Lifespan and testosterone

SIR — Coronary heart disease and atherosclerosis are the most frequent causes of death among men. Because these conditions occur more frequently in men than in women, it is often assumed that androgens play a causative role. Indeed, correlations between circulating testosterone levels and coronary heart disease have been made, as have negative correlations between blood lipids and androgens, in particular between high density lipoprotein cholesterol and testosterone[1].

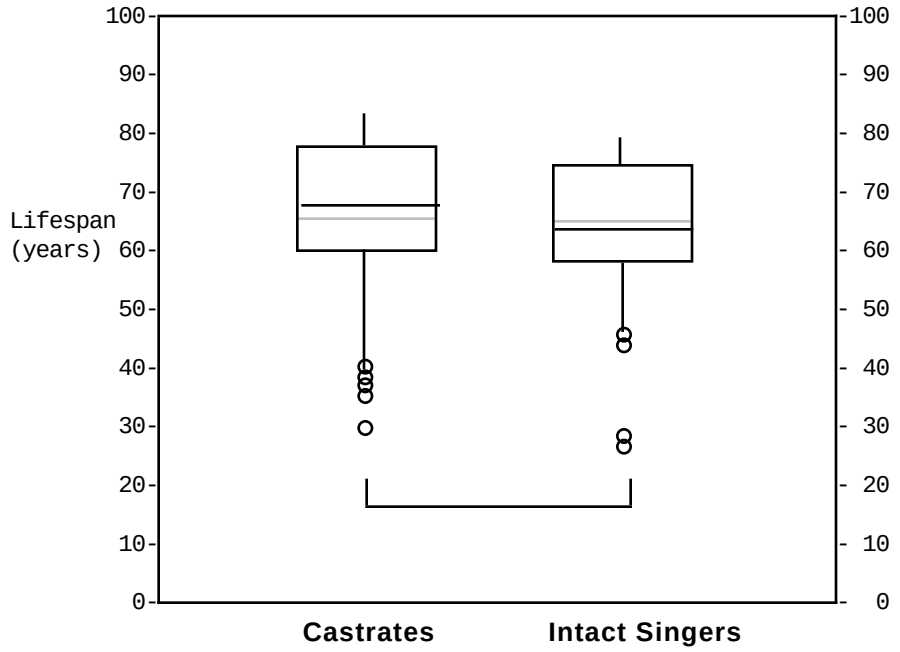
This hypothesis that testosterone may contribute to men's shorter lifespan is difficult to prove experimentally, not least because ethical considerations preclude many obvious tests. But from the sixteenth to the middle of the nineteenth century it was common practice in some European cultures to castrate prepubertal boys to prevent mutation of the voice and thus produce soprano and alto singers. Biographical data of these singers, when compared to those from a control group, should provide information on the influence of testosterone or — more generally speaking — the presence of testes on longevity.

From encyclopaedias and biographies[2–9], we have identified 50 castrates with outstanding reputations as singers born between 1581 and 1858. To establish a control group, we extracted from the same sources the names and dates of 200 'intact' male singers (bass, baritone or tenors) born during the same period and who had achieved comparable fame during their careers. We paired 50 of these singers with the castrates by selecting those with the most closely matched year of birth.

Over 277 years (1581-1858), there was no trend towards a change in lifespan in either group. The mean lifespan of the castrate singers was 65.5 ± 13.8 years (mean ± s.d.) and of the intact singers 64.3 ± 14.1 years, which is not significantly different (see figure).

These data show that prepubertal removal of the testes had no influence on the longevity of men. (italics by jh) But because people living at the time of our sample lived less long, on average than today. the possibility remains that androgens and other testicular hormones may be of importance during the extra years of life expectancy this century. The latter assumption receives some support from a study reporting a longer lifespan in castrated than in intact mentally disabled patients living in this century[10]. But it is difficult to draw conclusions for the general population from a study of institutionalized mentally retarded patients. The possibility remains that the difference in life expectancy between men and women may be related to another factor, probably Y-chromosome mediated, rather than to testosterone and the testes.

Eberhard Nieschlag, Susan Nieschlag, Hermann M Behre
Institute of Reproductive Medicine. Westfalian Wilhelm's University,
Steinfurter Str. 107, 48149 Münster, Germany



Lifespan of castrate and intact singers displayed as box plots (matched pairs, 50 in each group). The boxes encompass the 25th-75th percentiles of the data, with the median shown as the solid horizontal line and the mean as the broken line within the box. Whisker caps indicate the 10th and 90th percentile points of the data. Data outside that range are displayed individually. The groups were compared by unpaired t-tests and the differences were not significant (P= 0.65).

