
A new approach for analyzing running and swimming data provides an interesting view of sex differences.

Assessing Time Trends in Sex Differences in Swimming & Running

Howard Wainer, Catherine Njue, and Samuel Palmer

Introduction

The Boston Marathon has been staged annually since 1897. Until the early 1960s only 200 to 300 runners competed, but the number of runners has increased steadily since then. Now race organizers impose stiff qualifying standards to limit the field. Even with such restrictions in effect, the Boston Marathon's starting field numbered over 38,000 in 1996. Before the 1970s few women ran for recreation, and amateur regulations on competitive racing barred them from distances longer than 4 km. All that has changed as shown by the increase in entrants in several for-women-only races that have drawn more than 8,000 starters. In Fig. 1 all of the winning times for the Boston Marathon are shown augmented by the best fitting linear trend lines.

Three natural questions that are suggested by such data are the following:

1. How far behind men's marathon performances are women?
2. How long before women catch up?
3. Why are women lagging?

Most would agree that the extrapolation implicit in Fig. 1 is unwarranted and does not contribute much toward understanding the issues that underlie these three questions. In this article we shall explore these issues more

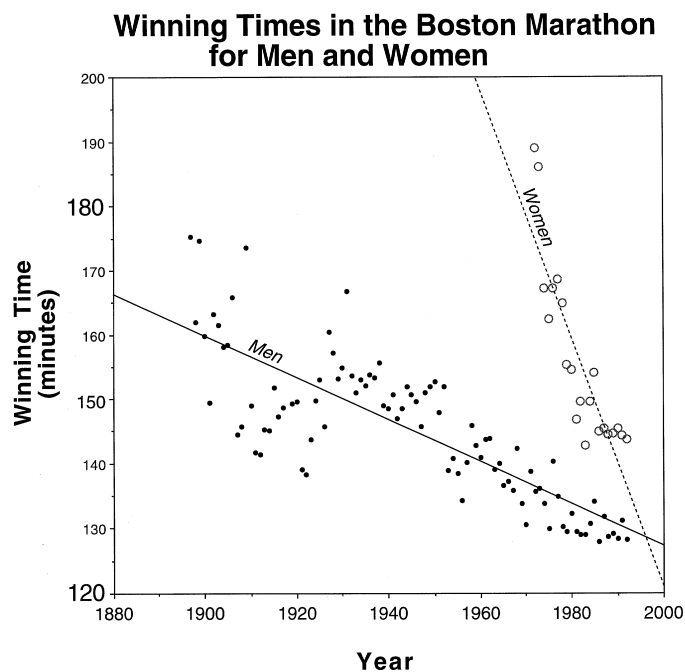


Figure 1. Winning times (in minutes) in the Boston Marathon since 1897 with linear trends indicated for both men and women.

carefully and suggest a measure that seems to provide help.

Over the past century women's participation in all competitive sports has increased markedly. At the same time the disparities in performance between men and women have shrunk substan-

tially. One cannot help but suspect that the former is a principal cause of the latter. To illustrate this phenomenon further, consider the changes in the world record for 100-m dash shown in Fig. 2 as mean speeds for the distance (in meters/second). We see

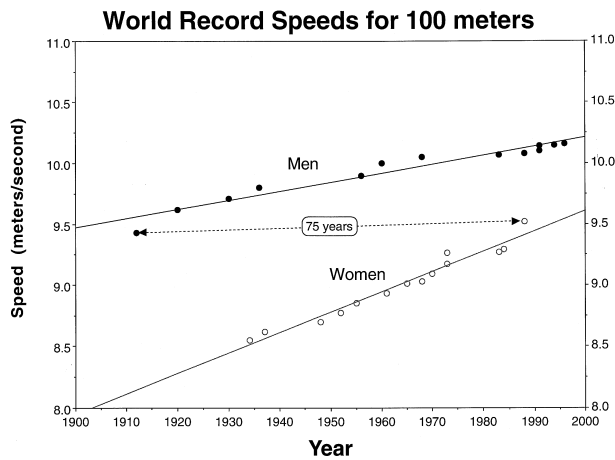


Figure 2. World record speeds for 100 meters for both men and women since 1910.

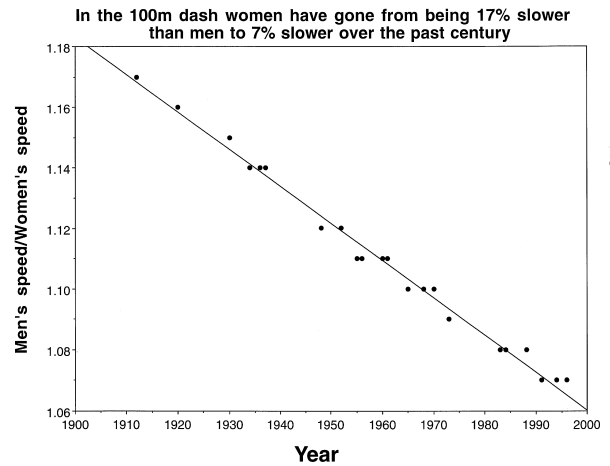


Figure 3. The ratio of women's world record speed in 100-meter dash to men's (derived from the linear fits in Fig. 2).

that men have increased their mean speed from less than 9.5 m/sec in 1910 to slightly over 10 m/sec by 1970. The women's record speed improved from about 8.5 m/sec in 1934 to almost 9.5 m/sec by 1990.

If we consider the ratio of these speeds, women to men, we find (in Fig. 3) that in 1934 the women's record was about 14% slower than the men's, but that this difference had shrunk to about 7% by 1990. Figure 3 is entirely derivative. It was obtained by fitting a linear function to each of the two time series in Fig. 2 and then dividing those two functions to yield the function shown as Fig. 3. Because a linear fit is rather good, this rough and ready approach is accurate so far as it goes, although it would be a capital mistake to extrapolate from it. In no way do we mean to imply that this ratio

will reach unity sometime in the middle of the next century. We recognize that the underlying trends are nonlinear and one possible explanation for the dramatic gains of women, relative to men, is that their records were taking place at a different portion of the curve. We shall return to this issue below, but for now merely note that the women's 100-m world record, set in 1984, is about the same as the men's record set 75 years earlier.

To gain momentum on what will be one of our major points, that women are improving at a faster rate in longer distances, consider the world records for the 800-m and 10,000-m runs shown in Figs. 4 and 5 respectively. These two plots are parallel in structure to Fig. 1 albeit with somewhat slower speeds and hence steeper slopes characterizing women's

improvement. Once again we can plot the ratio of men's to women's speeds, as we did in Fig. 3. This is shown for all three distances in Fig. 6.

