Comparing the Best Athletic Performances of the Two Sexes

29:4; 37-48, 2014

by Basil Grammaticos and Yves Charon

ABSTRACT

This paper addresses the question of the gender dependence of athletic performances. The comparison of women's to men's performances shows 1) there is a strong dependence of the ratio of the performance to the world ranking of the performer in both genders and 2) this dependence is smooth to the point that it allows the introduction of a simple model that makes it possible to recover the ratios of the world records with precision. The analysis shows a significant difference in gender dependence between running and jumping events. Finally, the temporal evolution of the ratio of men's-women's performances is examined. While in some cases the world record ratio is stationary, in disciplines like the triple jump, only relatively recently introduced for women, the ratio of the 50th performers shows a constant improvement of the results of the women compared to those of the men of the same rank. The results lay the basis for a systematic approach to the analysis of athletic performances and in particular to the comparison of the gender dependence of the latter.

AUTHORS

Basil Grammaticos, Ph.D., is a Director of Research at the National Scientific Research Centre (CNRS) in Paris and is currently head of the modelling team of the joint University of Paris VII – University of Paris XI Laboratory for Imaging and Modelling for Neurobiology and Cancer Studies (IMNC).

Yves Charon, Ph.D., is a professor at Paris VII University and Director of the joint University of Paris VII – University of Paris XI Laboratory for Imaging and Modelling for Neurobiology and Cancer Studies (IMNC).

Introduction

This article stems from a discussion between the two authors on the comparison of the performances of men and women. The first question one has to answer upon embarking on such an analysis is: what is the reference performance to be used in the comparison? We intend to present some elements towards possible answers to this fundamental question, laying down the methodological basis for performance comparisons.

Obviously the current world record cannot be used as a reference performance. This record can be due to an exceptional combination of talent and circumstances, like, for instance, Beamon's 1968 long jump record. It might also be due to lower standards as far as control of the competition is concerned (wind-speed limit measurement, application of anti-doping controls, etc.). In some cases there may exist a clear disparity between the two sexes, like in the case of the javelin throw, where it took more than 10 years for the women's javelin to be redesigned along the same lines as the men's.

The same arguments apply to the inadequacy of the Olympic record as a reference performance and, ***a fortiori***, to the performance of the Olympic or world champion of any specific year. Moreover, in the case of major, global, championships it is clear that the rank is much more important than the performance, which may, in particular in middle- and long-distance events, lead to tactical races and thus below-par performances. Weather and other unforeseeable or uncontrollable conditions may also have a non-negligible effect upon the results of major championships.

Another possibility would have been to use the official scoring tables in order to establish the equivalence between performances. One could thus compare men's and women's marks scoring the same number of points. Even disregarding the fact that the choice of the proper number of points is something non-trivial, the use of the scoring tables presents a substantial drawback. Even assuming that the tables provide a fair scoring (something that has been challenged in the past) they do get updated (and quite justifiably so) so as to make any study of temporal evolution more difficult.

After having presented these entires negative answers one understands that some serious analysis is in order. Clearly the reference performance can only be determined by analysis of a substantial volume of data. That been said, two main approaches appear possible at this stage: either implement some “standard'' statistical analysis or treat the body of data in the framework of some model. The latter is the approach we shall adopt in this paper.

In what follows we shall begin by an analysis of the performances of the world’s elite athletes, which will reveal a remarkable distribution of the performances versus the rank of the performer. This allows the representation of the results in terms of a model we have introduced in a previous work in which we propose of a fair system of scoring. Next we compare the performances of the two sexes using the statistics at our disposal and show again a strong but quite regular dependence on the rank of the performer. Finally we address the question of the time-evolution of the relative performances of the two sexes and show that in the relatively new disciplines for women there is a constant progression of the female performers compared to the male ones.

An Analysis of the World’s Elite Performances

The source of data we are going to use throughout this paper is the manual of the Association of Track and Field Statisticians[1]. The advantage of this manual (updated annually) is that as of 2011 it gives the all-time elite performers to a depth of 500. Only the best performance of each athlete is taken into account, contrary to what one can find in other “all-time best'' lists. The risk in this case is that the best performance of an athlete has profited from exceptional conditions (something that made the world record unsuitable as a reference performance). However the smoothness of the dependence of the performance on the rank of the athlete and the stability over time does justify this choice.

Figure 1 presents the best long jump marks for men and women respectively for the performers with rank 1, 10, 30, 50, 100, 200, 300, 400 and 500.

