Received: 14 March 2017 | Revised: 22 August 2017 | Accepted: 18 October 2017

DOI: 10.1002/ajhb.23075

ORIGINAL RESEARCH ARTICLE

WILEY American Journal of Human Biology

Accounting for measurement error in human life history trade-offs using structural equation modeling

Samuli Helle[1,2]

[1] Section of Ecology, Department of Biology, University of Turku, 20014, Finland

[2] Department of Biology, University of Turku, 20014, Finland

Correspondence
Samuli Helle, Section of Ecology, Department of Biology, University of Turku, 20014 Turku, Finland.
Email: sayrhe@utu.fi

Funding information
Koneen Säätiö, Grant/Award Numbers: 085253 and 086809

Abstract

Objectives: Revealing causal effects from correlative data is very challenging and a contemporary problem in human life history research owing to the lack of experimental approach. Problems with causal inference arising from measurement error in independent variables, whether related either to inaccurate measurement technique or validity of measurements, seem not well-known in this field. The aim of this study is to show how structural equation modeling (SEM) with latent variables can be applied to account for measurement error in independent variables when the researcher has recorded several indicators of a hypothesized latent construct.

Methods: As a simple example of this approach, measurement error in lifetime allocation of resources to reproduction in Finnish preindustrial women is modelled in the context of the survival cost of reproduction. In humans, lifetime energetic resources allocated in reproduction are almost impossible to quantify with precision and, thus, typically used measures of lifetime reproductive effort (e.g., lifetime reproductive success and parity) are likely to be plagued by measurement error. These results are contrasted with those obtained from a traditional regression approach where the single best proxy of lifetime reproductive effort available in the data is used for inference.

Results: As expected, the inability to account for measurement error in women's lifetime reproductive effort resulted in the underestimation of its underlying effect size on post-reproductive survival.

Conclusions: This article emphasizes the advantages that the SEM framework can provide in handling measurement error via multiple-indicator latent variables in human life history studies.

1 | INTRODUCTION

It is a commonly held view in evolutionary biology that causal inference is tied to experimental approach and that correlative data cannot be used to draw causal conclusions (Roff, 2002; Stearns, 1992). This problem concerns particularly human life history studies, because the manipulation of life history traits in humans is not feasible owing to ethical reasons. Most researchers in our field recognize that such inferential problems are due to the nonrandom selection of study subjects or to missing confounding variables not included in the analysis (Gagnon et al., 2009; Sear, 2007). But few have seemed to recognize that measurement error in independent variables, defined as the difference between a value measured and the true value of scientific interest, has also detrimental consequences on causal inference (Antonakis, Bendahan, Jacquart, & Lalive, 2010; Antonakis, Bendahan, Jacquart, & Lalive, 2014; Pearl, 2009). Measurement error, non-random selection, and omitted variables all undermine causal inference because they all introduce a correlation between independent variables and the model errors (e.g., of measurement error, please see the Appendix), thus violating a key assumption for any regression modeling (Antonakis et al., 2010, 2014; Pearl, 2009). Of these, the problem of missing variables is, however, the most severe in human evolutionary research as it is almost impossible to include all

variables, even the theoretically important environmental- and individual-level variables, into the same analysis.

There are several processes by which measurement error can contaminate the true scores of the target construct and hence reduce its reliability (i.e., the ratio of variance explained by true scores to the total variance). For example, the measurement instrument may not be precise enough or it may be sensitive to ambient environmental conditions or interobserver variability, thus introducing measurement error to the underlying construct of interest. Potentially far more important, but rarely considered thoroughly in evolutionary studies in general, is the validity of the measurements used to represent the constructs of interest (Houle, Pelabon, Wagner, & Hansen, 2011). Low validity means that the researcher has not directly measured (unintentionally or not) the biological construct she/he is interested in based on the underlying theory. Instead, the researcher has recorded a proxy or proxies of the construct that are then used in regression modeling as they would be perfectly reliable measures of these constructs.

In bivariate associations measurement error in an independent variable skews its regression or correlation coefficient toward zero and the standard errors of the point estimate are overestimated (Antonakis et al., 2010, 2014). Moreover, partial regression coefficients become causally invalid in an unpredictable manner and the overall statistical power decreases when the model structure and complexity increases due to the inclusion of intercorrelated predictors suffering from measurement error and when product terms of these variables (i.e., interactions and higher-order polynomials) are modelled (Bollen, 1989; Cole & Preacher, 2014; Freckleton, 2011; Pugesek & Tomer, 1995). Despite the detrimental consequences on the interpretation of regression coefficients in correlative data, measurement error in independent variables has received yet surprisingly little attention in life history studies (but see Adolph & Hardin, 2007; Freckleton, 2011; Kendall, 2015; Pugesek & Tomer, 1995). For example, Pugesek and Tomer (1995) clearly showed by simulations in the context of natural selection that even a small amount of measurement error in phenotypic traits, whether involving the trait of main focus or other traits in the model, will severely bias our conclusions on the magnitude of phenotypic selection. It is important to note that measurement error in the response variable is absorbed into the model error term and thus does not make the unstandardized regression estimates inconsistent, although it underestimates the amount of explained variance and reduces statistical power (Antonakis et al., 2010; Shipley, 2000).