1. Kalin, M. F. & Zumoff, B. Steroids **50** 330-352 (1990).
2. Slominsky. N. (ed.) Baker's Biographical Dictionary of Musicians (5th edn, Schirmer, New York, 1958: 6th edn. Macmillan, London, 1978).
3. Blume, F, Die Musik in Geschichte und Gegenwart 4 vols (Bärenreiter, Kassel. 1949-51).
4. Brown J.D.. Biographical Dictionary of Musicians (Hindesheim, New York, 1970).
5. Dean, W. & Knapp J.M. Handel's Operas 1704-1726 (Clarendon. Oxford 1987).
6. Ortkemper, H Engel wider Willen. Die Welt der Kastraten (Henschel, Berlin, 1993).
7. Riemann,H. Musik-Lexikon 10th edn (Hesse, Berlin, 1922)
8. Sadie. S. (ed.) The New Grove Dictionary of Music and Musicians 1-20 (Macmillan, London. 1980).
9. Sadie, S. (ed.) The New Grove Dictictionary of Opera 1-4 (Macmillan, London 1992).
10. Hamilton, J.B. & Mestler,G.E. J. Geront. 24, 395-411 (1969).

Longevity and testosterone

SCIENTIFIC CORRESPONDENCE • NATURE • VOL 368 • 31 MARCH 1994 • page 408

SIR—Nieschlag et al[1] concluded that testosterone may not be responsible for the difference in lifespan now present between men and women. But their data do not appear to me to support this tsatement.

With the standard deviations reported in ref. 1 the statistical analysis would not have the power to detect a difference in lifespan even of 5 years due to the small sample number. But anyhow there is a more important objection. The 'castrati' singers were almost exclusively from Italy[2], as were many famous singers in that period (this was the age of the Bel Canto[3]). I therefore consulted an Italian life table. Because there are no accurate data for the seventeenth and the eighteenth century I examined a historical period in which the life expectancy was similar to that of the group under study for an individual who had reached the age of 30.

In the year 1891 the expected age of death of an Italian man who had reached age 30 was 64.7 years and that of a woman 65.0 years[4]. The life expectancy in 1891 was indeed almost equal between Italian men and women at all ages. Even if the criteria with which I have selected this population are open to argument, it is clear that we cannot assume that there was a difference in mortality between the sexes at the time of the castrati.