The data show an extreme regularity and the same is true for all the other events we anal-

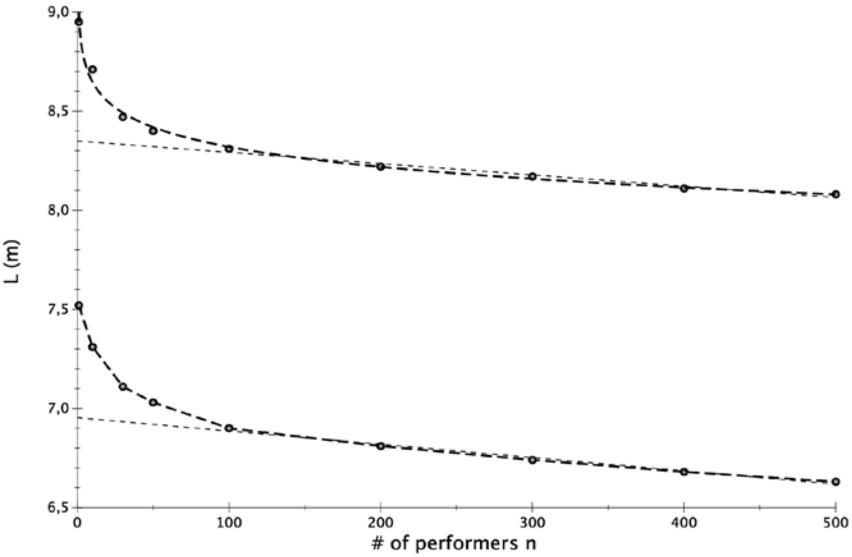

Figure 1: Performances of the all-time best 500 male (upper) and female (lower) long jump athletes versus their rank (The long-dash line corresponds to the best fit with the model (2) while the short-dash line is a linear fit over the athletes with rank form 100 to 500.)

ysed, which included the 100m, 400m, 1500m, 10,000m, marathon and all four jumps. We decided not to examine the throws in depth since any attempt at a comparison between the two sexes would have been thwarted by the choice of different weight implements. Still, a discussion of a possible comparison of men's and women's throwing performances will be presented in Section 3. The argument concerning throws applies, up to a certain point, to the hurdles events as well. The shape of the curve followed by the distribution of the performances over the population of athletes if seen (by imagining an exchange of axes) as the number of athletes realising a performance equal or superior to some value shows a fast, almost exponential, decrease towards the highest performances. This suggests immediately the use of the model introduced in GRAMMATICOS[2] based on the theories of HARDER[3] (his famous "apples to oranges" approach), in which we proposed a scoring scheme based on an analytical assumption for the distribution of performances of the form where *f* is the percentage of the population realising some

$$f(L) = \frac{A}{1 + Be^L} \qquad (1)$$

performance, which is represented here by the normalised (dimensionless) quantity *L*, length for jumps and velocity for races. The symbol e is the basis of the natural logarithms and *A, B* two constants. Since here we are giving the performance as a function of the number of athletes, we must invert (1). We thus obtain the expression

$$L = a + b\log(c/n - 1) \qquad (2)$$

where again, *L* is the normalised performance, *a, b, c* three constants and log is the symbol for the natural logarithm. Here *n* is the number of athletes and thus the constant *c* plays the role of the capacity (i.e. total number) of the elite population. The dashed lines in Figure 1 represent the best fit of expression (2) over the data. In practice we adjusted the parameters *a, b, c* so as to minimise some distance between the data points and the curve at the same abscissa (using the chi-square-based fit of our graphics application). As one can assess at a glance, the quality of the fit is excellent. The values of the parameters for the upper (men) and lower (women) parts of Figure 1 are respectively

a = 7.960m	b = 0.1358m	c = 1600
a = 6.655m	b = 0.1324m	c = 890.0

The value of ***c*** for men's long jump is one of the largest encountered in our fit of all the disciplines considered. Typical values are less than 1000.

The almost straight line along which lie the performances of the 100th up to 500th performer, suggests a possible reference performance to be used in comparisons between the two sexes. By fitting a straight line over these five points we obtain a performance corresponding to an abscissa n = 1, which can be used as a record corresponding to the elite once the top performers are excluded. In Figure 1 the straight-line fit is represented by the dotted line. We find thus a reference performance L = 8.35m for men and L = 6.95m for women. As a matter of fact, there exists another characteristic straight line, which could be used in order to introduce a reference performance - that of the tangent at the point where the curve defined by (2) has zero curvature.