Figure 6 shows several aspects of the phenomena clearly. Women world record holders run slower than their male counterparts, but over the past 30 years the difference between them has shrunk considerably. The longer the distance the more rapidly has the difference shrunk. Thus, we ought to expect that for really long races the slope of improvement should be steeper still.

What About Swimming?

Swimming seems somewhat different, at least superficially. The data shown in Fig. 7 are a parallel to those in Fig.

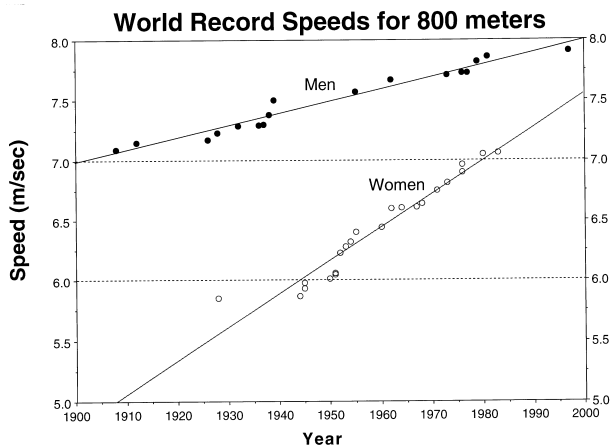


Figure 4. World record speeds for 800 meters for both men and women since 1910.

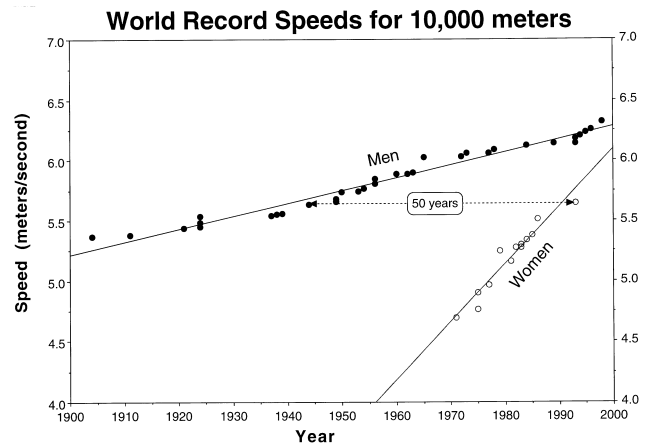


Figure 5. World record speeds for 10,000 meters for both men and women since 1910.

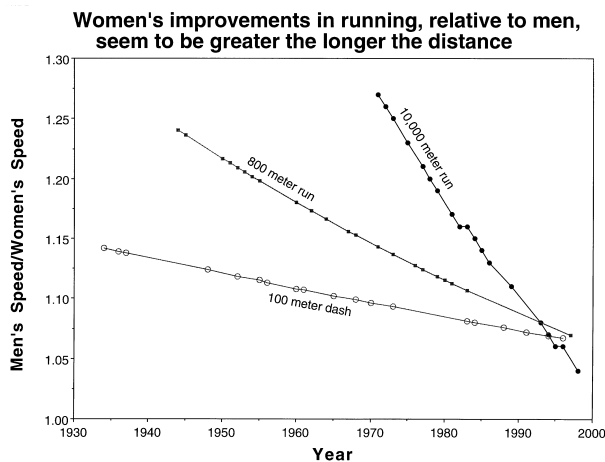


Figure 6. The ratio of women's world record speed in 100-meter, 800-meter and 10,000-meter runs to men's (derived from the linear fits in Figs. 2, 4, and 5).

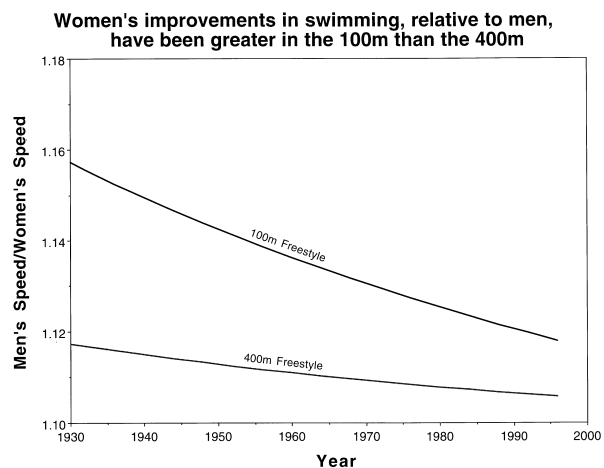


Figure 7. The ratio of women's world record speed in 100-meter and 400-meter freestyle swimming to men's record speeds.

6; there are similarities and differences. First we note that women's world record speeds for the 100-m and 400-m freestyle events has been approaching those of men, although somewhat faster for the 100-m event than for the 400-m event. Women's performances for the longer event, however, appear closer to men's performances.

Two Problems With This Approach

What can we infer from these results? Surely we can see that women are making substantial improvements in their athletic achievements, but can we project from these data what will be happening in the future. If we were to extrapolate from Figs. 6 and 7 we might say that women will be running and swimming faster than men within the next few decades. Most experts would be very surprised if this were to occur. Where have we been misled?

There are two problems. One is with the data that we have examined so far, and the second is with the method we have chosen to analyze them. Let us look at

each of these problems in turn and try to find a solution. First the method of analysis.

The trends over time of athletic record setting are nonlinear. When a sport (or event) is new, records improve very quickly, but as the sport matures the records change much more slowly. There are many causes for this, but the most obvious one is participation rate. When a sport is new the record holder might be the best among hundreds, whereas for a mature sport the world record holder might be the best among millions; as participation increases, the record improves apace.

We can see this effect clearly in data from the Boston Marathon, which has been run over essentially the same course for more than a century. In Fig. 8 we can see that the curve that describes the improvements in men's performance at the very beginning of the race, when participation among men was limited, is quite similar to the curve that describes the women's performances since 1972, when they were first allowed to compete. As women's participation increased so too did the speed necessary to win. How can we characterize this simply? We have duplicated the data points associated with the women's record speed and moved them back in time to where they roughly coincide with the men's records. This suggests a different approach to representing the difference between men's and women's performances.

