 16 years and

that the IPC Alpine Skiing Committee has extensive knowledge and vast experience of setting handicaps for this sport, we identified several weaknesses that should be addressed.

Three simple improvements could be implemented to improve the consistency of the factor system: base factors could be recorded to five decimal places rather than seven, categories could be redefined so that they contain similar numbers of skiers and factor adjustments for new best classes could be multiplicative rather than additive. However, other important issues concern us here. In particular, the updating algorithm is unnecessarily complicated and has some undesirable properties including discontinuities, asymmetries, inconsistencies and arbitrariness. We suggested some ways of simplifying the updating algorithm to resolve these weaknesses and recommend the procedure defined by (6) for this purpose. This would instantly simplify and unify the RHC–KREK system for updating base factors, without impacting upon the race results in particular competitions.

We also noted that the current system for calculating scoring times does not allow for unequal class sizes, avoid the influence of extreme performances or adapt for different racing conditions, and we indicated how these shortcomings might be addressed. Our main proposal to achieve this is to adopt (19), which evaluates factors by shrinking geometric means, for calculating scoring times from racing times. This has the advantage of removing the need for any base factors, together with their associated initialization and updating procedures. Our preliminary investigations suggest that the shrinkage parameter k can reasonably be set to 1 for simplicity.

An unresolved issue concerns whether to use trimmed means rather than complete means for evaluating the factors. If trimmed means are used, a problem arises in determining which times to include. We discussed this matter briefly in Section 4.2, but it is a difficult point to deal with properly and so we leave this open problem for further research.

We advise that more extensive, comparative analyses on actual racing times be performed before adopting our proposal above, to identify any shortcomings that we might have overlooked inadvertently. It might also be worth investigating James–Stein shrinkage to justify or improve upon our recommendation. We further believe that it should be possible to allow some manual adjustments in extreme circumstances, through a representative and committee specially elected for this purpose. Finally, we hope that these general concepts, models and analyses might also prove to be relevant for other sports competitions.

References

ABRAMOWITZ, M. & STEGUN, I. A. (1965) *Handbook of Mathematical Functions*. New York: Dover Publications.

BOLTON, R. N. & CHAPMAN, R. G. (1986) Searching for positive returns at the track: a multinomial logit model for handicapping horse races. *Manage. Sci.*, **32**, 1040–1060.

CLEATHER, D. J. (2006) Adjusting powerlifting performances for differences in body mass. *J. Strength Cond. Res.*, **20**, 412–421.

EDELMAN, D. (2003) A competitive horse-race handicapping algorithm based on analysis of covariance. *The Economics of Gambling* (V. Williams ed.). London: Routledge, pp. 106–113.

EVERSON, P. (2007) A statistician reads the sports pages. *Chance*, **20**, 49–56.

FAY, R. E. & HERRIOT, R. A. (1979) Estimates of income for small places: an application of James-Stein procedures to census data. *J. Am. Stat. Assoc.*, **74**, 269–277.

LEWIS, A. J. (2005) Handicapping in group and extended golf competitions. *IMA J. Manage. Math.*, **16**, 151–160.

LINTHORNE, N. P. (2007) Influence of the size of a nation's population on performances in athletics. *Mathematics in Sport* (D. F. Percy, P. A. Scarf & C. L. Robinson eds), The Institute of Mathematics and its Applications, UK, pp. 105–110.

PERCY, D. F. & SCARF, P. A. (2008) On the development of decision rules for bar quiz handicapping. *J. Oper. Res. Soc.*, **59**, 1406–1414.

STUART, A. & ORD, J. K. (1994) *Kendall's Advanced Theory of Statistics, Volume 1: Distribution Theory, 6th edn*. London: Hodder Arnold.

SZYMANSKI, S. (2003) The economic design of sporting contests. *J. Econ. Lit.*, **41**, 1137–1187.