Supporting Information

Underestimation of Thyroid Dysfunction Risk due to Regression Dilution Bias in a Long-Term Follow-Up: Tehran Thyroid Study (TTS)

Authors
A. Amouzegar¹, M. Beigy¹, S. Gharibzadeh¹,², F. Azizi¹

Affiliations
¹Endocrine Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, I. R. Iran
²Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, I. R. Iran

Correspondence
Fereidoun Azizi
Professor of Internal Medicine and Endocrinology
Endocrine Research Center
Research Institute for Endocrine Sciences
Shahid Beheshti University of Medical Sciences
P. O. Box 19395-4763
Tehran
I. R. Iran
Tel.: + 98/21/2243 2503
Fax: + 98/21/2240 2463
Azizi@endocrine.ac.ir
We used the Rosner regression model to obtain an estimated RDR for each TTS Ph. Let $y_i$ be the continuous outcome variable (for individuals $i = 1, \ldots, N$) such as FT4 and $w_i$ be the single measurement of predictor variable such as FT4 and TSH, which are susceptible to substantial day-to-day within-individual variation. Consider $x_i$ as the true long-life value of the predictor variable and $u_i$ as the random error for predictor variable. Thanks to the explanations of Frost and Thompson [7] we suppose that the association between the predictor variable $x_i$ and the outcome variable $y_i$ is described by a simple linear model where both $x_i$ and $u_i$ are normally distributed:

$$y_i = \alpha^* + \beta^* x_i + \delta_i \delta_i \sim N(0, \Phi^2),$$

$$w_i = x_i + u_i x_i \sim N(\mu, \sigma^2_{\delta}), u_i \sim N(0, \sigma^2_{\delta})$$

Since the true long-life value of predictor variable $x_i$ is not accessible for us, we are forced to estimate the relationship between single measurement of predictor variable $w_i$ and outcome variable $y_i$ through the following linear model of regression by assuming that $u_i, \delta_i, \text{ and } x_i$ are independently distributed:

$$y_i = \alpha + \beta w_i + \gamma_i \gamma_i \sim N(0, \psi^2),$$

$$\beta = \beta^* (\sigma^2_{\delta} / (\sigma^2_{\delta} + \sigma^2_{\psi})) \quad (1)$$

It is claimed, in this model, that $y_i$ and $w_i$ follow a bivariate normal distribution, and therefore, by regressing $y_i$ on $w_i$ instead of $x_i$, the estimated regression slope $\beta$ is decreased by a factor equal to the proportion of the total variance of $w_i$ which is between-individuals.

Let the $\lambda$ be correction factor that is employed to adjust the regression dilution bias:

$$\lambda = 1 + \frac{\sigma^2_{\psi}}{\sigma^2_{\delta}}$$

So that $\beta^* = \lambda \beta$ will enable us to adjust the estimated regression coefficient $\beta$ from the main study [7]. Actually random error in the corrected coefficient $\beta^*$ displays random error in the estimates of $\beta$ and $\lambda$ [7].

Here, we assume that within-person random errors are independently normally distributed. Let $w_{ij}$ be the $j$th repeat measurement of the predictor variable for the $i$th individual, thus:

$$w_{ij} = x_{ij} + u_{ij}, x_{ij} \sim N(\mu, \sigma^2_{\delta}) u_{ij} \sim N(0, \sigma^2_{\psi})$$

Rosner regression model for obtaining the estimated RDR for each repeated measure in TTS:

Let the $j$th repeated FT4 or TSH measurement of the $i$th individual in TTS be $Y_i (i = 1, \ldots, n)$;
\( j = 1, \ldots, m \) where \( n \) is the number of individuals in TTS, and \( m \) is the number of repeated measures or TTS Ph. The Rosner regression model will be then:

\[
Y_{ij} = \alpha_i + \beta Y_{i0} + \varepsilon_{ij}
\]

For each repeated measure \( m > 0 \), where \( \varepsilon_{ij} \sim N(0, \sigma_j^2) \) and \( \beta \) are the RDR.

As an example, the TTS phase 2 measure for TSH (\( w_{i2} \)) is regressed on the phase 1 (\( w_{i1} \)) with the regression coefficient \( \beta \), and \( \lambda \) will be estimated by \( \lambda = \beta^{-1} \). To obtain \( \lambda \) and its variance, first we consider \( y_i = w_{i2} \) (in equation 1) and assume \( \beta^* = 1 \) and \( \beta = (\sigma_b^2/(\sigma_b^2 + \sigma_W^2)) \), so the regression relationship between \( w_{i2} \) and \( w_{i1} \) will be [7]:

\[
w_{i2} = \alpha + \left( \frac{\sigma_b^2}{\sigma_b^2 + \sigma_W^2} \right) w_{i1} + \gamma_i \gamma_j \sim N(0, \psi^2)
\]

(2)

Therefore, \( \beta \) will be the unbiased estimator of \( \sigma_b^2/(\sigma_b^2 + \sigma_W^2) \) [7]. The asymptotically unbiased estimator of \( \lambda \) is:

\[
\lambda = \beta^{-1} = \frac{\sum (w_{i1} - \bar{w}_1)^2}{\sum (w_{i1} - \bar{w}_1)(w_{i2} - \bar{w}_2)}
\]

Finally, variance of \( \lambda \) can be estimated as:

\[
\text{var}(\lambda) = \frac{1}{\beta^2} \text{var}(\beta) \approx \frac{\lambda^2(\lambda^2 - 1)}{n}
\]