But why should we worry about measurement error in studies of human life history evolution? One of the most studied questions of human life history evolution has been whether women, who invest more direct energetic resources in reproduction than men, sacrifice their longevity for higher reproductive success. Such a link is the core of the evolutionary explanations of senescence (Kirkwood & Rose, 1991; Williams, 1957). In particular, the evolution of senescence based on disposable soma theory states that it is the amount of limited resources invested in reproduction, or reproductive effort, during the lifetime in relation to the environmental resources available that matters in terms of old-age survival (Kirkwood & Rose 1991). But so far, evidence for the predicted survival costs of reproduction are surprisingly scarce in humans (reviewed in Gagnon, 2015; Helle, Lummaa, & Jokela, 2005; Hurt, Ronsmans, & Thomas, 2006; Jasienska, 2009; Le Bourg, 2007) and the reasons why we do not generally see such costs in most of the human populations studied remain elusive.

Why should studies that measure lifetime reproductive effort in long-lived vertebrates with parental care like in humans mind measurement error? In ideal settings, one should attempt to measure women's physiological energy expenditure during every reproductive attempt proportional to the overall energy available over their whole reproductive lifespan to estimate the total amount of resources allocated in reproduction (Jasienska, 2009). Obviously, this is impossible in retrospective demographic studies, and likely almost impossible even in prospective studies, owing to the high amount of time and money needed to conduct such a study. Traditionally, lifetime reproductive effort in humans has been measured as the total number of offspring born or surviving to adulthood (Gagnon, 2015; Helle et al., 2005; Hurt et al., 2006; Jasienska, 2009; Le Bourg, 2007), and in some cases additionally as brood size (e.g., Helle, Lummaa, & Jokela, 2004) or skewed offspring sex ratio (e.g., Helle & Lummaa, 2013; Helle, Lummaa, & Jokela, 2002). While such demographic variables are associated with the energetic resources allocated in lifetime reproduction, they are unlikely to perfectly reflect the total lifetime reproductive effort, even in concert. That is, aside from the potential measurement error arising from the direct measurement of the life history traits in question, variables measuring merely reproductive output likely miss parental energetic resources allocated in offspring, for example, in terms of lactation and care during childhood. In other words, even though women's lifetime number of births could be counted without error, measuring their lifetime reproductive effort in the currency predicted by the theory (i.e., as energetic resources) probably cannot be done. As measures of post-natal allocation in offspring are commonly lacking from demographic data sets in humans (Jasienska, 2009), the discrepancy between the estimated and realized lifetime reproductive effort can thus be severe.

In this article, measurement error in female lifetime reproductive effort is addressed by applying structural equation modeling (SEM) with latent variables (Bollen, 1989; Kline, 2015). Latent variables are existing constructs but unmeasured using the current data at hand. SEM can handle measurement

TABLE 1 Descriptive statistics of variables used in this study ($n = 664$)

Variable	Mean	s.d.	Min	Max	Proportion
Survival beyond age of 50 year					
0 = did not survive					22.6%
1 = survived					77.4%
Number of offspring surviving	4.04	2.22	0	13	
Mean interbirth interval (yrs)	2.09	1.08	0	7.5	
Age at last reproduction (yrs)	39.3	6.06	14	49	

error in independent (and dependent) variables when the researcher has recorded several indicator variables of a hypothesized latent construct or has knowledge of the amount of measurement error in the variables of interest. This is accomplished by partitioning the overall variance of a latent variable to the "true variance" and to the "error variance," containing errors in measurement. Because the estimation of the amount of error variance without clear guidance (e.g., unavailability of repeated measures or test-retest approach) may be a very error prone task, the utility of SEM to handle measurement error with multiple-indicator latent variables is highlighted here. A comprehensive treatment of measurement error in covariates in statistics is given, for example, in Carroll et al. (2006) and in Buonaccorsi (2010).

As a simple example of the SEM approach, we consider here measurement error in lifetime reproductive effort in women, measured by three female reproductive life history traits. In order to demonstrate the effect of measurement error in predictors on the interpretation of regression estimates, we relate this latent lifetime reproductive effort to an individual's probability of surviving beyond menopause. The resulting regression estimate is then contrasted with the estimate obtained from a traditional regression approach where the single best proxy of lifetime reproductive effort available in the data is used to make biological inference.