We shall not go into the calculations related to this but give directly the result: once the fit of an expression (2) has been performed and the parameters ***a, b, c*** obtained, the reference performance is $L = a + 2b$. This would lead to values of 8.06m for men and 6.92m for women. Unfortunately these performances cannot be used for comparisons between the two sexes for a very pedestrian reason: we, the present authors, possess statistics including performances up to the 500th athlete only for the last two years. Thus the use of a reference performance based on data over 500 athletes would not allow any study of the evolution over time.

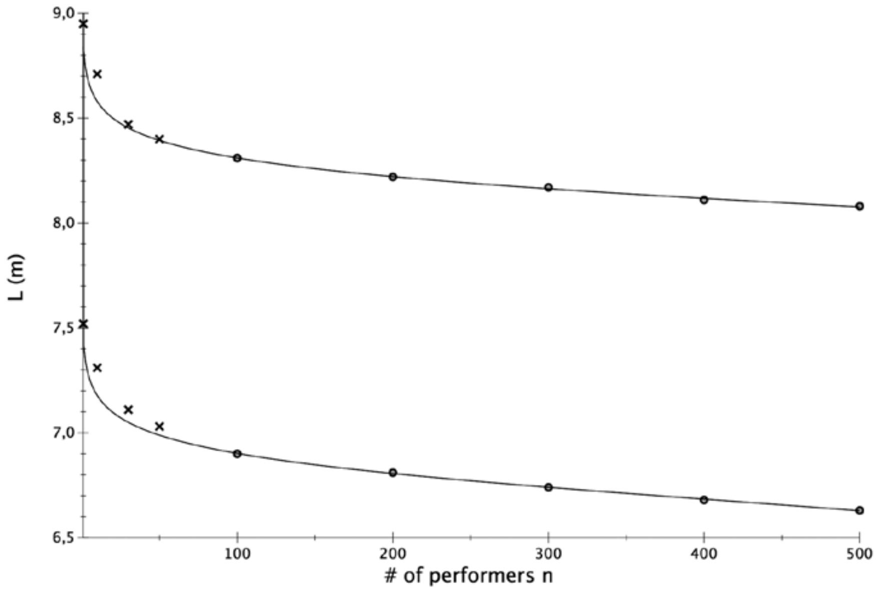

Figure 2: Performances of the all-time best 500 male (upper) and female (lower) long jump athletes versus their rank (The continuous line corresponds to the best fit with the model (2) based on the data on athletes with rank form 100 to 500 (dots). The crosses correspond to data on athletes with rank form 1 to 50, not used in the fit.)

Before proceeding to the comparisons section we would like to present a result, which, we must confess, we did not expect in the least. After having used the results for the population (100th to 500th) in order to fit a straight line we decided to make a small experimentation. Instead of fitting a straight line we tried to find the best expression (2) corresponding to the points n = 100,..., 500. The astonishing result is that, although the points n = 1,10,30,50 are absent from the fit, the resulting curve stays very close to their corresponding values, as one can see in Figure 2.

Of course, the values of the parameters have changed. We now have (always for the long jump)

a = 8.040m	b = 0.1129m	c = 1194
a = 6.675m	b = 0.1140m	c = 832.9

Using these values in expression (2) we find when n = 1, i.e. at the world record level, the predictions L = 8.84m for men and L = 7.44m for women, which are very close to the actual values of the current world records. Thus, in some sense, the distribution of the performances of the elite contains information about the top performers all the way up to the world record. Emboldened by this result, we proceeded to obtain the best fit of expression (2) for all disciplines analysed and the model elite-only-based prediction for the world records. Table 1 shows our findings.

We remark that the agreement with the existing (2013) records is quite spectacular, the largest deviation being of the order of seven percent in the case of the women's triple jump. There is also a systematic tendency for our method to overestimate the world records for women. A possible explanation to this is that the elite female population is less numerous than the male one and thus the slope of the curve linking the 100th to the 500th performer is steeper for women.

These extrapolated world records could constitute another reference performance but, again, they necessitate knowledge of 500-deep statistics, which prohibits their use in the current state of affairs.

Comparing Men's and Women's Performances

Having established the importance of data on elite athletes (and not just on the very top performers), we can now proceed to the main theme of this paper, namely the comparison of performances of men and women. Instead of hunting for some complicated expression involving the performances, we opted for a simple ratio of the men's mean velocities over those of the women in the case of track events and of distances, those of women over those registered by the men, in the case of field events. Just as in the analysis presented above, we have kept only the elite results, from 100th to 500th, although the results do not change much if we include the top performers as well (with the exception of the world records).