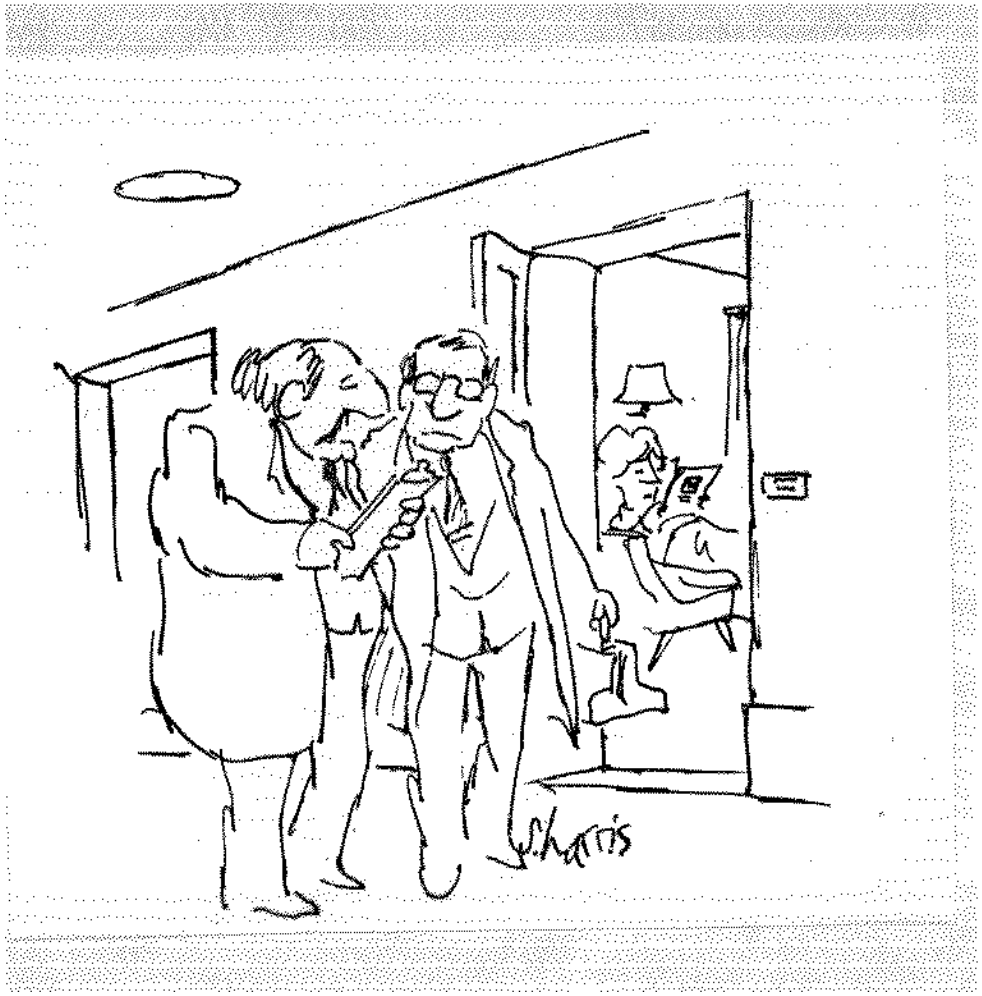
It is only since the First World War that increased female longevity is clearly present in all the developed countries[5]. There are many examples of small or inverted differences in the less developed countries even today[6] which cannot simply be explained by a high incidence of maternal death[5]. The most important cause of the difference, when present, is an increase in men of premature death from heart disease[7]. But in our 1891 Italian population, heart disease was listed as a cause of death only in 6% of the cases[8].

In conclusion, we cannot yet absolve testosterone. This is important not only for understanding the smaller incidence of coronary heart disease among women, but also because androgen hormones are sometimes taken for non- medical reasons (for example, enhancement of athletic performance) and their potentially damaging effects should not be underestimated.

Glovanal Paternostro
Department of Biochemistry, University of Oxford, Oxford OX1 3QU,

1 Nieschlag, E, Nieschlag, S. & Behre, H M. Nature 366, 215
2 Sadie, S. (ed) The New Grove Dictionary of Music and Musicians (Macmillan, London.
3 Scholes, P. A (ed.) The Oxford Companion to Music 10th edn (Oxford University Press,
4 Preston, S. H. Keyfitz, N & Schoen, R. Causes of Death: Life Tables for National Populations (Seminar Press, New York, 1972)
5 Stolnitz, G. J. Population Studies 10, No. 1. 17-42 (1956).
6 UN Statistiral Yearbook 1990/91 38th issue (New York 1993).
7 Johansson, S. Cardiovasc. Clin. 19 3-16 (1989).
8 Istituto Centrale di Statistica Sommario di Statistiche Storiche Italiane 1861-1955 (Roma, 1958).

THE NULL HYPOTHESIS



"Find out who set up this experiment. It seems that half of the patients were given a placebo, and the other half were given a different placebo

American Scientist 1982; 70:25.

Statistical scrotal effect

SCIENTIFIC CORRESPONDENCE • NATURE • VOL 368 • 31 MARCH 1994 • page 408

SIR—Nieschlag et al.[1] used historical data to examine the hypothesis that testosterone influences longevity in men. They examined mean life span in 50 castrati, famous as singers during the period 1581 – 1858, compared to the mean life span of 50 similar famous 'intact' singers born at the same times. The difference in mean life span (1.2 yr) was not significant in a two-tailed t-test (P=0.65), leading the authors to conclude: "The data show that prepubertal removal of the testes had no influence on the longevity of men." But the authors have fallen into a very common statistical trap—they conclude that a failure to demonstrate an effect implies that the effect does not exist, that an unrejected null hypothesis is true. A more reasonable conclusion from their data would have been that there is no evidence that the mean differences are not equal., which is not the same thing as concluding that the means are equal.

