Wills, A. K., Silverwood, R. J., & De Stavola, B. L. (2014). Comment on Tu et al. 2013. A critical evaluation of statistical approaches to examining the role of growth trajectories in the developmental origins of health and disease. International Journal of Epidemiology, 1662-1664. 10.1093/ije/dyu032

FOR PUBLICATION

Comment on Tu et al. 2013. A critical evaluation of statistical approaches to examining the role of growth trajectories in the developmental origins of health and disease.

From Andrew K Wills, Richard J Silverwood, Bianca L De Stavola

Tu et al. (1) give a useful overview of several common approaches used to examine growth trajectories in relation to a later health outcome. One approach covered is the use of general linear models where the later health outcome is expressed as a function of several repeated measures of size (all or a subset up to a critical time), or of linear functions of size. In their paper, these are described as the life course plot model, change scores models 1 and 2, and the conditional (size) model. The authors highlight that different conclusions are drawn depending on the model specification. As an example, they show that the model specified in terms of size measures (the life course plot) finds negative associations of later SBP with early life weight, whereas the model containing birth weight and subsequent changes in weight (change scores model 1) finds positive associations with birth weight and subsequent changes in weight. We would like to provide some clarification and expansion on the interpretation of these models.

First, it is worth noting that there is nothing contradictory about these results. While the analyses based on the life course plot, conditional size and change scores models are all re-parameterisations of the same underlying model (see appendix for their algebraic formulation and re-expression), the coefficients from each model are conditional on different sets of transformed variables and therefore address different questions. Hence we should indeed expect to get different answers.

As stated elsewhere, the interpretation of each coefficient in a model containing repeated measures of the same dimension at different ages can be difficult because of the conditioning (2). Generally, the interpretation of each coefficient in Table 3 of Tu et al. (1) takes the form of: "the expected difference in SBP for a 1SD greater weight at age *x* (or a 1SD greater increase in weight from age *x-1 to x* in the change scores models), conditional on all other sizes (or changes in size) in the model being equal". In some sense, this means that each coefficient now contrasts a trajectory rather than a measure at a single point in time. Understanding this trajectory contrast is important to understanding the results from each of the model re-parameterisations presented in Tu et al. (1) and reported elsewhere e.g.; (3-6).

Perhaps a useful visualisation of these model specifications is presented in figure 1. Using the example dataset from Tu et al.(1) , it illustrates the trajectory contrast tested by the coefficient for standardized weight at 8y in the life course model and conditional size model, and by the coefficient for change in standardized weight from 2-8y in each of the change scores models. The plot shows that in the life course model, the 8y coefficient represents the mean difference in SBP corresponding to a child whose trajectory includes a 1SD higher weight at age 8 and the same weight at all other ages. This model targets the question, what is the effect of carrying extra weight at only this one period in life? The corresponding coefficient in change scores model 1 represents the effect of a 1SD greater weight gain from 2 to 8y with the same birth weight and the same weight changes in all intervals other than 2 to 8y. This implies a 1SD difference in all future weights and hence asks a cumulative question: what is the effect of gaining extra weight in a particular interval and maintaining that weight differential in the future? The coefficient for 2 to 8y in change scores model

2 also represents the effect of a 1SD greater weight gain from 2 to 8y but compares against children with the same future weight. This implies a 1SD lower weight at earlier ages and asks the question: given the same future weight trajectory, does starting smaller with subsequent greater weight gain matter? The parameter in the conditional size model corresponding to the conditional score at age 8y is the same as that of a model that conditions only on past weight, hence the future is unspecified and the contrast in figure 1 ends at 8y. The plot shows that this coefficient represents the effect of a 1SD greater weight gain from 2 to 8y among those with the same earlier weight. As implied, this asks a prospective question at each age: what is the effect of current weight (or recent weight gain), independent of the past? Using path analysis terminology, the coefficients for conditional size represent the *total effect* of that particular variable on the outcome. i.e. the sum of its *direct* and *indirect effects* (through future weights). This shows the connection between path analysis and the conditional size and life course models. For example, a path analysis model where each measure of weight is allowed to influence the outcome both directly and indirectly through all future weights is expressed using the equation for the life course plot model (equation 1 in the appendix) and the equations to calculate the conditional scores (equations 4 in the appendix), no new information is added and no information is removed.