Let us ask the following question: "How many years ago was the men's record 2% faster than the women's record is today?" This question has two pieces to it: "How many years ago" is a translation of the curve to the left; "2% faster" is a translation upward. The key assumption here is that the improvement in performance for both men and women follows the same curve, so to make accurate comparisons

Women's Performances in the Boston Marathon Trails Men's by 75 years

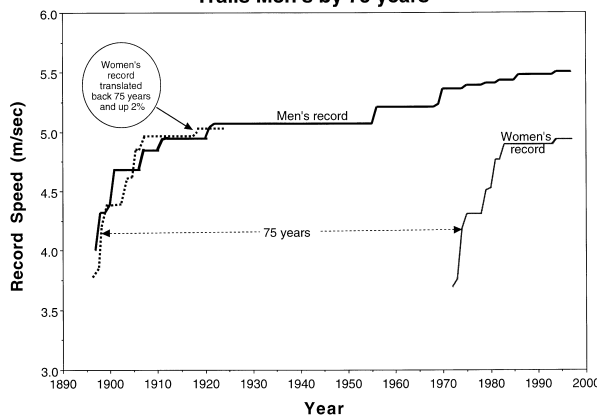


Figure 8. Winning speeds (in meters/second) in the Boston Marathon since 1897 with the women's curve translated left 75 years and upward 2%.

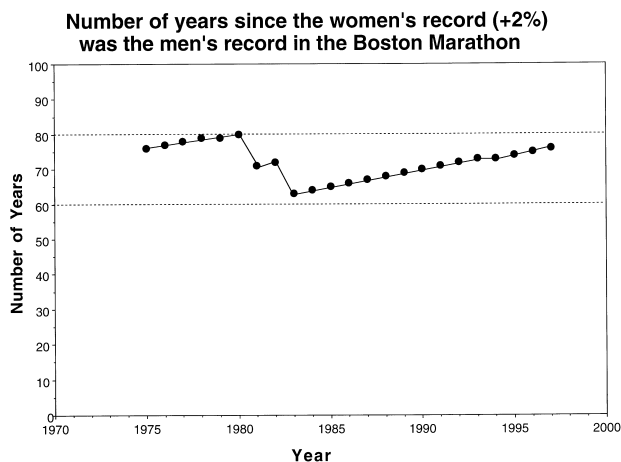


Figure 9. The number of years since the women's record in the Boston Marathon (+2%) was the men's record has remained roughly the same for the last 20 years.

we must translate one curve to overlay the other. The choice of 2% is somewhat arbitrary but provides a criterion for prediction. Cashmore (1999) estimated that this difference, which he ascribed to "the strength gap," is about 5%. We agree that this is a reasonable ballpark estimate but that there is some variation as a function of kind of exercise. We return to the issue of how to choose the size of this performance gap later.

When we plot this difference (in Fig. 9) we find that it is remarkably flat suggesting that between 60 and 80 years ago men were running the Boston Marathon about 2% faster than women and that the improvements that women have shown over the past 20 years were mirrored by improvements that men made over a half century earlier. This result more closely matches expert opinion in estimating when women will overtake men in the Boston Marathon.

The second problem is the statistical instability of inferences about the main body of a distribution that are made from an extreme value. Such values are always unstable and hence such inferences are bound to be equivocal. We could improve stability if we could get the entire distribution of world class performance. Sadly, such informa-

tion is not available. We note, however, that third-place times are also the median of the top five finishers. Thus, if we focus our attention on the performance of Olympic bronze medalists, we would have a measure that represents the median performance of the five best athletes in that Olympic year.

Some Results

In Fig. 10 the bronze medal times are shown for men and women for three track events (100 m, 800 m and 10,000 m) in all Olympics for which these events were run. Superimposed over each data series is the best fitting

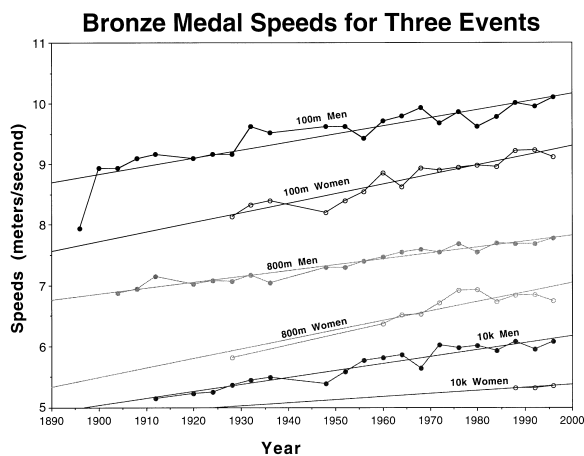


Figure 10. Bronze-medal speeds for three Olympic track events for both men and women, with linear fits.

straight line, which adequately represents the time trends seen in these events. Using the fitted values we can calculate the number of years ago that the women's bronze speed (+2%) was the men's bronze speed. This result is shown in Fig. 11. The results suggest quite a different inference than that we would have drawn from the results in Fig. 6 but are consistent with what we observed with the Boston Marathon data. Specifically, we see that women's performances in sprints (100 m) is closer to men's performances (less than 50 years behind) than for longer distances. We will not pay much attention to the 10,000 m results because the race has only been run three times, and hence we have too little information to draw reliable inferences about trends.

It is appropriate at this time in the narrative to describe briefly the analytic methodology used to assess time lags and why we chose 2% as the vertical difference in men's and women's performances. To illustrate both what we did and how it turned out, let us consider the Olympic bronze medal records in swimming 100 meters freestyle shown in Fig. 12.

The analytic problem is to translate the women's curve horizontally and verti-

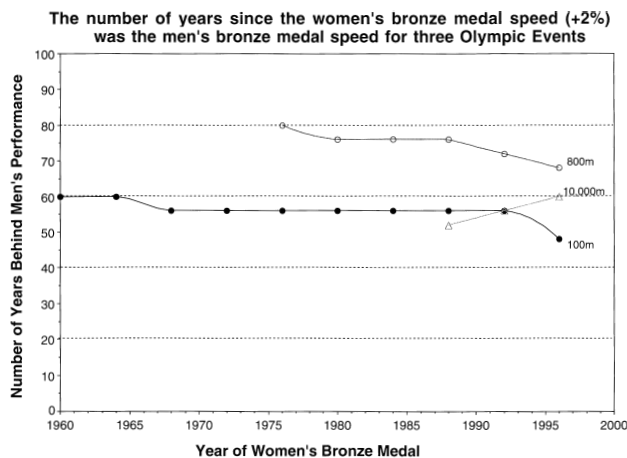


Figure 11. The number of years since the women's bronze-medal speed (+2%) was the men's bronze-medal speed in three track events.