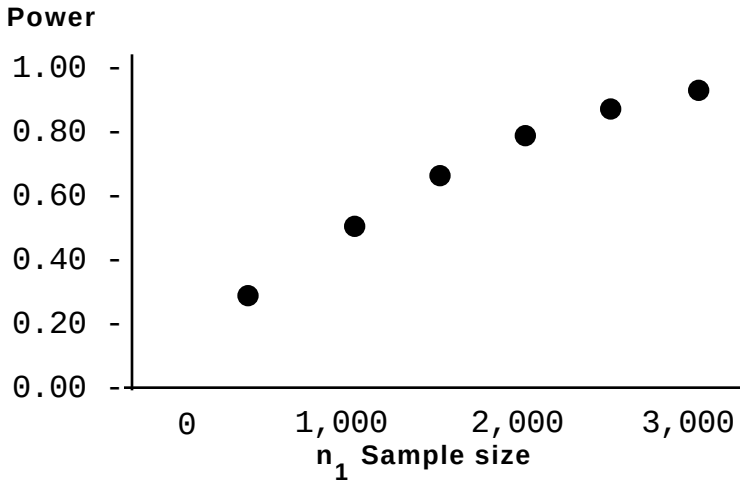
It is instructive to examine this data set from a power-analysis point of view. Let us assume, as the authors did, that (1) *age of death is normally distributed* † (this is almost certainly untrue); (2) the variances of age at death in castrati and intact singers are equal and estimable from the sample variances: and (3) the appropriate null hypothesis is that the mean age at death is the same (given the authors' preamble, a one-tailed hypothesis would have been more appropriate). Then. *if the real difference in age at death were 1.2 yr.* †† what is the probability (power) of detecting such a difference with a sample of size 50 in each group? Using a pooled variance of the ages at death, the probability is only 0.071 (see figure).

It is therefore not at all surprising that *the authors did not detect a significant effect* —they had almost no chance of doing so. lf the effect of castration were to change mean lifespan (in either direction) by 1.2 yr. then. given the observed variance structure. the authors would need a sample of size >2,120 (of each type of singer) to have an 80% chance of detecting such an effect with a maximal probability of 0.05 of making a type-1 error. Examined in another way – what is the smallest difference in mean age at death that could have been detected with this data set if the maximal probability of a type-l error was 0.05 and we wanted an 80% chance of detecting such a minimal difference if it existed? The answer is 9.1 yr. If the difference in longevity were that large. we probably wouldn't be engaged in statistical arguments about its valididity!

The best way, of course, to conclude that a null hypothesis is probably true is by replication; if repeated attempts to falsify the hypothesis always fail. then one begins to be increasingly convinced that the hypothesis is probably true. It is clearly ill advised to conclude the truth of a hypothesis based on a single sample. in the absence of a consideration of *what effect size was thought to be worth detecting,* or in the absence of a power analysis of *the effect size that was observed.* Readers wishing a gentle and informative introduction to power analysis should consult the wonderful paper by Peterman[2]; that seminal contribution should have what I term here the 'statistical scrotal effect"—it should cool the ardour of most hypothesis testers (testees?).

Stephen M Smith Dept Biology Univ. of Waterloo, Waterloo, Ont., Canada N2L3G1

1 Nieschlag, E . Nieschlag. S. & Behre H M. Nature 366 215
2 Peterman. R. M. Can. J Fish aquatic Sci. 47. 2-15 (1990)



Effect of sample size on power of a two-tailed t-test for independent means to reject a false null hypothesis if the difference in means = 1.2 yr (α = 0.05 ; s=13.95)

Commentary from jh in relation to some of the item italicized (by jh) in the text.

† It is a common fallacy to demand that data must be "normally distributed" in order for the t-test of means to be valid/accurate ... after all, the means of 65.5 and 64.3 in the original article were based on sample sizes of n=50 and n=50, surely big enough that they offset the skewness in the individual values i.e. the Central Limit Theorem should be considered here.

†† It is also a common pitfall to 'post-hoc' examine power in relation to the observed delta. What if --by chance -- the observed delta had been 0.0? That would have led to infinite sample size requirements!
The important point here is that the "delta of interest" is the "delta that would make a difference". First, it is a parameter value, not a statistic. Second, it is a matter of judgement, and not at all related to the study or the data at hand. Thus "a power analysis of *the effect size that was observed.* " makes no sense (the study produced an estimate of the true effect size. What would be important if it existed and what exists are two separate items.

See HANLEY JA: "Confidence limits vs power calculations" letter. Epidemiology , %(2) 264-266, 1994.