

Let y_{mir} denote the r th replicate measurement on the i th item by the m th method, where $m = 1, 2$; $i = 1, \dots, N$; and $r = 1, \dots, n_i$. When the design is balanced and there is no ambiguity we can set $n_i = n$. The LME model underpinning Roy's approach can be written

$$y_{mir} = \beta_0 + \beta_m + b_{mi} + \epsilon_{mir}. \quad (1)$$

Here β_0 and β_m are fixed-effect terms representing, respectively, a model intercept and an overall effect for method m . The model can be reparameterized by gathering the β terms together into (fixed effect) intercept terms $\alpha_m = \beta_0 + \beta_m$. The b_{1i} and b_{2i} terms are correlated random effect parameters having $E(b_{mi}) = 0$ with $\text{Var}(b_{mi}) = g_m^2$ and $\text{Cov}(b_{1i}, b_{2i}) = g_{12}$. The random error term for each response is denoted ϵ_{mir} having $E(\epsilon_{mir}) = 0$, $\text{Var}(\epsilon_{mir}) = \sigma_m^2$, $\text{Cov}(\epsilon_{1ir}, \epsilon_{2ir}) = \sigma_{12}$, $\text{Cov}(\epsilon_{mir}, \epsilon_{mir'}) = 0$ and $\text{Cov}(\epsilon_{1ir}, \epsilon_{2ir'}) = 0$. Additionally these parameter are assumed to have Gaussian distribution. Two methods of measurement are in complete agreement if the null hypotheses $H_1: \alpha_1 = \alpha_2$ and $H_2: \sigma_1^2 = \sigma_2^2$ and $H_3: g_1^2 = g_2^2$ hold simultaneously. ? uses a Bonferroni correction to control the familywise error rate for tests of $\{H_1, H_2, H_3\}$ and account for difficulties arising due to multiple testing. Additionally, Roy combines H_2 and H_3 into a single testable hypothesis $H_4: \omega_1^2 = \omega_2^2$, where $\omega_m^2 = \sigma_m^2 + g_m^2$ represent the overall variability of method m .

? also use a LME model for the purpose of comparing two methods of measurement where replicate measurements are available on each item. Their interest lies in generalizing the popular limits-of-agreement (LOA) methodology advocated by ? to take proper cognizance of the replicate measurements. ? demonstrate statistical flaws with two approaches proposed by ? for the purpose of calculating the variance of the inter-method bias when replicate measurements are available. Instead, they recommend a fitted mixed effects model to obtain appropriate estimates for the variance of the inter-method bias. As their interest mainly lies in extending the Bland-Altman methodology, other formal tests are not considered.

? develop their model from a standard two-way analysis of variance model, refor-

mulated for the case of replicate measurements, with random effects terms specified as appropriate. Their model can be written as

$$y_{mir} = \alpha_m + \mu_i + a_{ir} + c_{mi} + \varepsilon_{mir}. \quad (2)$$

The fixed effects α_m and μ_i represent the intercept for method m and the ‘true value’ for item i respectively. The random-effect terms comprise an item-by-replicate interaction term $a_{ir} \sim \mathcal{N}(0, \varsigma^2)$, a method-by-item interaction term $c_{mi} \sim \mathcal{N}(0, \tau_m^2)$, and model error terms $\varepsilon_{mir} \sim \mathcal{N}(0, \varphi_m^2)$. All random-effect terms are assumed to be independent. For the case when replicate measurements are assumed to be exchangeable for item i , a_{ir} can be removed. The model expressed in (2) describes measurements by m methods, where $m = \{1, 2, 3 \dots\}$. Based on the model expressed in (2), ? compute the limits of agreement as

$$\alpha_1 - \alpha_2 \pm 2\sqrt{\tau_1^2 + \tau_2^2 + \varphi_1^2 + \varphi_2^2}$$

? notes that, for $m = 2$, separate estimates of τ_m^2 can not be obtained. To overcome this, the assumption of equality, i.e. $\tau_1^2 = \tau_2^2$ is required.

There is a substantial difference in the number of fixed parameters used by the respective models; the model in (1) requires two fixed effect parameters, i.e. the means of the two methods, for any number of items N , whereas the model in (2) requires $N + 2$ fixed effects.

Allocating fixed effects to each item i by (2) accords with earlier work on comparing methods of measurement, such as ?. However allocation of fixed effects in ANOVA models suggests that the group of items is itself of particular interest, rather than as a representative sample used of the overall population. However this approach seems contrary to the purpose of LOAs as a prediction interval for a population of items. Conversely, ? uses a more intuitive approach, treating the observations as a random sample population, and allocating random effects accordingly.