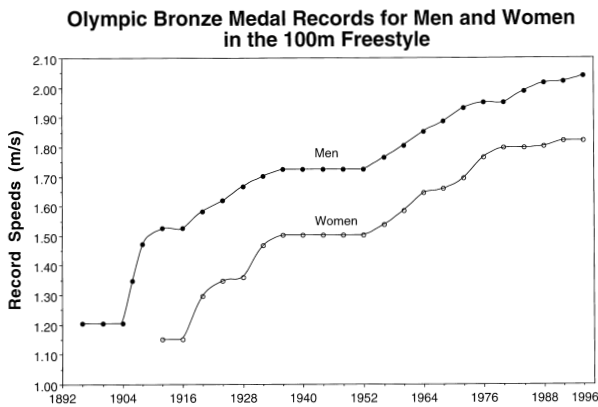


Figure 12. Olympic bronze-medal speeds for the 100-meter freestyle swimming event for both men and women.

cally until it coincides as closely as possible to the men's curve. We did a conditional minimization by sequentially first moving the women's curve to the left four years and then upward until the mean squared difference between the two curves was minimized; then we moved it back four more years and repeated the minimization. When this was done we found (summarized in Table 1) that the error was minimal when there was a lag of between 8 and 12 years and with women swimming about 8% slower than men.

This result is somewhat different from what we found for track using the same methodology, where it yielded a lag of about 75 years with a 2% difference. Data limitations as well as the rather flat response surface requires us to interpret these results broadly. Obviously the estimates of the two parameters of the curve translation will

be negatively correlated, so interpretations are muddled a bit. But it seems clear that the data suggest that, in swimming sprints, women seem to be limited to achieving speeds about 8% slower than men's and that women's improvement curve lags men's by about a decade. In Fig. 13 we show the trend in the number of years that women's speeds (+9%) have lagged men's in the 100-m freestyle. There is clearly a time trend suggesting substantial gains over the last few decades (as well as an anomalous lump surrounding World War II). We shall suggest one plausible explanation for the gains in the next section.

When we repeat the same analysis (see summary in Table 2) for the 400-m freestyle event, we find very similar results except that minimal error of fit seems to occur with only a four

year delay. Once again (see Figs. 14 and 15) we see a remarkably similar structure in the improvements in bronze medal times (Fig. 14) as well as a swooping reduction in the time lag between the sexes. One interpretation of this result is that women's training in distance swimming seems to have reached approximately the same level as men's.

Discussion

In this article we have examined sex differences in performance in track and swimming over time. We noticed that the men's and women's curves of improvement in performance were very similar, except that the women's curves were offset horizontally by a

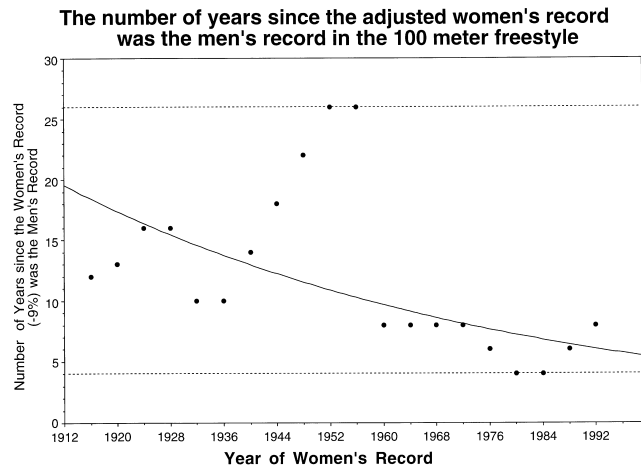


Figure 13. The number of years since the women's Olympic bronze-medal speed (+9%) was the men's Olympic bronze-medal speed in 100-meter freestyle swimming event.

Table 1 — Results of Optimization Curve Registration for 100-m Freestyle

Number of years back	Percent upward	Mean squared error ($\times 1,000$)
4	13%	5.47
6	11%	3.95
8	9%	2.86
12	7%	2.97
16	5%	4.59

Table 2 — Results of Optimization Curve Registration for 400-m Freestyle

Number of years back	Percent upward	Mean squared error ($\times 1,000$)
0	11%	2.16
4	9%	1.86
8	7%	2.11
12	5%	2.53
16	3%	3.22
18	0%	3.78
20	0%	4.49

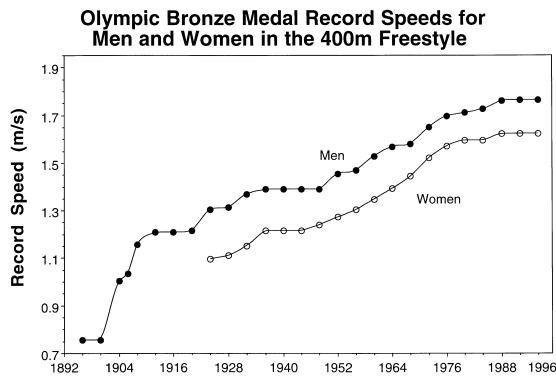


Figure 14. Olympic bronze-medal speeds for the 400-meter freestyle swimming event for both men and women.

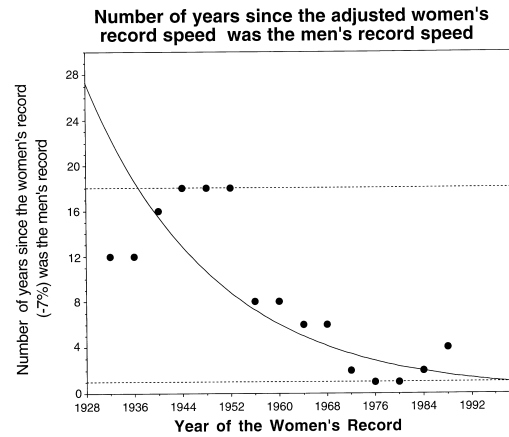


Figure 15. The number of years since the women's Olympic bronze-medal speed (+7%) was the men's Olympic bronze-medal speed in 400-meter freestyle swimming event.

varying number of years and vertically by some percentage. The vertical offset seems to vary between 2% (for the marathon) and 9% (the 100-m freestyle). The range of this offset melds well with the average 5% estimate that others have found (e.g., Cashmore 1999). This offset is usually interpreted as representing some sort of physiological strength difference. Such an interpretation is supported by our findings. It is generally agreed that the biggest sex differences in strength are in the upper body, so we should expect to see larger differences in the vertical offset in swimming than in running because the former uses upper body strength to a greater degree. Our analyses only partially support the emerging consensus that sex differences are small to nonexistent in physical endurance. We found smaller sex differences in the longer swimming event, but in running women seem to do relatively better in the sprints.