2 | MATERIAL AND METHODS

2.1 | Demographic data

We used demographic data collected from historical Finnish parish registers kept by the Lutheran church that allow individuals to be followed from birth, through their reproductive history and adult life, to death (Luther, 1993). These data, based on family reconstructions, were collected from four parishes, namely Enontekiö, Inari, Utsjoki, and Sodankylä, located in northern Finland (Helle et al., 2014). In this area, the systematic collection of birth records started in the 1730s in Sodankylä and in the 1750s in Enontekiö, Inari, and Utsjoki. That is, the records include all individuals born in those parishes since then, and their parents and other, potentially non-kin family members born before. These data consist of two distinct ethnic groups: the indigenous Sami and the settled Finns, but mixed marriages between these groups were rare during the study period (Enbuske, 2008). Sami practised mainly hunting, fishing and reindeer herding but also small-scale agriculture whereas settled Finns relied mainly on agriculture but also on traditional subsistence (Enbuske, 2008). Owing to the historical nature of the data, it contains individuals who lacked modern birth-control methods and advancements of modern medical care.

From these data, we recorded information on three demographic life history traits that should reflect the amount of women's lifetime reproductive effort: offspring number surviving to adulthood (age 18) (OFS), mean inter-birth interval (MIBI), and age at last reproduction (ALR; see Table 1 for descriptive statistics). We also recorded women's lifespans to determine whether a woman survived to the post-menopausal period, defined here as the age of 50 years, or not. Note that the cut-off value of the age of 50 years has been commonly used in studies of human reproduction-longevity trade-offs (Gagnon, 2015; Helle et al., 2005; Hurt et al., 2006; Jasienska, 2009; Le Bourg, 2007; but see Helle & Lummaa, 2013). The original sample size was 3,054 women. However, owing to the inclusion of women who married only once and to the missing values of their life history traits and birth year (i.e., listwise deletion was used), the final sample size was 664 women. Please note that the variables used here are selected for illustrative purposes only, and this example is not intended to represent a comprehensive life history analysis of the survival costs of reproduction in human females. A far more complete treatment of the subject will be given elsewhere (Helle, unpublished manuscript).

2.2 | Structural equation modeling

SEM with multiple-indicator latent variables was used to model measurement error in women's lifetime reproductive effort (Bollen, 1989; Kline, 2015). This model assumed an unobserved theoretical construct, labeled as women's

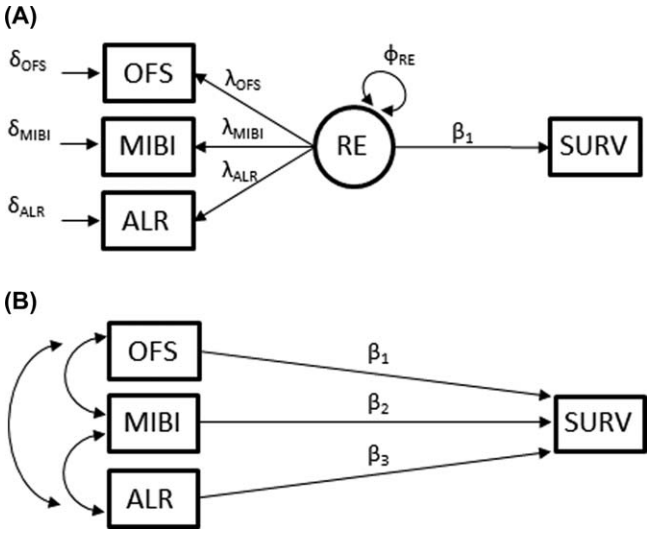

FIGURE 1 A graphical representation of the SEM with multiple-indicator latent variable (A) and the multiple regression model (B). In SEM, women's lifetime reproductive effort is measured by OFS, MIBI, and ALR, which are modeled as effect indicators (i.e., their variation is caused by the latent variable). Observed variables are represented as boxes and unobserved latent variables as circles. Single-headed arrows are used for three purposes: (i) when pointing from a latent to observed indicators, they represent factor loadings (λ's); (ii) when pointing at observed or unobserved response variables, they represent structural path coefficients (β's) and; (iii) when pointing at nondiscrete indicators of latent variables, they represent their residual variances (δ). Double-headed arrows represent the variances of latent variables (ϕ) or covariances among independent variables. Please note that parameters for the location of variables (intercepts and thresholds for continuous and discrete variables, respectively) are omitted for simplicity

lifetime reproductive effort. This continuous-scale latent variable was assumed to cause variation in multiple reproduction-related variables observed by the researcher, but that by themselves (either separately or jointly) cannot be taken to represent error-free measures of women's lifetime reproductive effort (Kline, 2015). In other words, the reproduction-related life history traits were assumed to result from an individual's lifetime reproductive effort and, thus, to have some shared variance (termed as common variance). Hence, the life history traits are called effect indicators, caused by the latent variable, and they are modelled as response variables in structural equations (Bollen, 1989). Because of the theoretical expectation that effect indicators are caused by the underlying latent construct, the indicators should show positive and at least moderately high intercorrelations (Kline, 2015). The error variances of such latent variables represent the amount of unobserved omitted causes contributing to the variance of those latents (Bollen, 1989). Based on the indicators included in the analysis, SEM thus partitions the overall variance of a latent variable into its underlying "true variance" and into nuisance "error variance" that includes errors in measurement. Residual variances of the effect indicators, which are also considered as latent variables, represent the amount of variance specific to that indicator not accounted for by the latent variable (termed as unique variance). These residual variances are by default assumed to be uncorrelated among the indicators, but that assumption can be relaxed if theoretically justified (Bollen, 1989; Kline, 2015). If only one latent variable is estimated in the model, one needs at least three freely estimated indicators in order to have an identified model (Kline, 2015).