Returning to the 'contradictory' conclusions highlighted by Tu et al. (1). The suggestion from their life course plot (see their figure 4 and Table 3 (1)) is that being small at birth and in early life (although weak associations), and having a higher current weight are associated with a higher SBP. Change scores model 1 on the other hand, suggests that being heavy or gaining weight at any stage from birth is linked to a higher SBP. These are different contrasts, not contradictory findings. The parameters in the first model are comparisons against children with the same size at all other ages, whereas those in the second model compare against children who remain smaller at all future ages. Referring to the estimated age-specific coefficients from change scores model 1 (see Table 3 in Tu et al. (1)) also shows that those who were smaller at birth, stayed smaller and experienced weight gain later in life tended to have a higher SBP than those of the same current weight who were born bigger (obtained by performing the calculation from their table: -1SD*β(zwt_0)+1SD*β($zwt_{19\text{-}15}$)= -1*2.58+1*4.24 =+1.66mmHg), this is also the contrast made by the coefficient for $zwt_{19\text{-}15}$ in change scores model 2 and so we get the same result. As mentioned, the coefficients from each model can be re-expressed as coefficients from the other models (see appendix) and in doing so we then just force them to make the same contrast and hence ask the same question. Note also that the life course model, change scores model 1 and conditional change model have the same estimated regression coefficient for the last weight measure (i.e. standardized weight at 19y or the change in weight between 15 and 19y). This is no coincidence but algebraic equivalence (2)(see appendix), highlighting the interpretation to be given to the most proximal weight measure, i.e. conditional on past values. This shows that if the coefficients are interpreted with respect to the conditioning in the model so that the contrast and question posed by each is explicitly referenced, then seemingly contradictory conclusions about the role of life course weight can be easily reconciled.

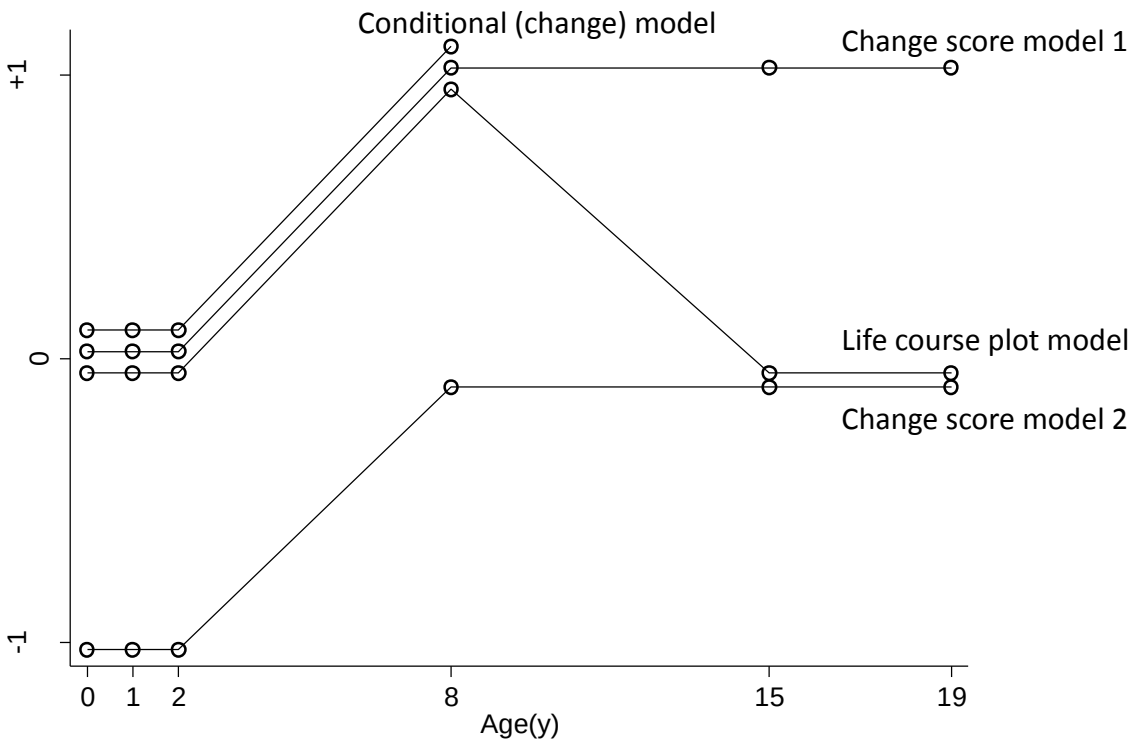

Figure 1. An illustration of the trajectory contrast made by the coefficient for weight at 8y in the life course model and conditional (size) model, and the coefficient for weight change from 2 to 8y in the change scores models as presented in Tu et al (1). The lines represent the difference in weight (z-score) at each age versus the reference trajectory. The thin separation between trajectories is done for clarity; in reality they should be superimposed.

Appendix

Using the same model terminology as Tu et al. (1), generally for a set of *k* standardised repeated measures of size (x), the life course plot model can be written as:

$$E(Y) = \beta_0 + \sum_{j=1}^{k} \beta_j x_j \qquad [1]$$

Change score model 1 (conditioned on initial size x_1) as:

$$E(Y) = \gamma_0 + \gamma_1 x_1 + \sum_{j=2}^{k} \gamma_j (x_j - x_{j-1}) \qquad [2]$$

Change score model 2 (conditioned on final size x_k) as:

$$E(Y) = \delta_0 + \sum_{j=1}^{k-1} \delta_j (x_{j+1} - x_j) + \delta_k x_k \qquad [3]$$