After making the vertical physiological adjustment, we used, as our principal measure of women's improvements, the number of years that have elapsed between when the adjusted women's record was the men's record. This measure provided some interesting insights. Most notable was the shift in the message seen in Fig. 1 to that seen in Fig. 8; women's performances in the marathon seem to lag men's by about 75 years, not the 5 or 6 that a naive linear extrapolation might suggest. The lag is not much greater in the sprints; Florence Griffith Joiner's world record for 100 m (10.54 seconds), though much better than any other competitor, is about the same Harold

Abraham's 1924 Olympic time.

In swimming the lag is much shorter, being measured in years not decades. Moreover it has been shrinking quickly. Why? Palmer (in press) suggested that changes in attitudes toward what women are physically capable of has given rise to vastly increased rigor of training. Women's training in swimming now is indistinguishable from men's in terms of both content and frequency. But, more important is participation. Palmer (in press) reports that "girls start swimming younger than ever before. There has been a dramatic increase in the enrollment in swim clubs and other swim teams to the point that now fully 60% of swimmers under age 17 are female. Secondly, more girls than ever before are sticking with the sport. Since Title IX was enacted in 1972, more women have been able to continue swimming through college, and not have to stop at the age of eighteen due to lack of opportunity."

We are well aware of the fragility of the analytic procedure of curve registration we used to translate one performance curve to another. It relies on the untested assumption that the size of the talent-pool within each sex is time shifted, but otherwise essentially equal. Moreover, the response surface is quite flat, and other solutions, with quite different interpretations may be almost as likely. We are currently doing a more rigorous sensitivity analysis that should help illuminate this issue. But nonetheless, we believe that the results obtained provide us with a much clearer view of sex differences in athletic performance. We anticipate

that others will augment both our data and our methodology to yield better answers yet.

[Howard Wainer's work on this project was partially supported by the Senior Scientist Award given to him by the trustees of the Educational Testing Service. Catherine Njue's work on this project was principally done while she was a summer pre-doctoral fellow at the Educational Testing Service. Samuel Palmer's work was done as part of an independent study project in statistics at Princeton High School. We would like to thank Phil Whitten of *Swimming World Magazine* for the bronze-medal swimming data, Cathy Durance of USA Swimming for the data on women's participation, and John Tukey.]

References and Further Reading

Bronze medal Olympic track times for men and women in 100 m, 800 m, and 10,000 m were obtained from *University of Exeter Centre for Innovation in Mathematics Teaching on the World Wide Web*: www.ex.ac.uk/cimt/data/.

Cashmore, E. (1999), "Women's Greatest Handicaps: Sex, Medicine, and Men," *British Journal of Sports Medicine*, 33, 76.

Palmer, S. (in press), "Sex Differences in Swimming Then and Now," *Swimming World*, 35.

Track World Record progression data for men and women in 100 m, 800 m and 10,000 m were obtained from *Mika Perkiomaki's Great Track and Field Statistics Page on the World Wide Web*: www.utu.fi/~mikkoski/ylinks.html.

Comment: Studying Trends in Sport Participation by Modeling Results of Elite-Level Athletic Performance

David E. Martin

Study of the results of athletic competition is useful for identifying the contemporary limitations of human performance (Hill 1924). Sport sociologists in particular will thus delight in reading Wainer, Njue, and Palmer's "Assessing Time Trends in Sex Differences in Swimming and Running." We all know that women's participation in sports has lagged behind that of men, and sociologists have offered many explanations: Cultural and religious influences are but two. Sport psychologists suggest that women as athletes are equally trainable in comparison to men, with similar competitiveness and enthusiasm. Thus, as the many barriers to women's participation in sports gradually disappear, the increasing size of the gene pool that is participating has substantially increased the rate of performance improvement. The challenge has been to identify and portray these changing dynamics over time.

Wainer, Njue, and Palmer provide a novel approach for quantifying such changing patterns, and illustrate their concepts using selected swimming and running events that have been contested since the modern Olympic Games began. Those interested in other sports will enjoy using this strategy to study other Olympic events contested annually over this past century that have had both male and female participation. Results for each Olympic Games are easily obtained (Wallechinsky 1996), and each sport publishes its own annual top performance sum-

maries. There are a few points of possible misinterpretation or confusion in Wainer et al.'s presentation, however — not in their study strategy but in their data collection and interpretation — that deserve mention.

First, although the Boston Marathon does in fact impose qualifying time standards that must be met by those wishing to run, the example used to indicate the number of such competitors — namely, the 35,868 finishers in 1996 — gives an inflated picture. That was the race's centenary, and a large number of race participants were permitted to run as a result of lottery selection without meeting the qualifying standard. A better indication of recent race participation at Boston is given by examining the number of finishers during those recent years in which participation was invitationally restricted — 1995 (8,258), 1997 (8,893), 1998 (10,289), and 1999 (11,274).

Second, although it is correct that the Boston Marathon has been run continuously since 1897 over *basically* the same route, extensive road construction has produced four different course distances. Figure 1 of Wainer et al. plots winning times from 1897 through 1999, and Fig. 8 plots pace in meters per second over the same period. For these plots to portray meaningful comparisons of performance, the finish times and race paces need to be normalized to a single appropriate distance. Logically, this ought to be the standard marathon distance used

presently (26 miles, 285 yards; 42,195 meters). Awareness of these course variations is not mentioned in the article. The various distances, and the years these distances were used (Martin and Gynn 1979) are as follows:

1897 through 1923: 39,742 meters; 24 miles 1,232 yards

1924 through 1926: 42,025 meters; 26 miles, 209 yards

1927 through 1952: 42,195 meters; 26 miles, 385 yards

1953 through 1956: 41,083 meters; 25 miles, 938 yards

1957 through 1999: 42,195 meters; 26 miles, 385 yards

Third, it is mentioned that "third place (Olympic) times are also the median of the top five finishers," and thus, we should "focus our attention on the performance of the Olympic bronze medallists" for developing "a measure that represents the median performance of the five best athletes in that Olympic year." Though it seems reasonable to assume that bronze-medal times are more representative of the entire distribution of world class performances than gold-medal times, the bronze-medal times are suspect as well. For one thing, some Olympic tracks (or swimming pools) are better suited to fast times than are others. Thus, in some Olympic years the best world class performances don't occur at the Olympics at all. The data problems identified here point out the intrinsic difficulties in studies of this type.