How the observed effect indicators are associated with the latent variables are called "loadings" and they are estimated by the means of linear regression from a latent variable to its indicators (Bollen, 1989; Kline, 2015). Therefore, these loadings tell how much the values of indicators change when the latent variable changes one unit. If the indicators are measured on different scales, variance-standardized solutions are usually also reported (Kline, 2015). Because latent variables are arbitrary constructs, they have no inherent metric and usually the most important effect indicator (i.e., a marker indicator) is set to have a fixed unstandardized loading of 1 to set the scale for the latent variable (Hayduk & Littvay, 2012). Another option is to fix the latent variance to unity, so that all the unstandardized loadings of effect indicators can be estimated. Anatically, these two approaches produce identical results and model fit (Bollen, 1989).

The equations for a model with one latent variable with, let's say, three indicators are

$$\gamma_1 = \alpha_1 + \lambda_1\eta_1 + \varepsilon_1$$
$$\gamma_2 = \alpha_2 + \lambda_2\eta_1 + \varepsilon_2$$
$$\gamma_3 = \alpha_3 + \lambda_3\eta_1 + \varepsilon_3$$

where γ's are observed variables, or effect indicators, that depend on the latent variable η_1, and α's are intercepts for the indicators. The parameters λ index the association between the indicators and the latent variable, and ε's represent the unique variances of indicators that are assumed to have zero covariance (Kline, 2015).

The SEM used to examine the association between women's lifetime reproductive effort and its influence on whether they survived to the post-menopausal period is shown in Figure 1A. The SEM has two parts: the measurement model part that describes how the observed effect indicators load onto a latent variable and the structural model part that describes the assumed causal links between the variables of main interest, whether latents or observed. In SEM, these two are combined with simultaneous equations to accomplish unbiased estimation of structural parameters while accounting for the influence of measurement uncertainty in constructs by the measurement model (Bollen, 1989; Kline, 2015). It is important to realize that in ecological and evolutionary studies we are usually interested in structural parameters. That is, using the current example, the researcher is now interested in how the latent

TABLE 2 Correlation matrix of the life-history traits used as effect indicators

	1	2	3
Number of offspring surviving (1)	1.000		
Mean interbirth interval (2)	0.446	1.000	
Age at last reproduction (3)	0.406	0.473	1.000

variable measuring lifetime reproductive effort in women is related to their survival probability to the postmenopausal period and not how the individual indicators of lifetime reproductive effort are related to the outcome as in a regular multiple regression approach (Figure 1).

The effect indicators reflecting women's lifetime reproductive effort, modelled as a continuous latent variable, were the number of offspring surviving to adulthood (OFS), mean interbirth interval (MIBI), and ALR (Figure 1A). The rationale for using the number of offspring surviving to adulthood as an effect indicator of women's lifetime reproductive success is rather self-explanatory since it is widely used as the best single proxy of their lifetime reproductive effort. Hence, it was used as a marker indicator to set the scale for the latent lifetime reproductive effort. This also helps to contrast the results with the multiple regression model because the regression coefficients from both models are now on the same scale. Moreover, MIBI and ALR reflect women's lifetime reproductive effort because birth spacing is associated with the offspring's postnatal mortality and the duration of lactation, and because ALR characterizes the length of women's "reproductive window." And, as assumed by the measurement model, all these effect indicators show at least moderate positive intercorrelations (Table 2). Please note that women's age at first reproduction would not be a suitable for an effect indicator here, because it is negatively correlated with the number of offspring surviving to adulthood (Helle et al., 2005). The reliability of the latent variable, women's lifetime reproductive effort, was estimated by calculating its composite reliability (Raykov, 2004). Whether a woman survived to the postmenopausal period or not was treated as an observed dichotomous response variable (Figure 1).

The regression estimate of women's latent lifetime reproductive effort on whether they survived to the postmenopausal period or not was compared to a regular regression estimate obtained from a traditional multiple regression model. In this model, the single best indicator of women's lifetime reproductive effort in these data, their number of adult offspring, is regressed on whether women survived to the post-menopausal period or not (Figure 1B). To make the comparison appropriate to the SEM, MIBI and ALR were added into the model as continuous covariates (Figure 1B).