Figure 3 shows the behaviour of the ratio of velocities for the 100m (similar results being obtained for all the events analysed). The tendency is systematic: the 200th female performer is doing less well, compared to the 200th male, than the 100^{th} female compared to the 100^{th} male and the same goes on as we move to athletes with lower ranks. This is probably due to the elite female population being less numerous than the male one. If that were the main reason for the observed tendency, the effect would have been more important in relatively "new" disciplines like triple jump and pole vault, something which is not incompatible with our results and the extrapolated world records presented in Table 1.

The data in Figure 3, but also in all the other cases analysed, show a very smooth dependence of the ratio on the rank of the athlete. This has led us to try to fit the curve with a simple function of the form

$$r = \frac{f}{n^k} \qquad (3)$$

Table 1: Predictions for world records (the percentage of difference from the records as of 31 December 2013, a negative sign indicating that our prediction overestimates the world record)

Event	Men	%	Women	%
100m	9.68 sec	1.0	10.48 sec	-0.1
400m	43.46 sec	0.6	47.30 sec	-1.0
1500m	3 min 26 sec	0.0	3 min 43 sec	-0.7
10,000m	25 min 57 sec	-1.3	29 min 18 sec	-0.7
Marathon	2 hr 03 min 32 sec	0.1	2 hr 12 min 57 sec	-1.7
Long Jump	8.84m	1.2	7.44m	1.1
Triple Jump	18.63m	-1.9	16.58m	-7.0
High Jump	2.45m	0.0	2.14m	-2.4
Pole Vault	6.30m	-2.3	5.18m	-2.4

where r is the ratio of women's to men's performances and n is the rank of the athlete. In Table 2 we give the results of the fit for all the events analysed.

Using expression (3), for n = 1 one gets r = f and thus f can be interpreted as the model prediction for the ratio of the world records. In fact, the value of r thus obtained overestimates the actual value of the ratio of the world records by something of the order to one percent.

Quite astonishingly, even the uptick of the ratio for the marathon, compared to the 10,000m, is correctly reproduced. A cursory glance at results in Table 2 shows that while the ratios for track events are grouped around 0.90, those for the field events have values clustered around 0.85. (We are talking here about the best possible ratio, i.e. the model-predicted one for the world record. As shown in Figure 3, the ratio decreases when we move to athletes with lower rank).

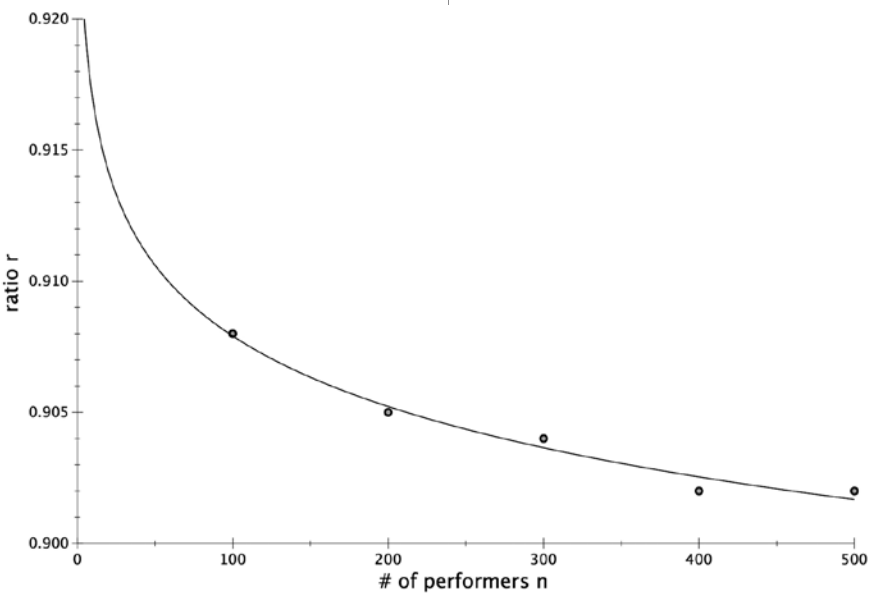

Figure 3: Ratio of the mean velocities over 100m of women over men for the all-time best athletes with ranks from 100 to 500 versus their rank (The continuous line corresponds to the best fit with the model (3).)

Table 2: Ratio of top women's performances to top men's performances

Event	f	k
100m	0.923	0.0043
400m	0.925	0.0087
1500m	0.905	0.0050
10,000m	0.901	0.0064
Marathon	0.925	0.0107
Long Jump	0.856	0.0064
Triple Jump	0.886	0.0139
High Jump	0.861	0.0044
Pole Vault	0.842	0.0193

We shall not attempt here any deep analysis of the origins of the disparity of women/men performance ratios between the track events and the jumping events. Clearly this can only be done in the framework of a serious biomechanical model taking into account the anatomical and physiological differences between the two sexes, something well beyond the scope of this article.