Finally, although it is quite interesting from a sport sociology viewpoint to compare the changing dynamics of *participation* by women and men, it is inappropriate to extend the comparison to *performance* dynamics. This was alluded to briefly in suggesting correctly that extrapolation of the data in Fig. 1 to predict that women will run faster than men by mid-2000 is nonsensical. Others have suggested that this will indeed occur (Whipp and Ward 1992), and have been duly criticized (Julin 1992). One must be careful when using such phrases as “when women will catch up to men” or “why women are lagging behind” lest they be misinterpreted to mean that women ought to be *able* to “catch up” to men in sport performance. There’s ample reason why they should not: They are two physiologically different populations, not two subsets of a single population, one of which is somehow “lagging behind.” A brief review of how the physiological differences contribute to the performance difference in both the strength- and endurance-oriented events makes this clear.

In the shorter sprint distances contested by both runners and swimmers, the performance differences between the sexes are affected by two independent variables. One is the larger skeletal muscle mass among men, caused by both human growth hormone and testosterone, which exist in higher concentration in men than in women. In addition, however, women are under a greater estrogenic influence than men, and this hormone promotes greater fat storage in tis-

ues. A champion male 100-meters-dash man may have 6% body fat, but his female counterpart will have 11%. This fat slows both athletes because it must be carried along as excess baggage, but the slowing effect in women is greater. Among swimmers, however, the increased body fat among women makes them more buoyant, giving them a performance advantage. This explains the smaller male/female performance difference for swimming records than for running records. Unfortunately, the variable extent of clandestine steroid hormone usage among both sexes in both running and swimming over the years has added a confounding element to meaningful analysis of record-quality data.

As the competition distance lengthens, in both swimmers and runners, although the muscle and fat differences remain, the endurance component of performance contributes to performance more critically. The name of the game in successful longer distance competition is sustained high-volume oxygen utilization to generate energy, with minimal performance inhibition due to accumulating tissue acidity. In the bloodstream, 98.5% of the oxygen is transported via attachment to hemoglobin molecules located in red blood cells. Hemoglobin is a large protein

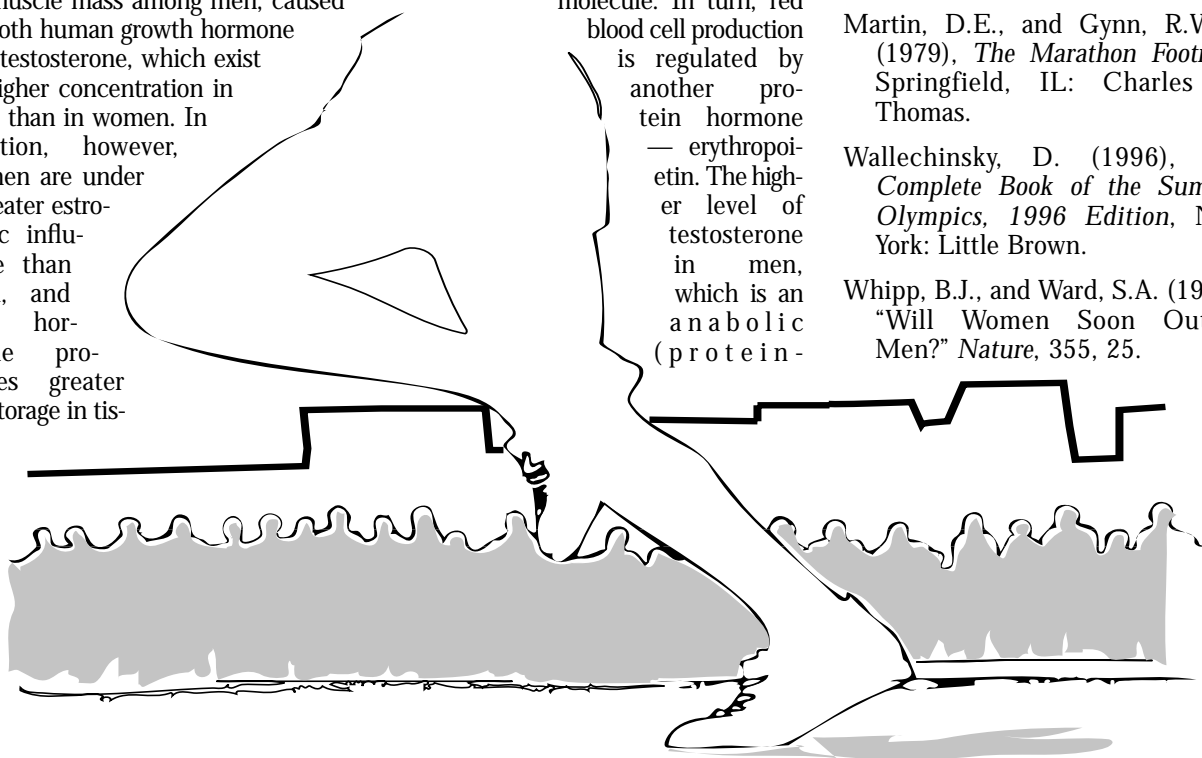
molecule. In turn, red blood cell production is regulated by another protein hormone — erythropoietin. The higher level of testosterone in men, which is an anabolic (protein-

building) hormone, provides them with both a measureably greater total hemoglobin and red-blood-cell mass. A typical mean value for blood hemoglobin concentration in men is 15 gm/dl; for women this is 13.5 gm/dl — a 10% difference. Calculation of blood oxygen (O_2) transport using these values gives 17.3 ml O_2 /dl of blood for women and 19.2 ml O_2 /dl for men — again a 10% difference. In addition to the blood oxygen *concentration* difference, there is a blood *volume* difference. Women have roughly 66 ml of blood volume/kg body weight (BW), compared to 77 ml/kg BW for men — a 14.3% difference.

Physiologists enjoy using such facts to estimate the influence of these sex differences on the existing performance differences seen in sport. Such differences explain clearly why women will never “catch up” to men, and indeed, why there is no need to. They are different people — vive la difference!

References and Further Reading

- Hill, A.V. (1924), “The Physiological Basis of Athletic Records,” *Lancet*, 2, 481-486.
- Julin, A.L. (1992), “No Improvement in Running?” *Nature*, 356, 21.
- Martin, D.E., and Gynn, R.W.H. (1979), *The Marathon Footrace*, Springfield, IL: Charles C. Thomas.
- Wallechinsky, D. (1996), *The Complete Book of the Summer Olympics, 1996 Edition*, New York: Little Brown.
- Whipp, B.J., and Ward, S.A. (1992), “Will Women Soon Outrun Men?” *Nature*, 355, 25.