Because the response variable was dichotomous, logit link function was used as in a regular logistic regression analysis to model the probability that a woman survived to the postmenopausa period. To avoid selection bias among the individuals born prior to systematic data collection, study parish, and birth cohort were combined to form unique stratum that were used in disproportionate stratification with differing sampling weights (i.e., proportion of individuals sampled differed between the stratum before systematic data collection) to obtain less-biased parameters and their standard errors (Stapleton, 2013). Owing to the use of sampling weights, a pseudo maximum likelihood estimator with robust standard errors was applied to estimate the model parameters. Please note that no evaluation of global model fit to the data was conducted as commonly done in SEM. This is because the aim here is not to biologically interpret the estimates from this simple example and because multiple regression models are always saturated, thus providing a perfect fit to the data. Instead, the focus here was on the local estimation of the coefficient from women's lifetime reproductive effort or their number of adult offspring to their survival probability to the postmenopausal period. Analyses were conducted with Mplus version 7.4 (Muthén & Muthén, 2015).

3 | RESULTS

The results of the SEM are shown in Table 3. The measurement part of the model shows that all indicators loaded significantly and positively onto the latent lifetime reproductive effort (Table 2). A one unit increase in women's lifetime reproductive effort, scaled by the number of adult offspring, was associated with 0.51 (95% CIs = 0.41, 0.60) year increase in their MIBI and with 3.86 (95% CIs = 2.76, 4.96) year increase in their ALR (Table 3). A standardized solution for these loadings shows that women's ALR responded most strongly to variation in their lifetime reproductive effort, followed by MIBI and the number of surviving offspring. The proportion of unique indicator variance explained by women's lifetime reproductive effort was 32.7, 35.5, and 65.7% for OFS, MIBIs, and ALR, respectively. The composite reliability of women's lifetime reproductive effort was 0.66 (95% CIs = 0.55, 0.77), meaning that the indicators used here explained on average 66% of the variance in women's lifetime reproductive effort.

The structural part of the SEM indicated that one unit increase in women's lifetime reproductive effort was associated with 2.67-fold (95% CIs = 1.59, 3.75) higher odds of surviving to the postmenopausal period (Table 3).

The results of the multiple regression model are shown in Table 4. According to this model, the number of offspring surviving to adulthood was not statistically associated with the women's probability to survive to the post-menopausal period (Table 4): one additional offspring surviving to

TABLE 3 The results of SEM examining the influence of women's lifetime reproductive effort on whether they survived to post-reproductive period or not

	β	β_{stdXY}	S.E	*z*	*P*-value
Factor loadings (λ)					
Lifetime reproductive effort					
Number of adult offspring	1.000	0.572			
Age at last reproduction	3.860	0.810	0.560	6.887	<0.0001
Mean interbirth interval	0.505	0.595	0.049	10.379	<0.0001
Structural path coefficients (β)					
Survival to postreproductive period					
Lifetime reproductive effort	0.983		0.206	4.771	<0.0001
Intercepts					
Number of adult offspring	4.045		0.106	38.123	<0.0001
Age at last reproduction	39.312		0.290	135.341	<0.0001
Mean interbirth interval	2.095		0.055	38.066	<0.0001
Thresholds					
Survival to postreproductive period	−1.556		0.164	−9.519	<0.0001
Variances (ϕ)					
Lifetime reproductive effort	1.619		0.312	5.185	<0.0001
Residual variances (δ)					
Number of adult offspring	3.339		0.359	9.300	<0.0001
Age at last reproduction	12.618		3.495	3.611	<0.0001
Mean interbirth interval	0.751		0.114	6.570	<0.0001

Note: All the loadings of lifetime reproductive effort are on raw scale while the structural coefficient from lifetime reproductive effort to survival to postreproductive period is on logit-scale. Distributional locations are given as intercepts and thresholds for continuous and discrete responses, respectively. Standardized loadings (β_{stdXY}) represent fully standardized solutions for effect indicators. The total number of parameters estimated was 11.

adulthood increased the women's odds of surviving to post-reproductive period by 2.4% (95% CIs = 0.89, 1.16) only.

4 | DISCUSSION

Revealing causality is the ultimate goal for scientific research (Pearl, 2009; Shipley, 2000). This task becomes very demanding when an experimental approach cannot be undertaken. In correlative research, our understanding of the complex model systems under study is often limited to an extent that confounding of causal inference owing to missing variables is very likely. Further development of underlying theory can assist in overcoming such obstacles to some degree. When additionally dealing with hard-to-measure constructs, statistical models that explicitly incorporate that uncertainty

TABLE 4 The results of a multiple regression model examining the influence of women's number of offspring surviving to adulthood, mean interbirth interval, and age at last reproduction on whether they survived to the postreproductive period

	β	β_{stdXY}	S.E	*z*	*P*-value
Partial regression coefficients (β)					
Number of adult offspring	0.024	0.027	0.070	0.374	0.729
Age at last reproduction	0.172	0.471	0.032	5.330	<0.0001
Mean interbirth interval	0.010	0.064	0.130	0.074	0.941
Thresholds					
Survival to post-reproductive period	5.413		1.123	4.821	<0.0001

Note: All the multiple regression coefficients are on logit-scale. Standardized loadings (β_{stdXY}) represent fully standardized solutions for effect indicators. The total number of parameters estimated was 4.

into constructs of interest can further bring us closer to the causal effects we are searching for.

