

Online Resource 1: Details and discussion of the literature review and the re-analysis of the

Damon & Thomas (1967) study

Article title: A curvilinear effect of height on reproductive success in human males

Journal: Behavioral Ecology and Sociobiology

Gert Stulp^{1,2},
Thomas V. Pollet¹,
Simon Verhulst²,
Abraham P. Buunk^{1,3},

1 Department of Social Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712

TS, Groningen, The Netherlands

2 Department of Behavioural Biology, University of Groningen, P.O. Box 11103, 9700 CC,

Groningen, The Netherlands

3 Royal Netherlands Academy of Arts and Sciences, The Netherlands

Author for correspondence:

Gert Stulp

g.stulp@rug.nl

Details and discussion of the literature review

We searched for studies on the relationship between male height and reproductive success using specific search terms (male, height, stature, reproductive success, number of children) in electronic databases (PubMed and WebOfScience) and by checking references of relevant papers. Only studies in which the number of live born children or the number of surviving children was used as a measure of reproductive success were included. Ideally, we would have carried out a meta-analysis but unfortunately too few studies reported the required estimates of effect size necessary to conduct such an analysis (in particular the effects of height squared were rarely tested or reported).

We identified seventeen scientific publications reporting the relationship between height and reproductive success measured as number of (living) children, of which one article studied two different populations (Kirchengast & Winkler 1994). Including the present study this brings the total to nineteen studies (Table 1). A variety of effects of male height on reproductive success were reported, including positive ($n=3$), negative ($n=2$), no ($n=8$) and curvilinear effects as in the present study ($n=6$). There can be different causes for this variation. One possibility is that there is true variation in the selection pressure among populations or over time, which is in itself not unusual. For instance, Siepielski et al. (2009) found considerable year to year variation in both strength and direction of selection for morphological traits, and this may also hold true for height.

Alternatively, but not mutually exclusive, conclusions may differ for methodological reasons, for example due to differences in sampling procedure, in the variables considered in the statistical analysis, or in the sample size and hence statistical power. As mentioned above, too few studies reported sufficient statistical details to allow for a meta-analysis to test these hypotheses. Instead we therefore discuss each of the possible methodological explanations for the variation in results among the eighteen studies.

With respect to the sampling procedure: A conspicuous difference among the studies is that not all studies were restricted to men who were at least close to have completed their reproductive careers, e.g. over fifty years in developed countries. Thus when the association between male height and reproductive success is mostly determined at a later age, than effects of height are difficult to detect when using a sample of younger men. Some studies used samples that were clearly not representative of the population (e.g. only healthy men, men from low socio-economic class, or ‘troubled boys’), but in most cases it is not clear to

what extent and in what way this would affect the results. An exception is the study of Mueller & Mazur (2001), who sampled men from the US military Academy at West Point with military careers of 20 years or more, and found a clear positive relationship between height and reproductive success. This sample is intentionally not representative of the whole population with respect to physical health and condition. More importantly, the selection procedure for this academy is likely to be stronger on tall men, because for biomechanical reasons it is more difficult for tall men to meet physical requirements of the military such as the minimum number of eight correct pull-ups and 54 push-ups in two minutes (Mueller & Mazur, 2001), and tall men that do meet those requirements may be exceptionally fit even compared to shorter men that meet the same requirements. Thus the discrepancy in results between the present study and the study of Mueller & Mazur (2001) may well be due to differences in sampling. Studies that used samples from non-western societies may be representative as such, but more difficult to compare with our results. However, also among non-western populations the results are mixed.

With respect to the statistical analysis we find that the studies vary in the variables controlled for when testing the effect of height, and in whether or not height squared was tested. Height is associated with both education and income, which are also associated with reproductive success. As education and income have opposite effects on reproductive success (negative and positive respectively), it is important to control for both of these measures instead of using a combined social status measure. Only two other studies controlled for education and (proxies of) income (Nettle 2002; Fielding et al. 2008). Note however that in our study it made little difference whether or not education and income were controlled for, because the effects of height were largely independent of education and income. Not controlling for these

parameters is therefore unlikely to have affected tests of curvilinear effects except that this would have slightly increased the statistical power.

Tests of non-linear effects were reported in only 5 studies (including the present study), but we cannot exclude the possibility that there were unreported non-significant results. Four out of these five studies did find non-linear effects, with Sear (2006) being the exception.

Although not reported in his article, Nettle (personal communication) also tested for non-linear effects, and did not find curvilinear effects. One possible reason for the discrepancy between the study from Nettle (2002) and our results, is that Nettle (2002) used a sample of men who might not yet have ended their reproductive careers (i.e. all men of 42 years of age). Furthermore, the average number of children and the variance was much lower in his sample compared to ours (1.81 ± 1.33 versus 2.54 ± 1.53 children), potentially making it more difficult to find an effect because of the lower variance. Mueller & Mazur (2001) did not test for non-linearity, but visual inspection of the data suggests this was also unnecessary as graphs clearly displayed a positive linear effect of height on number of children. This may be due to the biased nature of their sample as discussed above, and hence we consider it justified to ignore this result in this context. The importance of testing for non-linear effects becomes clear when considering the re-analyses of two studies in the table. Mitton (1975) re-analyzed the data of Clark & Spuhler (1959) and Damon & Thomas (1967) and found a curvilinear effect in both data sets. Mitton (1975) excluded single and married men without children from his re-analyses of Damon & Thomas (1967), potentially biasing the outcomes.