Comment: Can We Trust a Method Just Because We Like Its Predictions?

Phillip N. Price

If the analysis and conclusions of Wainer, Njue, and Palmer (2000) are right, then we live in a special time: After 70 years in which men's and women's performances (both absolute and relative to one another) could be accurately extrapolated using straight lines, this method will stop working very soon. I happen to agree. But I wonder if there has ever been a time when people didn't feel that the runners and swimmers of their time were finally approaching the limits of human performance. Who knows, perhaps changes in training or diet, or in the rules regarding the use of steroids and other drugs, will allow improvements to continue this way for several more decades. I doubt it, but history should make us cautious.

An interesting feature of the analysis by Wainer et al. is that it attempts to predict differences between men's and women's speeds (in some future year in which records for both sexes have stabilized) without predicting the limiting speed for either sex. That's a very good thing, as it's clearly impossible to extract a reliable plateau speed based on historical data, at least for running: Just take a look at Fig. 10! In fact, as can be seen in Fig. 10, and as noted by Whipp and Ward (1992), not only are running speeds (still) improving linearly for most running events,

but they are doing so at about the same rate at all distances — of the order of 7 m/min per decade for men, and 16 m/min for women, within about 10% in each case. (These estimates do not quite agree with Fig. 10, which shows bronze-medal Olympic speeds rather than world records, but they are close enough). Of course, these improvements cannot keep up forever and perhaps not even for long, though the 70- to 100-year stretch covered in the figure is quite impressive as it is. In the swimming data shown in the article, there is evidence of a plateau already being approached, and the marathon records also appear to be stabilizing (the winning times for both men and women in the 1999 Boston Marathon were little changed from the 1992 records shown in Fig. 1, in contrast to the extrapolation that this should have been the year the women beat the men).

By coincidence, I began reading *Full House* (Gould 1996) shortly after being asked to comment on this article. *Full House* is about some consequences of physiological plateaus in athletics and other contexts, and Gould's Chapter 8 discusses men's and women's athletic performances. Gould agrees with most of the points of Wainer et al., particularly with regard to the error of predicting women's ver-

sus men's performances based on linear extrapolation from the past: he calls extrapolation "a dangerous, generally invalid, and often foolish game," which perhaps overstates the case a bit. He also says "In the most popular and established men's [running] events, we note the pattern of rapid initial improvement followed by flattening of the curve" and suggests that the same will happen soon in the women's events. Judging from Fig. 10, I think Gould and I must have different ideas about what "flattening of the curve" means.

In any case, Wainer et al. have set themselves a difficult task — to extract sensible estimates of relative future performances for the sexes without any way of predicting absolute performances. This is analogous to Cox's proportional hazards model (Cox 1972), in which it is possible to draw inferences about treatment and covariate effects without having to specify or estimate a baseline hazard function that quantifies the probability of dying before a given age.

Wainer et al. suggest a clever method for extrapolating relative performances without delving into the more difficult issue of absolute performances: Shift the women's record-versus-year curve in both dimensions (speed and year) to extract two para-

meters rather than one. These shifts have simple interpretations: the shift in speed represents the amount by which women are intrinsically slower (or faster) than men, and the shift in year represents the amount by which women's approach to their top speed lags that of men due to dramatically unequal participation and training of men and women in the past and, possibly, the present. I think this separation of the variation into two components, a lag in years and a gap in speed, is a substantial conceptual advance.

All of which is fine as far as it goes, but the actual application is more in the realm of curve fitting than model fitting. I certainly agree that the rapid increases in speed in each sex's first 15 years of the Boston Marathon are largely attributable to increases in the pool of participants so that qualitatively these 15-year periods are comparable. But the method suggested by Wainer et al. requires more than qualitative comparability: It attempts to extrapolate based on the exact shapes of the speed-versus-year curves. Does the growth in the number and skill of male Boston Marathon participants around 1900 really match that of female participants 80 years later so well that we can use the race results from these widely separated periods to estimate differences in male–female future performances that are of the order of 2%? I'm rather skeptical, even though the results are in line with expert opinion (as they are for the other races considered in the article).

Furthermore, notice the standard that I (along with Wainer et al.) am applying here: we are judging the applicability of the method not by comparison to data, but through comparison to expert opinion regarding the future speed gap. Of course, there is nothing wrong with including expert opinion in making predictions, and indeed this is often done in Bayesian modeling by using informative prior distributions. But there is a distinction: in Bayesian models, the prior distribution can affect the inferences directly, but that is not the case here. For example, suppose we obtain data for some race not considered by Wainer et al. (the 1500-m run, say, or the 100-m backstroke), apply their method, and discover that the best fit is obtained if

the women's curve is shifted back 40 years and down 5% rather than up. Experts would judge it implausible that women will eventually be 5% faster than men in these races, but the current method has no way of allowing this judgment to influence the prediction. Would we then conclude that the experts are wrong and that the model results are right and women will be 5% faster than men in these races? I doubt it; at least, I wouldn't bet that way.

It is intriguing, though, that the results of applying Wainer's method to a variety of races, in both running and

Wainer et al. have set themselves a difficult task — to extract sensible estimates of relative future performances for the sexes without any way of predicting absolute performances.

swimming, all yield reasonable results (in the sense of being in line with expert opinion) for persistent male–female performance differences. This is particularly interesting because the method doesn't just tolerate curvature in the speed-versus-year curve; it demands it: if improvements for both sexes were truly linear, then the trade-off between left-right and up-down shifts would be undefined. There is some evident downward concavity in the swimming data presented, but none (to my eye) in the running data, so I find it surprising that the method generates sensible results for those cases too. Thus, in spite of being skeptical for the reasons mentioned in the previous paragraph, I admit the possibility that Wainer et al. might be onto something.

And of course, it's easier to criticize than to suggest improvements. Suppose I wanted to estimate future male–female performance differences,

what would I do? Realistically, for now I would poll the experts, just as, if I were interested in predicting the outcome of a future football game, I would rely on the point spread rather than a statistical model. But that's a pretty unsatisfying answer to publicly state in a statistics magazine. And of course, just because someone is an expert, he or she is not necessarily right. How many track and field experts who witnessed Bannister breaking the four-minute mile would have thought that the mile record would someday be under 3:44, as it is now?