Inference problems owing to measurement error in independent variables has been known in statistics for more than a century now (Spearman, 1904) and the solution provided by SEM to this problem has been previously highlighted in biological sciences (Grace, 2006; Pugesek, Tomer, & von Eye, 2003; Shipley, 2000), but to date few studies have taken a full advantage of a multiple-indicator latent variable approach. Although an approach based on an *a priori* estimate of measurement error of constructs can also be adopted in a SEM framework by using latent variables with single indicators and fixed error variances (e.g., Hayduk & Littvay 2012), the use of multiple indicators to account for the imperfect measurement of latent constructs does not require researchers to predefine its error variance. Such approaches may be problematic if these uncertainties cannot be quantified with reasonable accuracy, particularly if construct validity is in question (Pugesek & Tomer 1995), which is likely the case in estimating lifetime reproductive effort in humans used as an example in this article. Naturally, SEM with multiple-indicator latent variables is not a solution in all situations. When there are multiple sources of measurement error that are not well captured by the indicators, and thus not properly modelled, structural estimates from SEM can be biased (DeShon, 1998).

In addition to these approaches, there are statistical approaches that do not assume that independent or dependent variables are measured without error, such as (reduced) major axis regression (or type 2 regressions) used in studies of allometry (Warton, Wright, Falster, & Westoby, 2006). However, the applicability of these regression methods to study biological questions has been recently heavily critized (Hansen & Bartoszek, 2012; Smith, 2009). Some of the most recent approaches consider measurement error in variables as a missing data problem (Blackwell, Honaker, & King, 2015). However, a major advantage of SEM for human life history research lies in its great flexibility to accommodate, for example, non-Gaussian multilevel and longitudinal models as well as a Bayesian approach (e.g., Hoyle, 2012; Song & Lee, 2012).

The discrepancy between how lifetime reproductive effort is defined in life history theory and how the predictions involving lifetime reproductive effort have currently been tested in correlative human data provides an example where measurement error likely plays a role. That is, previous investigations on the trade-off between lifetime reproductive effort and postreproductive survival, a core explanation for the evolution of senescence (Kirkwood & Rose, 1991; Williams, 1957), have failed to recognize the mismatch between what is exactly predicted by the theory and whether the statistical approaches currently in use in our field can appropriately test these predictions with the available data. This mismatch inevitably leads to unpredicted inconsistency in causal effects due to measurement error in women's lifetime reproductive effort (Antonakis et al. 2010, 2014; Cole & Preacher, 2014), which may partly explain the mixed results found in human studies examining costs of reproduction (Gagnon, 2015; Helle et al., 2005; Hurt et al., 2006; Jasienska, 2009; Le Bourg, 2007). Although there have also been recent efforts to examine the costs of reproduction at the genetic level in humans (Gögele et al., 2011; Wang, Byars, & Stearns, 2013), and presumably such studies are likely to increase in number in the future owing to greater data availability, our current knowledge in this area relies heavily on phenotypic associations that are particularly vulnerable to methodological problems like measurement error in correlated predictors.

The example presented here was designed for illustrative purposes only, and thus should not be taken as a scientific investigation of a trade-off between lifetime reproductive effort and post-menopausal survival in women. Because the current analysis fully ignored the potential confounding owing to differences in resource availability among women and other potentially important covariates for this question, the finding that increased lifetime reproductive effort in women increased, rather than decreased, their odds of surviving to the postmenopausal period is expected at the phenotypic level but unlikely to represent a causal effect of reproductive effort on somatic senescence (van Noordwijk & de Jong, 1986). Nevertheless, the current modeling exercise clearly shows the statistical consequences of not accounting for measurement error in underlying constructs and how this can profoundly alter the conclusions drawn from the data. Had we used a multiple regression model assuming that women's lifetime reproductive effort is fully captured by their number of adult offspring, and is thus measured without error, we would have concluded that women's lifetime reproductive effort was unrelated to women's probability to survive to the post-menopausal period. This example thus demonstrates the classical attenuation of regression coefficients due to measurement error in independent variables (Antonakis et al., 2010, 2014; Spearman, 1904).

It is pivotal to understand that the adverse consequences of measurement error in independent variables on (partial) regression estimates does not primarily depend on the data analyzed *per se*, but on the critical assumption of whether the researcher believes he/she has accurately measured the construct of interest or not. That is, do the variables included in the analysis reliably represent the theoretical constructs they should be measuring (i.e., validity of measurements)? For example, the argument and the analysis presented here relied on the premise that it is the amount of limited energetic resources allocated in reproduction during the lifetime (i.e., lifetime reproductive effort) that causes variation in old-age survival (Kirkwood & Rose, 1991), but that in humans

those resources cannot be measured without error, particularly using demographic data. However, if reproductive costs in women are mainly mediated via other than energetic pathways responding to environmental resource availability (e.g., Edwards & Chapman, 2011) and if those pathways can be properly measured by the existing demographic data, or if the number of offspring surviving to adulthood precisely enough captures variation in women's lifetime reproductive effort, then it would be completely legitimate to ignore measurement error in this construct and to rely on the results obtained from a regular multiple regression model.