Therefore, we re-analysed the Damon & Thomas (1967) data, with the help of tables provided in Mitton (1975), Vetta (1975), and Damon & Thomas (1967), including single men and married men without children (for details see below). Using Poisson regression, we found a significant curvilinear effect of height on number of children. So after re-analysis we

find 6 out of 8 studies in which was tested for non-linearity, show a curvilinear effect of height on reproductive success.

Two studies mention curvilinear effects without actually testing for them. Goldstein and Kobylansky (1984, p.42) conclude 'According to our data, the peak of fertility tends to be related with modal parental morphological traits'. Similarly, Mueller et al. (1981) conclude on the basis of their data: '...non-linear associations of anthropometrics and fertility are more likely than directional selection.' (p. 315). In total, thus 8 studies appear to support a curvilinear relationship of male height on reproductive success from the 10 studies considering non-linear effects.

An additional methodological issue is the low statistical power for detecting an effect due to insufficient sample sizes. Selection gradients are typically low (Kingsolver et al. 2001), and therefore substantial samples are required to detect an effect. We calculated the N needed to detect the effect size from our study with a power of 0.8 and a p-level of 0.05, using G*Power 3 (Erdfelder et al. 1996). We used an effect size of $r=0.06$ which was taken from a linear regression of number of children on height and height² using the data of the present study (obviously this is a very conservative effect size, as studies with much lower samples sizes observed effects of height). Linear regression was used to determine the effect size rather than the Poisson regression applied in the present study, to facilitate comparison with the few studies that performed a regression analyses, because these studies used linear regressions exclusively. Given these parameters, an N of 2,680 was needed to obtain a power of 0.80 to detect a curvilinear effect of height. In addition to our study, Nettle (2002) is the only study with a sample size that exceeds this number. All other studies reporting null findings had much lower sample sizes, with the largest sample being 303 (Sear, 2006). With

this sample size, an effect size of $r=0.06$, and a p-value of 0.05, this study had a power of 0.14 to detect a curvilinear effect of height. Thus, all studies (except Nettle 2002) reporting no effect of height had a power equal or lower than 0.14. It is therefore not surprising that many studies did not observe selection on male height even when it was tested.

The re-analysis of the Damon & Thomas (1967) study

We re-analyzed the Damon and Thomas (1967) data, with the help of tables provided in Mitton (1975), Vetta (1975), and Damon and Thomas (1967), including single men and married men without children. On the basis of the means and standard deviations of these tables, we generated the data using random number generators. Depending on the underlying distribution of the variable to be generated, we used normal or Poisson random number generators. Using Poisson regression, we found a significant curvilinear effect of height on number of children (Table 5). In our re-analysis the height associated with the optimum number of children (177 cm) was close to the average height (173 cm). To compare the Poisson regression parameter estimates between those of our study (Table 3) and those of the re-analysis of the Damon and Thomas (1967) data, a test of the equality of was done using the formula $Z=(b_1-b_2)/\sqrt{(SEb_1^2 + SEb_2^2)}$ (Paternoster et al. 1998). Parameter estimates were not significantly different (Table 5).

Table 5. Comparison of the Poisson regression parameter estimates (\pm s.e.; p value in brackets) of the effects of height and height² on number of children ever born for our study and the re-analyses of Damon and Thomas (1967).

	Our study ^a	Damon & Thomas (1967)	Difference estimates (Z, p)
Intercept	$-1.76*10^1 \pm 5.63$ (0.001774)	-29.63 ± 9.65 (0.002136)	
Height	$2.09*10^{-1} \pm 6.29*10^{-2}$ (0.000897)	$3.39*10^{-1} \pm 1.11*10^{-1}$ (0.002312)	Z=-1.02, p=0.31
Height ²	$-5.89*10^{-4} \pm 1.76*10^{-4}$ (0.000813)	$-9.53*10^{-4} \pm 3.20*10^{-4}$ (0.002917)	Z=1.00, p=0.32
N	3,578	2,616	

^a We used the parameter estimates from the analyses on the number of children ever born, as Damon & Thomas (1967) also used this measure of reproductive success

^b The difference between the estimates of our study (Table 3) and the re-analysis is expressed in the z statistic using the formula $Z=(b_1-b_2)/\sqrt{(SEb_1^2 + SEb_2^2)}$ (Paternoster et al. 1998)

Additional references (not listed in the main document)

Erdfelder E, Faul F, Buchner A (1996) G*POWER: A general power analysis program. Behav Res Meth Ins C 28:1-11

Paternoster R, Brame R, Mazerolle P, Piquero A (1998) Using the Correct Statistical Test for the Equality of Regression Coefficients. Criminol 36: 859-866

Vetta A (1975) Fertility, Physique, and Intensity of Selection. Hum Biol 47:283-293