However, before concluding this section we would like to devote a few lines to the discussion of the situation for throws. As a matter of fact, we performed the analysis described in the first section of this article for the disciplines of shot put and discus throw and the results were in perfect agreement with those obtained for the other disciplines. However, due to the fact that implements of different weights are used for the two sexes it does not make much sense to proceed to explicit comparisons. Given that the world records for the two mentioned disciplines are very close one could conclude that the historical choice of weights is optimal and that when it comes to upper-body force comparisons there exists a factor of almost two between men and women. In fact, there exist several studies[4] that seem to corroborate this.

We are of a different opinion. First, when one analyses the weightlifting records for the heaviest category one finds a ratio of roughly 0.7 between women's and men's performances. Second, in a previous work[5] we have shown that decreasing the weight of an implement does not lead to a proportional increase of the performance. We concluded that work with a speculative analysis on the results of Olympic and World Champion Valerie Adams (NZL) and surmised that (according to our model predictions) she should be able to put a men's shot at around 15m. Given a reasonable men's top performance of 22m this would give a women/men ratio of around 0.7, which we believe to be more realistic when it comes to elite comparisons.

On the Evolution of Relative Performances Between Men and Women

One of the works that had motivated the discussion over gender difference in athletic performance was a recent article by DUPUY[6], who claimed that the women's/men's world record ratio follows an S-shape curve over time, and, in particular for the newer disciplines of triple jump and pole vault. The S-shape evolution means that at the beginning the records of women progress slowly, if at all, with respect to the men's records, then they enter a phase of fast progress and finally this progress levels-off.

The argument of the author is that the pattern of the world record ratio is due to a sudden drop is the social barrier for women to participate in new disciplines, which leads to an enhancement of ability self-selection. While we believe that the arguments of the author are essentially valid, we cannot accept his results at face value since they are based on a comparison of world records - something that does not necessarily represent the true state of a discipline. In order to illustrate this we are going to compare a long-established discipline to some new' ones, including the triple jump and pole vault analysed by DUPUY.

In Figure 4 we present the results for long jump, a discipline with an official history going back to the beginning of the 20th century and for which we are in possession of detailed data. It is remarkable that not only is the ratio of the world records constant (since 1991) but also the ratios of the 50th, 100th and 200th performers is practically constant on the scale on which newer disciplines show a substantial evolution. Long jump is a discipline that has reached full maturity.

We turn now to the “new" disciplines of 5000m and 10,000m, triple jump and pole vault and analyse the evolution of the relative women’s/men’s performances in the light of the analysis presented in the previous section. Given the relative paucity of statistical data, which does not allow us to use any of the reference performances proposed in the first section, we opted by considering the performance of the 50th ranked performer. Figures 5, 6 and 7 give the parallel evolution of the world record ratio and the ratio of the 50th performers for men and women. The resulting behaviours are quite different for the four disciplines, in particular as far as the world record ratios are concerned.

In Figure 5 we are representing together the results for the 5000m and 10,000m. We remark readily that the behaviour of the velocity ratios of the world records is quite different from that of the 50th performers. In the case of 10,000m, a singular women's record established in 1993, and unbeaten since, is at the origin of the down-sloping evolution of the ratio (full triangles in Figure 5, joined by long dashes). The record of 5000m, on the contrary, was last established in 2008, which explains the up-surge of the corresponding curve (triangles joined by short dashes). Meanwhile the ratios for the 50th performers are evolving regularly with a slow-down in the last decade compared to that of the 1990s.

The case of the triple jump, Figure 6, is more interesting. While the world record ratio has not changed since 1995, the ratio of the 50th performer continues to grow. Of course, the rate today is less pronounced compared to that of the 1990s, but the ratio does increase. This is a perfect sign that the discipline is evolving, as far as women are concerned. In any case the saturation predicted by the analysis of DUPUY is absent once one considers the top elite athletes instead of just the world-record holders.

