

Commentary on the article "improving the prediction of maturity from anthropometric variables using a maturity ratio"

Item Type	Journal article
Authors	Nevill, A;Burton, RF
DOI	10.1123/pes.2017-0201
Publisher	Human Kinetics
Journal	Pediatric Exercise Science
Download date	2026-03-17 06:39:36
License	https://creativecommons.org/licenses/by-nc-nd/4.0/
Link to Item	http://hdl.handle.net/2436/622404

1 **Commentary on the article “Improving the prediction of maturity from anthropometric**
 2 **variables using a maturity ratio”.**

3 Alan Nevill¹ and Richard F. Burton²

4 ¹Faculty of Education, Health and Wellbeing, University of Wolverhampton, Walsall, UK.

5 ²School of Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

6

7 We applaud the attempt of Fransen et al. (4) to improve on the original maturity offset article of
 8 Mirwald et al. (5). Both articles derive equations for predicting age at peak height velocity (APHV) but
 9 both are at best misleading and at worst fundamentally flawed. As their response variables, Mirwald
 10 et al. (5) used ‘maturity offset’ (CA-APHV) where CA is chronological age, and Fransen et al. (4) used
 11 ‘maturity ratio’ (CA/APHV). The problem is that their equations contain the subject’s CA in both sides
 12 of the prediction equations. In the statistical analyses, this will result in spuriously high values of R^2 .

13 For ease of reference, the equation given by Fransen et al. (4) is reported below. This, and
 14 subsequent equations, are re-written with fewer unnecessary decimal places/significant figures.

15 *Maturity ratio*

$$16 = 6.99 + (0.116 \cdot CA) + (0.00145 \cdot CA^2) + (0.00452 \times Body\ Mass) - (0.0000341 \times Body\ Mass^2) - (0.152$$

$$17 \times Stature) + (0.000933 \times Stature^2) - (0.00000166 \times Stature^3) + (0.0322 \times Leg\ Length) - (0.000269 \times$$

$$18 Leg\ Length^2) - (0.000761 \times [Stature \times CA])$$

19

20 The equation was derived for ages 11-16 years. Over this range the quadratic expression $\{(0.116 \cdot CA)$
 21 $+ (0.00145 \cdot CA^2)\}$ proves to be numerically equal, within 0.7%, to $\{0.154 \cdot CA - 0.242\}$. This fact might
 22 be used to make the above equation simpler and more “user-friendly”.

23

24 Pearson (10) and later Neyman (9) warned that spuriously high correlations will be found between
 25 some indices that have a common component. In both of the articles above (4, 5), the authors have
 26 made the mistake of regressing the maturity offset difference or ratio (CA-APHV or CA/APHV) with
 27 predictors that include CA (i.e., CA is common to both the response and predictor variables), a
 28 calculation that will always lead to a spuriously high correlation (2). To illustrate the inevitable danger

29 of this effect, consider the following example adapted from Nevill et al. (7). Table 1 includes two
 30 randomly generated normally distributed columns of data (population means \pm standard deviations
 31 being 10 ± 1), arbitrarily named *CA* and *APHV*.

32 Table 1 about here

33 Clearly, there is no significant correlation (or regression) between the two random variables ($r = -$
 34 0.060 ; $P = 0.85$). However, if we correlate the “maturity offset” difference ($CA - APHV$) with either the
 35 *CA* or *APHV* values, we obtain significant but spurious correlations $r = 0.757$ ($P = 0.004$) or with
 36 *APHV* $r = -0.697$ ($P = 0.012$). Similarly, if we correlate the maturity ratio ($CA/APHV$) with either *CA* or
 37 *APHV*, once again we obtain significant but spurious correlations, with *CA* $r = 0.782$ ($P = 0.003$), or
 38 with *APHV* $r = -0.666$ ($P = 0.018$). These significant correlations would lead to the erroneous
 39 conclusion that maturity-offset differences or maturity ratios are meaningfully and positively
 40 associated with *CA* or negatively associated with *APHV*.

41 Moore et al. (6) have proposed much simpler prediction equations than those of Mirwald et al. (5) and
 42 Fransen et al. (4), but have made the same mistake of including *CA* on both sides. Their equations
 43 are nevertheless worth looking at here in providing a simple object lesson as to how spuriously high
 44 R^2 will occur when a common variable appears on both sides of the regression equation.

45 For boys they have:

$$46 \quad CA - APHV = -8.1 + 0.0070 \times (CA \times \textit{sitting height})$$

47 For a representative sitting height of, say, 80 cm,

$$48 \quad CA - APHV = -8.1 + 0.0070 \times (CA \times 80) = -8.1 + 0.56.CA$$

49 For girls they have:

$$50 \quad CA - APHV = -7.7 + 0.0042 \times (CA \times \textit{height})$$

51 For a height of, say, 150 cm,

$$52 \quad CA - APHV = -7.7 + 0.0042 \times (CA \times 150) = -7.7 + 0.63.CA$$

53 In both cases *CA* makes a major contribution to both sides of the equation. Unsurprisingly, the values
 54 of R^2 for the full (original) equations are therefore high, namely 0.906 and 0.898 respectively. So high
 55 are these that the addition of other predictors to the equations was found to increase R^2 by less than
 56 1%. For either sex, ($CA-APHV$) is thus being predicted only from *CA* and one other variable, either
 57 sitting height of total height. Yet, as Mirwald et al. (5) illustrate, the ratio of leg length to sitting height

58 (and so also of height to sitting height) tends to be maximum when CA tends towards APHV. One
59 would therefore expect the equations to include two of the three variables height, sitting height and
60 leg length (but not all three, since height is the sum of the other two).