The life history traits used here as effect indicators, causally reflecting women's lifetime reproductive effort, were mainly chosen for illustrative purposes in mind. As indicated by the reliability of the latent variable, 34% of the variance (but plausibly as much as 45%) in women's lifetime reproductive effort was left unexplained by OFS, MIBIs, and ALR. In other words, women's lifetime reproductive effort as a latent variable still contained a fair amount of measurement error. Naturally, a researcher could use other effect indicators if available, or preferably include more effect indicators to increase the reliability of the latent variable of interest. However, one should not try to blindly add more life history traits as effect indicators to model latent lifetime reproductive effort: not all the life history traits available from demographic data may satisfy the assumption that lifetime reproductive effort has a causal effect on the trait. There are other alternative, more flexible ways to specify measurement models in SEM that can handle a more complex set of observed variables by, for example, allowing negative or zero associations among indicators as well as more complex causal structures (Bollen & Bauldry, 2011). We were also assuming that the indicators recorded here were truly measuring women's lifetime reproductive effort. In theory, this may not have been the case and we might have actually measured some other construct or constructs, causing variation in the life history traits used here (Kline, 2015).

This article has hopefully increased the awareness of several, usually hard-to-meet assumptions that commonly used regression approaches make. Ignoring those assumptions will severely limit their usefulness to make valid biological conclusions from the data. In addition to measurement error in the independent variables discussed here, reciprocal causality between dependent and independent variables in cross-sectional data, missing confounding variables, and autocorrelated and nonindependent residuals also undermine causal inference in regression modeling (Antonakis et al., 2010, 2014). Some even consider the resulting "associations" from studies unable to confront to these assumptions scientifically uninteresting since the magnitude and direction of such associations can easily be spurious (Antonakis et al., 2010). Because human life history research rests mainly on correlative data, the researchers should pay more attention to measurement theory and how imperfect measurements can be incorporated into the statistical analyses.

ACKNOWLEDGMENTS

Thanks to Matti Leiviskä and Kimmo Oikarinen for collecting the data. This study was funded by Kone Foundation (grants no. 085253 and 086809).

REFERENCES

Adolph, S. C., & Hardin, J. S. (2007). Estimating phenotypic correlations: Correcting for bias due to intraindividual variability. *Functional Ecology*, *21*, 178–184.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*, 1086–1120.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2014). Causality and endogeneity: Problems and solutions. In D.V. Day (Ed.), *The oxford handbook of leadership and organizations* (pp. 93–117). New York: Oxford University Press.

Blackwell, M., Honaker, J., & King, C. (2015). A unified approach to measurement error and missing data: Overview and applications. *Social Methods and Research*, 1–39.

Bollen, K. A. (1989) *Structural equations with latent variables*. Wiley: New York.

Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, *16*, 265–284.

Buonaccorsi, J. P. (2010) *Measurement error: Models, methods, and applications*. New York: Chapman & Hall.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006) *Measure-ment error in nonlinear models: A modern perspective*, 2nd ed. London: Chapman & Hall.

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, *19*, 300–315.

DeShon, R. P. (1998). A cautionary note on measurement error corrections in structural equation models. *Psychological Methods*, *3*, 412–423.

Edwards, D. A., & Chapman, T. (2011). Mechanisms underlying reproductive trade-offs: Costs of reproduction. In T. Flatt & A. Heyland (Eds.), *Mechanisms of life history evolution. The genetics and physiology of life history traits and trade-offs* (pp. 137–152). New York: Oxford University Press.

Enbuske, M. (2008). *Vanhan lapin valtamailla: Asutus ja maankäyttö kemin lapin ja enontekiön alueella 1500-luvulta 1900-luvun alkuun*. Helsinki: Suomen Kirjallisuuden Seura.

Freckleton, R. P. (2011). Dealing with collinearity in behavioral and ecological data: model averaging and the problems of measurement error. *Behavioral Ecology and Sociobiology*, *65*, 91–101.

Gagnon, A. (2015). Natural fertility and longevity. *Fertility & Sterility*, *103*, 1109–1116.

Gagnon, A., Smith, K. R., Tremblay, M., Vézina, H., Paré, P.-P., & Desjardins, B. (2009). Is there a trade-off between fertility and longevity? A comparative study of women from three large

historical databases accounting for mortality selection. *American Journal of Human Biology*, *21*, 533–540.

Gögele, M., Pattaro, C., Fuchsberger, C., Minelli, C., Pramstaller, P. P., & Wjst, M. (2011). Heritability analysis of life span in a semi-isolated population followed across four centuries reveals the presence of pleiotropy between life span and reproduction. *The Journals of Gerontology: Series A, Biological Sciences and Medical Sciences*, *66A*, 26–37.

Grace, J. B. (2006). *Structural equation modeling and natural systems*. Cambridge, UK: Cambridge University Press.