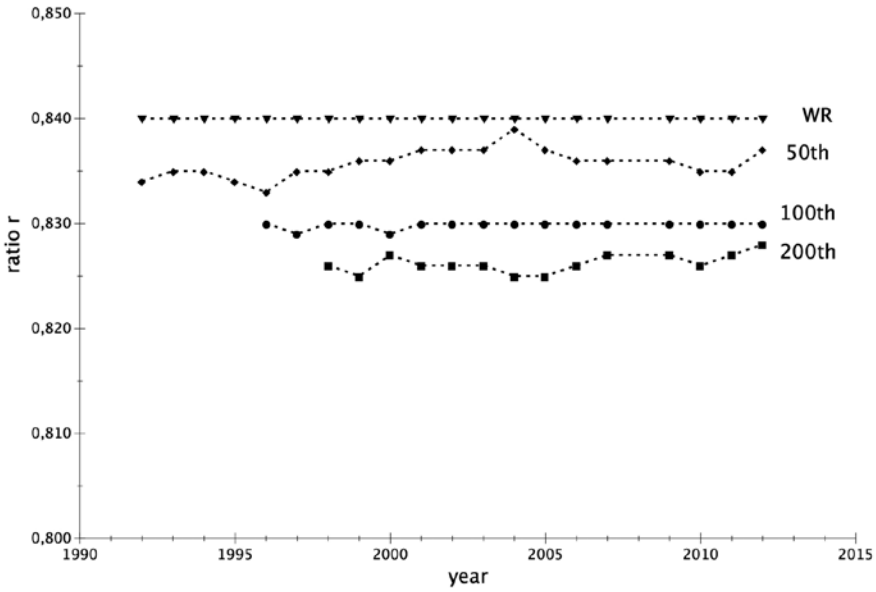

Figure 4: Ratio of long jump lengths of women over those of men for the world record holders and for the athletes with ranks 50, 100 and 200 in the all-time lists

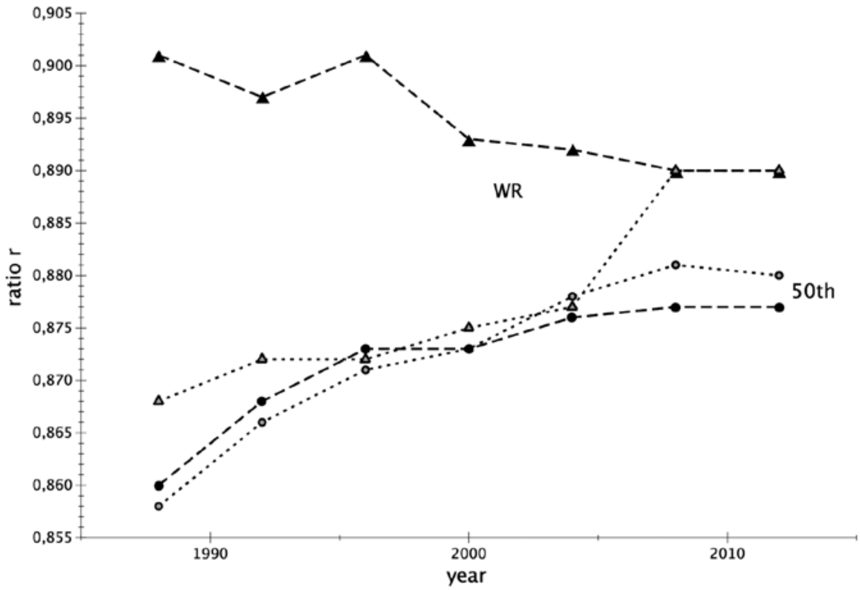

Figure 5: Ratio of mean velocities of women over those of men for the world record holders (upper curves) and for the athletes with rank 50 in the all-time lists (lower curves) as a function of time for the disciplines of 5000m (open symbols) and 10,000m (full symbols)

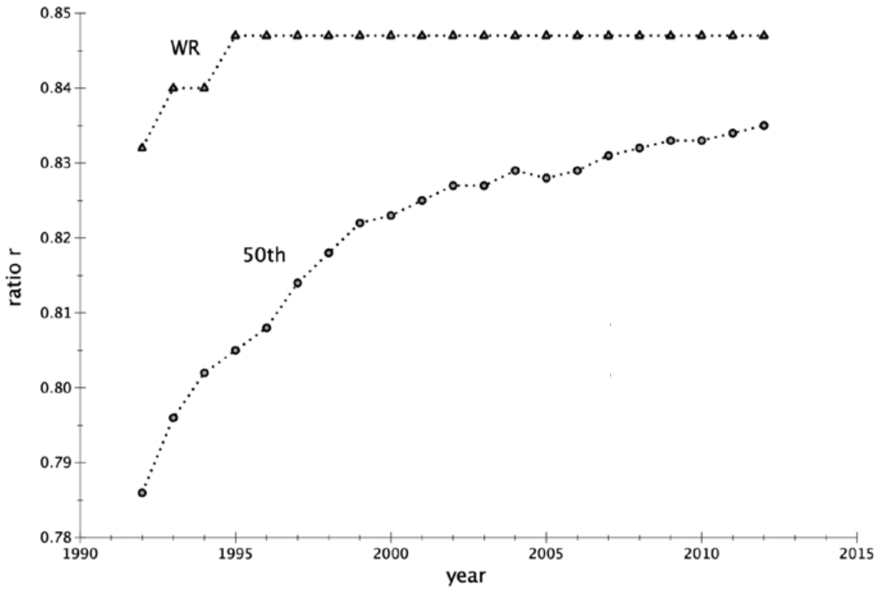

Figure 6: Ratio of the triple jump lengths of women over those of men for the world record holders (upper curve) and for the athletes with rank 50 in the all-time lists (lower curve) as a function of time