OK, so I'm not convinced of Wainer's method, nor do I want to just rely on expert opinion. What's the alternative? I'll start by suggesting some future work that builds on the work of Wainer et al. No prediction is complete without an estimate of uncertainty. Even if the method of Wainer et al. happens to be perfectly suited to the actual application, stochastic variation in the performances will lead to variation in the predictions depending on which years of records are considered. The sensitivity to such stochastic variation could be assessed by examining the stability of the predictions for various sets of data: What are the predictions if silver-medal speeds are used instead of bronze, if data after 1980 (or prior to 1960) are excluded, and so on. The resulting variation in predictions would set a lower bound on the uncertainty; of course, it wouldn't address the major issue of model (or rather method) misspecification.

Beyond that, I can only suggest the more difficult task of modeling physiology (or at least the relevant male–female physiological differences) directly. Whipp and Ward (1972) discussed relationship between speed improvements and changes in metabolic rate, a possibly helpful way of looking at the issue. Wainer et al. point out in their discussion that there appear to be inherent differences in upper-body strength between men and women, and that this ought to translate into speed differences, particularly in upper-body-intensive sports such as swimming. A more precise estimate of the persistent male-female strength difference

or power might be obtained (how?) and converted into an expected difference in swimming speed by regressing speed on strength for male and female swimmers (it would be interesting if the regression slopes turn out to be different). Alternatively, scaling laws based on weight, muscle mass, fat mass, and so forth, could perhaps be used, somewhat along the lines of Schmidt-Nielsen (1972), which is, by the way, the best short data-based book ever written.

None of these suggestions have the elegance and simplicity of the method suggested by Wainer et al., and in fact, they aren't likely to provide much better estimates of eventual sex differences in performance either. It's just a tough problem. I think that Wainer et al. have made a significant contribution to this area of study by suggesting that performance differences should be decomposed into two components (year and speed) rather than one, but I don't really trust the results of their suggested method. I suspect we won't have better data-based estimates (as opposed to expert opinions) until some events start showing evidence of approach to a speed asymptote, a phenomenon that must surely occur sometime, and possibly soon, but which (at least for running) is not apparent in the data so far.

References and Further Reading

- Cox, D.R. (1972), "Regression Models and Life-Tables," *Journal of the Royal Statistical Society, Ser. B*, 34, 187-220.
- Gould, S.J., (1996), *Full House: The Spread of Excellence From Plato to Darwin*, New York: Three Rivers Press.
- Schmidt-Nielsen, K. (1972), *How Animals Work*, New York: Cambridge University Press.
- Wainer, H., Njue, C., and Palmer, S.J. (2000), "Assessing Time Trends in Sex Differences in Swimming and Running," *Chance*, 13(1), 10-15.
- Whipp, B.J., and Ward, S.A. (1992), "Will Women Soon Outrun Men?" *Nature*, 355, 25.

Harvard University Press
2/3 page ad
art supplied
printer to strip in

Sex and Sports: A Rejoinder

Howard Wainer, Samuel Palmer, and Catherine Njue

We would like to thank Professors Martin and Price for their careful reading and encouraging suggestions. They have provided for us an inkling of what Picasso must've felt like when a recreational scribbling of his was sold for thousands of dollars. We are delighted that something that was so much fun to do has elicited the praise and thoughtful consideration of such accomplished scholars.

Professor Martin has provided important physiological facts that can be used to explain the effects of our model. It might be useful in the future to try to include them as covariates and see to what extent the accuracy of the model's predictions are improved. In addition Professor Martin's facts help to answer Professor Price's concern about the efficacy of our approach. Price feels that the principal value of our approach is that it seems to match expert opinion (not a bad first step, we reckon). Martin's description of the inexorable chain of physiological events that begins with sex differences in hemoglobin and testosterone concentrations, affecting differences in muscle mass and oxygen-carrying capacity, yielding differences in speed and endurance, provides a compelling supporting argument for our separation of the effects of physiology and participation.

Implicit in Professor Price's praise of our two-parameter conceptualization is the notion that if sufficiently good data can be obtained (see Palmer (in press) for a fuller discussion of this) we can get better adjustments for participation and so isolate the truly physiological effect. At this point trying to connect the estimates of measured physiological variables with adjusted performance could be enormously informative.

Of course, participation rates are inadequate to accurately measure what is fundamentally a self-selected process (Wainer, Palmer, and Bradlow 1998; Wainer 1986), but they do provide a beginning. When coupled with the sort of mixture modeling proposed by Rubin (1977), they can help us to obtain good estimates of the precision of our parameters, which reflect more than what is available from traditional sampling error.

Price also correctly points to the flat response surface explicit in our fitting procedure as a source of plausibly substantial sampling variability. As this is being written, we are trying to debug a computer program designed to assess this variability. It uses a specialized variation of the jackknife and when complete should provide us with a fuller notion of the stability of our estimates. Unfortunately, it appears that the publishing deadline will reach us before the results.

What if the women's curve needs to be shifted down 5%? Professor Price suggests this as an outcome that would invalidate our model. Because such an outcome (if it were not the result of a sampling anomaly) would violate the physiological implications that form the basis of our model, he is correct. But any saltworthy model must be potentially disprovable. The model then gains credence as additional data are gathered that are still supportive. Note, however, that our model does allow women to exhibit superior performances to men when they have greater participation and the gains due to participation outweigh the losses associated with physiological differences. An example of such a situation might be in synchronized swimming, which is almost totally dominated by

women. If men suddenly decided to compete, we would expect that, for a while at least, women's performances would be superior.

Of course, we have focused on sporting events that are scored in a completely objective way (a stop watch), but it is not inconceivable that sports based on human judges could also fit into this mold. Our preference for extending these results would not be in this direction but rather toward those sports (like archery) in which strength and endurance play a vastly reduced role. Here we would expect that participation rates would be the sole determinant of relative performance.

We are not unaware of the plausible applicability of this methodology to the study of sex differences on performance in various academic skills.

References and Further Reading

- Palmer, S. (in press), "Sex Differences in Swimming Then and Now," *Swimming World*, 35.
- Rubin, D.B. (1977), "Formalizing Subjective Notions About the Effects of Nonrespondents in Sample Surveys," *Journal of the American Statistical Association*, 72, 538-543.
- Wainer, H. (1986), *Drawing Inferences From Self-selected Samples*, New York: Springer-Verlag.
- Wainer, H., Palmer, S.J., and Bradlow, E.T. (1998), "A Selection of Selection Anomalies," *Chance*, 11(2), 3-7.