The Conditional size scores are the residuals ε_2 *to* ε_k estimated from the following models:

$$x_2 = \zeta_{2,0} + \zeta_{2,1} x_1 + \varepsilon_2$$

$$x_3 = \zeta_{3,0} + \zeta_{3,1} x_1 + \zeta_{3,2} x_2 + \varepsilon_3$$

…

$$x_k = \zeta_{k,0} + \sum_{j=1}^{k-1} \zeta_{k,j} x_j + \varepsilon_k \qquad [4]$$

where ε_j is the conditional size or growth score for period *j-1* to *j*. The analysis model used to estimate the conditional size effects is then:

$$E(Y) = \eta_0 + \eta_1 x_1 + \sum_{j=2}^{k} \eta_j \varepsilon_j \qquad [5]$$

Coefficients from change model 1 (equation 2) can be re-expressed in terms of coefficients from the life course plot model (equation 1) as follows:

$$\gamma_j = \sum_{i=j}^{k} \beta_i \qquad \text{for all } j \geq 1 \qquad [6]$$

Likewise, coefficients from change model 2 (equation 3) can be re-expressed in terms of coefficients from the life course plot model (equation 1):

$$\delta_j = -\sum_{i=1}^{j} \beta_i \qquad \text{where } 0<j<k \qquad [7]$$

$$\delta_j = \sum_{i=1}^{k} \beta_i \qquad \text{where } j=k$$

Coefficients from the life course plot (equation 1) can also be re-expressed in terms of those from the models to estimate the conditional size effects (equations 4 and 5), as follows:

$$\beta_j = \eta_j - \sum_{i=j+1}^{k} \zeta_{i,j} \eta_i \qquad \text{for all } j \geq 1 \qquad [8]$$

Equation [8] is probably best understood by considering the equivalence of the *conditional size* coefficients (η_j's in equation 5) with the *total effects* from a saturated path analysis model containing repeated measures – by saturated we mean that the outcome is conditioned on all available repeated measures and each measure is allowed to influence the outcome directly, and indirectly through all future measures. For example, the saturated path model for serial measures of standardised weight in the example dataset of Tu et al., (1) would be expressed using equations [1] (life course plot model) and [4] (equations used to calculate the conditional size scores), where the number of repeated measures *k,* is 6. In path analysis jargon, the *total effect* of weight at age *j* is the sum of the *direct effect* at age *j* and all *indirect effects* through future weights. The *direct effects* from a path analysis model are equal to the β_j's in the life course model (equation 1), the *indirect effects* are equal to the sum of the paths emanating from weight at age *j* to the outcome via intermediate variables, in this case subsequent weight measures. Here the effect of each path is estimated by the product of the coefficients along that path. Therefore the *direct effect* at age *j* (β_j) can be re-expressed as the *total effect* at age *j* (η_j) minus the sum of indirect effects from age *j* ($\sum_{i=j+1}^{k} \zeta_{i,j}\, \eta_i$), which gives equation [8]. Note though that this decomposition in path analysis is appropriate only if there are no unmeasured confounders between each growth measure and the outcome, no interactions among the growth measures and if all variables are continuous (7)

References

1. Tu YK, Tilling K, Sterne JA, Gilthorpe MS. A critical evaluation of statistical approaches to examining the role of growth trajectories in the developmental origins of health and disease. Int J Epidemiol. 2013 Oct;42(5):1327-39. PubMed PMID: 24038715.
2. De Stavola BL, Nitsch D, dos Santos Silva I, McCormack V, Hardy R, Mann V, et al. Statistical issues in life course epidemiology. American journal of epidemiology. 2006 Jan 1;163(1):84-96. PubMed PMID: 16306313.
3. Adair LS, Fall CH, Osmond C, Stein AD, Martorell R, Ramirez-Zea M, et al. Associations of linear growth and relative weight gain during early life with adult health and human capital in countries of low and middle income: findings from five birth cohort studies. Lancet. 2013 Aug 10;382(9891):525-34. PubMed PMID: 23541370. Pubmed Central PMCID: PMC3744751. Epub 2013/04/02. eng.
4. Cole TJ. Modeling postnatal exposures and their interactions with birth size. The Journal of nutrition. 2004 Jan;134(1):201-4. PubMed PMID: 14704319. Epub 2004/01/06. eng.
5. Gamborg M, Andersen PK, Baker JL, Budtz-Jorgensen E, Jorgensen T, Jensen G, et al. Life course path analysis of birth weight, childhood growth, and adult systolic blood pressure. American journal of epidemiology. 2009 May 15;169(10):1167-78. PubMed PMID: 19357327. Pubmed Central PMCID: 2732973.
6. Skidmore PM, Hardy RJ, Kuh DJ, Langenberg C, Wadsworth ME. Life course body size and lipid levels at 53 years in a British birth cohort. Journal of epidemiology and community health. 2007 Mar;61(3):215-20. PubMed PMID: 17325398. Pubmed Central PMCID: 2652912.
7. MacKinnon D. Introduction to Statistical Mediation Analysis. New York: Taylor & Francis; 2008.