Finally, for the pole vault, the situation, as can be assessed from Figure 7, is even clearer: no saturation of the womens'/mens' performance ratio can be seen, even at the world record level. Moreover, the gap between the world-record ratio and the one of the 50th performer did widen over the past decade (the unique reason for this, and given the fact that the men's record did not progress over the last 20 years, being the appearance of one exceptional athlete). All in all, the ratio for the 50th performer is slightly lower than that of the other jumps, but given that at the level of the world record the ratio is not particularly small (despite the fact that the men's world record is an impressive one) we believe that there is a substantial progression margin for female pole vaulters.

Thus our conclusion is that when one focuses on some reference performance (be it not optimal as argued above) rather than the world record one can appreciate much better the state of a discipline and assess in a more precise way its level of maturity.

Conclusions

In this article we set out to address the question of the comparison of the athletic performances of men and women with a particular emphasis on the proper methodology to be applied. Before attempting any comparison one must decide on the reference performance to be used. This is something highly non-trivial, and as we have argued throughout the article, the customary use of the world record as reference performance may be misleading (at least as far the relative evolution of a discipline is concerned).

The approach we opted for is one based on mathematical modelling based on existing statistical data. We have thus analysed the results of the top-500 performers of all time, as compiled by the Association of Track and Field Statisticians. We have shown that the distribution of the performances can be perfectly described by a model introduced in GRAMMATICOS[2], based on the theories of

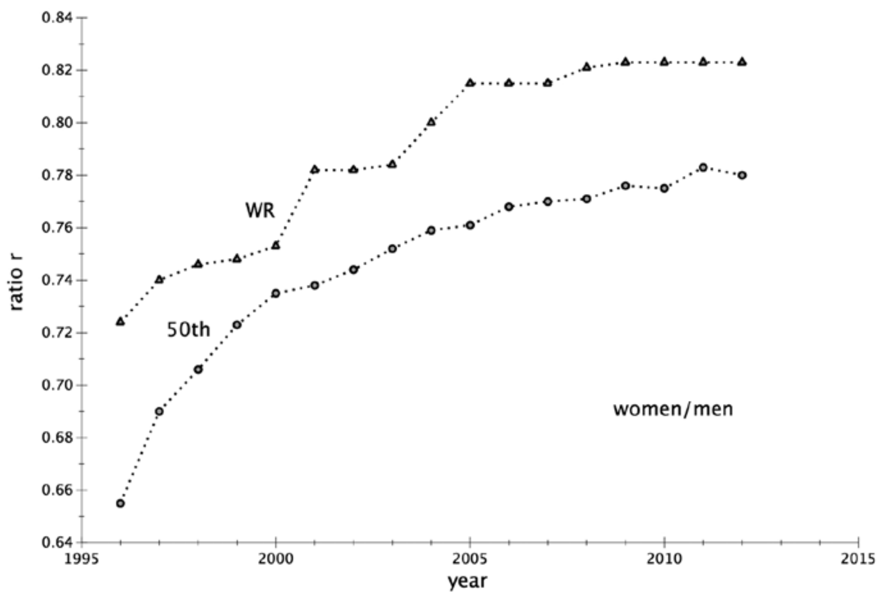

Figure 7: Ratio of the pole vault heights of women over those of men for the world record holders (upper curve) and for the athletes with rank 50 in the all-time lists (lower curve) as a function of time

HARDER[3], which aimed at the construction of efficient scoring tables. The main difference between the work in GRAMMATICOS[2] and the model presented here is that here the model is tailored to a description of an elite population of, at best, a few thousands of individuals, contrary to a scoring approach, which must encompass a much wider population. Our statistical treatment and model-based analysis allowed us to propose reference performances that could be used for comparisons. However we were not able to continue in this direction since the necessary statistical data are only available for the last two years.

An astonishing result of the model-plus-data approach was the fact that using just the data for the elite population from the 100th to the 500th performers we were able to extrapolate backwards towards a hypothetical world record, which turned out to be fairly close to the actual value, the worse discrepancy being of the order of a few percent.

