Linear models can't keep up with sport gender gap

Will women runners ever overtake men at the Olympics? Don't hold your breath.

Sir — Women sprinters may one day overtake men, according to A. J. Tatem and colleagues ("Momentous sprint at the 2156 Olympics?" *Nature* **431,** 525; 2004). As the holder of the world and European high-jump records for women over 50, I must say that this statistical prediction has been greeted with much laughter in athletic circles. The authors show linear regression lines and state that this model does not differ significantly from nonlinear approaches. I'm prepared to believe that. But much criticism is possible on both the data set and the logic behind the model.

The data set is very small: only one sprint result per Olympiad, with a large variation in results. Taking the mean of the best 10 per year provides 40 times more data, leads to much less deviation and clearly shows nonlinearity (see www.antenna.nl/weia/Progressie.html).

A logical critique goes like this: an athlete can improve greatly by training three times instead of twice a week and can improve further by adding a fourth training session, and so on — but each additional session will give less improvement than the one before. It follows that the sport as a whole will show a similar nonlinear improvement. When statistics, nevertheless, point to linear development, there must be something wrong. Most likely the 'linear' graph in fact consists of more nonlinear parts. For example, one part for the period when athletes were adding ever more training sessions, one part for when they reached a ceiling in adding sessions (around 1980), one part for when drug users were filtered out, and so on.

In which Olympiad will the form of the real nonlinear development become clear? I dare not guess.

Weia Reinboud
Simon Bolivarstraat 87, NL 3573 ZK Utrecht, the Netherlands

Sprint research runs into a credibility gap

Sir — A. J. Tatem and colleagues calculate that women may outsprint men by the middle of the twenty-second century (*Nature* **431,** 525; 2004). They omit to mention, however, that (according to their analysis) a far more interesting race should occur in about 2636, when times of less than zero seconds will be recorded.

In the intervening 600 years, the authors may wish to address the obvious challenges raised for both time-keeping and the teaching of basic statistics.

Kenneth Rice
MRC Biostatistics Unit, Institute of Public Health, Forvie Site, Robinson Way, Cambridge CB2 2SR, UK

Biology students find holes in gap study

Sir — We are students aged 16–18 in a Texas high school. Our biology teacher Vidya Rajan asked us to comment on the paper by A. J. Tatem and colleagues (*Nature* **431,** 525; 2004); we believe the projection on which it is based is riddled with flaws.

The idea of women running faster than men — although not novel (see B. J. Whipp and S. A. Ward *Nature* **355,** 25, 1992; and Correspondence *Nature* **356,** 21, 1992) — is interesting, but one cannot draw these conclusions based on generalization by extrapolation. Tatem *et al.* used a domain of 104 years to extrapolate to a domain of 252 years. It is not logical to say that the first 104 years will have data with exactly the same regression as the next 148 years. Using similar reasoning in 1992, Whipp and Ward suggested that women would run the marathon faster than men by 1998. This has still not happened.

In Tatem and colleagues' study, men were measured for 32 more years than women. This ignores the possibility that women might be reaching a plateau: had women's times been unexpectedly high before 1934, one could trace a decreasing rate of change for post-1934 Olympians.

Improvements due to the increase in numbers of women running are likely to level off as the rate of increase in participation slows down (see www.olympics.org.uk/olympicmovement/olympicissueswoman.asp).

Finally, both men and women may reach a physiological limit beyond which they cannot progress.

With these factors taken into consideration, the predictions made from the extrapolation seem less than sound.

Advanced Placement Biology Class
A&M Consolidated High School, College Station, Texas 77840, USA

Mind the gap: women racers are falling behind

Sir — Trend extrapolation can be an inexact science, especially in sport. A. J. Tatem and colleagues (*Nature* **431,** 525; 2004) suggest, counterintuitively, that a future woman may run faster than her male counterpart over 100 metres. It is worth noting that the 'fastest human on the planet' is usually the world-record holder for the 200 m, not the 100 m. For example, Michael Johnson's running velocity for his current world record 200 m in 19.32 s was 10.35 m s^{-1}, whereas Tim Montgomery's for his 100 m in 9.78 s was 10.22 m s^{-1}.

However, sports physiologist Stephen Seiler has analysed Olympic and world championship running results and found that the mean performance gender gap in the world records has actually increased from 10.4% in 1989 to 11.0% now (C. Holden *Science* **305,** 639–640; 2004). This held for seven of the eight events from 100 m upwards. The exception was Paula Radcliffe's marathon, which narrowed the gap from a relatively vulnerable 11.9% to 8.4%. Hence, in general, the gap has widened during the past 20 years.

Nevertheless, Tatem and colleagues make the very good point that only a minority of the world's population of women has been able to compete. Were China and India, with their vast populations, to come fully onstream in track and field sports, they could bring with them statistical outliers of both sexes who would demolish current records. But it is likely that there would still be a gender gap in the range of 7–10% in favour of the biologically advantaged men.

N. C. Craig Sharp
Sport Sciences, Brunel University West London, Uxbridge, Middlesex UB8 3PH, UK

A. J. Tatem and colleagues reply — We appreciate the interest generated by this light-hearted piece.

We were, of course, careful to caveat predictions with "if current trends continue". But if we were to follow all the advice we have received, we would be both correct and incorrect to fit a linear model. If we were incorrect, we should instead have fitted a two-part spline, a lowess curve, not a lowess curve, a rational function of polynomials, a quadratic model (predicting times regressing back to 1900 levels by 2100), a cubic model and an exponential curve. We should have both removed and added points, and were both correct and incorrect in our use of confidence intervals!

Will the gender gap continue shrinking? We look forward to finding out.