Hansen, T. F., & Bartoszek, K. (2012). Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. *Systematic Biology*, *61*, 413–425.

Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methods*, *12*, 159.

Helle, S., Brommer, J. E., Pettay, J. E., Lummaa, V., Enbuske, M., & Jokela, J. (2014). Evolutionary demography of agricultural revolution in preindustrial northern Finland. *Proceeding of the Royal Society: Biological Sciences*, *281*(1794):20141559.

Helle, S., & Lummaa, V. (2013). A trade-off between having many sons and shorter maternal post-reproductive survival in preindustrial Finland. *Biol. Lett. 9*, 20130034.

Helle, S., Lummaa, V., & Jokela, J. (2002). Sons reduced maternal longevity in preindustrial humans. *Science*, *296*, 1085.

Helle, S., Lummaa, V., & Jokela, J. (2004). Accelerated immunosenescence in preindustrial twin mothers. *Proceedings of the National Academy of Sciences USA*, *101*, 12391–12396.

Helle, S., Lummaa, V., & Jokela, J. (2005). Are reproductive and somatic senescence coupled in humans? Late, but not early, reproduction correlated with longevity in historical Sami women. *Proceeding of the Royal Society: Biological Sciences*, *272*, 29–37.

Houle, D., Pelabon, C., Wagner, G. P., & Hansen, T. F. (2011). Measurement and Meaning in Biology. *The Quarterly Review of Biology*, *86*, 3–34.

Hoyle, R. H. (2012). *Handbook of structural equation modeling*. New York: The Guilford Press.

Hurt, L. S., Ronsmans, C., & Thomas, S. L. (2006). The effect of number of births on women's mortality: Systematic review of the evidence for women who have completed their childbearing. *Population Studies*, *60*, 55–71.

Jasienska, G. (2009). Reproduction and lifespan: trade-offs, overall energy budgets, intergenerational costs, and costs neglected by research. *American Journal of Human Biology*, *21*, 524–532.

Kendall, B. E. (2015). A statistical symphony: instrumental variables reveal causality and control measurement error. In G.A. Fox, S. Negrete-Yankelevich & V.J. Sosa (Eds.), *Ecological statistics: Contemporary theory and application* (pp. 149–167). New York: Oxford University Press.

Kirkwood, T. B. L., & Rose, M. R. (1991). Evolution of senescence: Late survival sacrificed for reproduction. *Philosophical Transactions of Royal Society of London: Series B, Biological Sciences*, *332*, 15–24.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*, 4th ed. New York: The Guilford Press.

Le Bourg, E. (2007). Does reproduction decrease longevity in human beings? *Ageing Research Reviews*, *6*, 141–149.

Luther, G. (1993). *Suomen tilastotoimen historia*. Helsinki: WSOY.

Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Pearl, J. (2009). *Causality: Models, reasoning and inference*, 2nd ed. Cambridge, England: Cambridge University Press.

Pugesek, B. H., & Tomer, A. (1995). Determination of selection gradients using multiple regression versus structural equation models (SEM). *Biometrical Journal*, *37*, 449–462.

Pugesek, B. H., Tomer, A., & von Eye, A. (2003). *Structural equation modeling. Applications in ecological and evolutionary biology*. Cambridge: Cambridge University Press.

Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy*, *35*, 299–331.

Roff, D. A. (2002). *Life history evolution*. Sunderland, MA: Sinauer Associates Inc.

Sear, R. (2007). The impact of reproduction on Gambian women: Does controlling for phenotypic quality reveal costs of reproduction?. *American Journal of Physical Anthropology*, *132*, 632–641.

Shipley, B. (2000). *Cause and correlation in biology. A user's guide to path analysis, structural equations and causal inference* (Vol. *140*, pp. 476–486). Cambridge, UK: Cambridge University Press.

Song, X.-Y., & Lee, S.-Y. (2012). *Basic and advanced bayesian structural equation modeling. With applications in the medical and behavioral sciences*. Singapore: Wiley.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72–101.

Stapleton, L. (2013). *Multilevel Structural Equation Modelin with Complex Sample Data*. In G.R. Hancock & R.O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 521–562), 2nd ed. Charlotte, NC: Information Age Publishing, Inc.

Stearns, S. C. (1992) *The evolution of life histories*. Oxford: Oxford University Press.

Wang, X., Byars, S. G., & Stearns, S. C. (2013). Genetic links between post-reproductive lifespan and family size in Framingham. *Evolution, Medicine, & Public Health*, *1*, 241–253.

Warton, D. I., Wright, I. J., Falster, D. S., & Westoby, W. (2006). Bivariate line-fitting methods for allometry. *Biological Reviews*, *81*, 259–291.

Williams, G. C. (1957). Pleiotropy, natural selection and the evolution of senescence. *Evolution*, *11*, 398–411.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

How to cite this article: Helle S. Accounting for measurement error in human life history trade-offs using structural equation modeling. *Am J Hum Biol*. 2017; e23075. https://doi.org/10.1002/ajhb.23075