In order to proceed to a comparison of the performances women versus men we decided to use simply the ratio of mean velocities for track events and the ratio of lengths for field events. As argued in GRAMMATICOS[2], velocities and lengths are directly related to the energy expenditure during the effort and thus do characterise not only the objective performance but also the subjective, physiological, one. The interesting result when one analyses the data for the elite population is that the ratio of women's/men's performances decreases with the rank of the athlete. We do not believe this to be just a recruitment effect: old, well-established, disciplines show it too. It turned out that the dependence of the ratio on the rank is very smooth and can be described by a very simple model.

Moreover, the model did allow us to extrapolate backwards from the elite population composed of the performers with ranks 100 to 500 and obtain the hypothetical ratio of the world record, which turned out to be very close to the actual value, with a 1% accuracy. The comparison (at the level of the extrapolated world record) clearly showed two groupings. For track events, the top female performers have mean velocities that are roughly 90% of those of their male counterparts. For jumps on the other hand, the ratio of distances (or heights) is, at maximum, of the order of 0.85. (It is even slightly smaller in the case of the pole vault, but given the fact that the discipline is quite new we cannot tell whether there exists a physio-anatomical reason for this). We have also addressed the question of throws, through the example of shot put, where it appears clearly that elite women perform at no more than 70% of their male counterparts.

One more topic treated in this work was the time-evolution of records, in view of the results of DUPUY[6], who claimed that the ratio of women's/men's world records follows an S-shaped curve over time. While we abide by his, essentially sociological, analysis we pointed out the inadequacy of using the ratio of world records instead of data based on elite performers. We have analysed four disciplines, each rather recently introduced for women, namely 5000m, 10,000m, triple jump and pole vault. We have shown that for all four the ratio obtained from the performers who ranked at the 50th position is constantly increasing. In the case of pole vault, the most recently introduced of the four, both the ratios, that of the world record and that of the 50th performer, are increasing, however the latter is doing so faster than the former. The open question for this discipline is, of course, whether the ratio of world records will reach the same asymptotic value of 0.85 seen in the other three jumps.

Our analysis has, hopefully, shown how detailed statistical data are important in order to formulate useful models for the athletic performance. Had we had at our disposal even more deep lists of the best performers say up to the 1000th, ranked, we would have been able to put our models to more stringent tests. A follow-up of such lists over 10 or more years would also have allowed a much finer analysis. Concerning the domain of throws, which has been largely ignored in this paper, what would have been most useful is a collection of anecdotal evidence of women throwing men's implements (and also of the opposite, although we are aware of the practical difficulties due to the length of the throws, with the exception of the shot put). Finally, one direction to which the present study can be extended is that of the comparison of the performances of men and women as a function of age. Best performance lists do exist for Masters athletes, so such a study does not seem impossible, although recruitment effects, in particular for the more advanced age categories, are expected to be significant.

Please send all correspondence to:

Basil Grammaticos
grammaticos@imnc.in2p3.fr

REFERENCES

1. MATTHEWS, P. (2014). Athletics: The International Track and Field Annual, published annually, recent editions by Sportsbooks.

2. GRAMMATICOS, P. (2007). The physical basis of scoring the athletic performance. *New Studies in Athletics*, 22 (3): 47.

3. HARDER, D. (2001). Sports comparisons: You can compare apples to oranges, Castro Valley, CA, USA: Education Plus. The web site of D. Harder is www.strengthhopedia.com.

4. MILLER, A.E.J.; MACDOUGALL, J.D.; TARNOPOLSKY, M.A. & SALE, D.G. (1993). Gender differences in strength and muscle fiber characteristics. *European Journal of Applied Physiology*, 66 (3): 254.

5. GRAMMATICOS, B. (2014). Shot put with lighter implements. *New Studies in Athletics*, 29 (1): 75-81.

6. DUPUY, A. (2012). An Economic Model of the Evolution of the Gender Performance Ratio in Individual Sports. *International Journal of Performance Analysis in Sport*, 12 (1): 2

7. JANSSEN, I.; HEYMSFIELD, S.B.; WANG, Z.M. & ROSS, R. (2000). Skeletal muscle mass and distribution in 468 men and women aged 18-88 yr. *Journal of Applied Physiology*, 89 (1): 81.

8. LEYK, D.; GORGES, W.; RIDDER, D.; WUNDERLICH, M.; RUTHER, T.; SIEVERT, A. & ESSFELD, D. (2007). Hand-grip strength of young men, women and highly trained female athletes. *European Journal of Applied Physiology*, 99: 415.