61 The equations of Fransen et al. (4) and Mirwald et al. (5) contain more terms. However, the
62 remarkably high values of R^2 associated with these must also be spurious—once again due to the
63 presence of CA on both sides of the equations.

64 Another major concern with the article of Fransen et al. (4) is that the authors are analysing repeated
65 measures data (that contains both between- and within-subject errors). Each subject has just one
66 APHV but a series of repeated observations over time where predictor variables such as leg length,
67 height and CA are repeatedly recorded over their growth cycle, that are incorporated as predictor
68 variables. These should be analysed using a multilevel modelling software approach that will
69 accommodate the hierarchical or nested observational units associated with these data, as
70 recommended and adopted by Baxter Jones et al. (1) and Nevill et al. (8).

71 Finally, the use of multiplicative allometric models (log-linear) rather than additive polynomials would
72 almost certainly improve the fit and overcome the obvious heteroscedastic errors (often referred to as
73 the shot-gun effect) seen clearly in Figures 3 and 4 (4). This approach was demonstrated to be
74 superior on several data sets associated with modelling the developmental changes in strength and
75 aerobic power in children (8). Note that the data structure reported in Nevill et al. (8) is hierarchical or
76 nested, very similar to the structure reported by Fransen et al. (4).

77 We see from the three papers (4-6), that estimating the APHV is practically very useful, and that a
78 valuable body of relevant data exists. However, the proposed equations, with their inflated values of
79 R^2 are likely to be misleading and may well be flawed. We believe that the predicted variable should
80 simply be AHPV. It is also desirable that any finally recommended formula should look simple enough
81 that people actually use it. It would also be good if it obviously reflected, or suggested, known
82 relationships amongst potential predictor variables. These might include the tendency of the ratio of
83 sitting height to total height to rise from a minimum at 12-15 years and for the Rohrer Index, (body
84 mass)/height³, to rise after ~12-16 years (3). The ratio of leg length to sitting height tends to be
85 highest when CA equals APHV (5).

86

87 **References**

- 88 1. Baxter-Jones A, Goldstein H, Helms P. The development of aerobic power in young athletes. *J*
89 *Appl Physiol* 1993, 75:1160-1167.
- 90 2. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against
91 standard method is misleading. *Lancet* 1995, 346, 8982:1085-1087.
- 92 3. Burton RF. Sitting height as a better predictor of body mass than total height and (body
93 mass)/(sitting height)³ as an index of build. *Ann Hum Biol* 2015, 42:210-214.
- 94 4. Franssen J, Bush S, Woodcock S, Novak A, Deprez D, Baxter-Jones ADG, Vaeyens R, Lenoir M.
95 Improving the prediction of maturity from anthropometric variables using a maturity ratio. *Ped*
96 *Exerc Sci* 2017, 30(1):xx-xx
- 97 5. Mirwald RL, Baxter-Jones ADG, Bailey DA, Beunen GP. An assessment of maturity from
98 anthropometric measurements. *Med Sci Sports Exerc* 2002, 34(4):689-694.
- 99 6. Moore SA, McKay HA, Macdonald H, Nettlefold L, Baxter-Jones ADG Cameron N, Brasher PMA.
100 Enhancing a somatic maturity prediction model. *Med Sci Sports Exerc* 2015, 47(8):1755-1764.
- 101 7. Nevill A, Holder R, Atkinson G, Copas J. The dangers of reporting spurious regression to the
102 mean. *J Sports Sci* 2004, 22(9):800-802.
- 103 8. Nevill AM, Holder RL, Baxter-Jones ADG, Round J, Jones DA. Modeling developmental changes
104 in strength and aerobic power in children. *Journal of Applied Physiology* 1998, 84:963-970.
- 105 9. Neyman J. *Lectures and Conferences on Mathematical Statistics and Probability*, 2nd ed. pp.
106 143-154. Washington DC: US Department of Agriculture.1952
- 107 10. Pearson K. Mathematical contributions to the theory of evolution.--on a form of spurious
108 correlation which may arise when indices are used in the measurement of organs. *Proc Roy Soc,*
109 *Lond* 1896;60(359-367):489-498.

110

111 Table 1 Two randomly generated normally distributed columns of data (population means \pm standard
 112 deviations being 10 ± 1), arbitrarily labelled *CA* and *APHV*. Also shown are values of *CA-APHV* and
 113 *CA/APHV*.

<i>CA</i>	<i>APHV</i>	<i>CA-APHV</i>	<i>CA/APHV</i>
11.9	9.7	2.2	1.23
11.7	10.7	1.0	1.09
9.5	11.7	-2.2	0.81
10.7	12.4	-1.7	0.86
8.3	10.5	-2.2	0.79
8.2	11.0	-2.7	0.75
9.6	9.5	0.1	1.01
9.9	10.7	-0.8	0.93
9.7	10.1	-0.4	0.96
9.8	10.1	-0.3	0.97
9.8	9.0	0.8	1.09
9.7	11.8	-2.1	0.82

114