SUPPLEMENTARY DATA

Supplementary material to "Recommendations for the Definition of the Clinical Responder to Support Efficacy Claims in Insulin Preservation Studies Submitted To US or European Regulatory Agencies"

Beam, Gitelman, Palmer

In this supplemental material, we introduce the statistical model underpinning our approach, detail how characteristics of a responder definition are thus defined, and describe how we implemented the model in practice.

1. Statistical Model

Broadly speaking, the model is a multivariate normal "errors in variables" model in which some of the variables are unobservable and the measurement error variance is assumed known. The errors in variables model has been proposed for responder analysis in the past, but only from a univariate[1] or bivariate[2,3] standpoint. We extend these past approaches via a multivariate model having unobservable variables. The assumption of known error variance has been utilized in the literature concerning misclassification arising from measurement error in which an independent study is conducted to estimate this variance (Byonaccorsi, JASA 1990).

For subject i measured at time t=1,2 we observe the apparent value y which arises from measurement error of the actual value x:

$$y_{ti} = x_{ti} + \epsilon_{ti}$$

We assume that the vector $X_i \equiv [x_{1i}, x_{2i}]'$ is multivariate normal having parameters

Equation 1

$$E(X_i) = [\mu_1, \mu_2]'$$

$$V(X_i) \equiv \Sigma_{xx} = \begin{bmatrix} \sigma_x^2 & \sigma_{x_1x_2} \\ \sigma_{x_1x_2} & \sigma_x^2 \end{bmatrix}$$

Result 1

Let $Y_i \equiv [y_{1i}, y_{2i}]'$. Assuming that $\epsilon_{ti} \sim iid\, N(0, \sigma_\epsilon^2)$ are independent of x_{ti}, the vector $[X_i, Y_i]$ is multivariate normal having parameters

$$E([X_i, Y_i]) = [\mu_1, \mu_2, \mu_1, \mu_2]'$$

$$V([X_i, Y_i]) = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xx} \\ \Sigma_{xx} & \Sigma_{xx} + \sigma_\epsilon^2 I \end{bmatrix}$$

Proof: This result follows immediately from the assumptions:

$$E(y_t) = E(x_t + \epsilon_t) = \mu_t$$

$$V(y_t) = V(x_t + \epsilon_t) = \sigma_x^2 + \sigma_\epsilon^2$$

$$Cov(y_1, y_2) = E\big((y_1 - \mu_1)(y_2 - \mu_2)\big)$$

$$= E(((x_1 + \epsilon_1) - \mu_1)((x_2 + \epsilon_2) - \mu_2)))$$

$$= E\Big(\big((x_1 - \mu_1) + \varepsilon_1\big)\big((x_2 - \mu_2) + \varepsilon_2\big)\Big)$$

$$= E((x_1 - \mu_1)(x_2 - \mu_2) + (x_1 - \mu_1)\varepsilon_2 + (x_2 - \mu_2)\varepsilon_1 + \varepsilon_1\varepsilon_2 = \sigma_{x_1x_2}$$

$$Cov(x_1, y_1) = E\big((x_1 - \mu_1)(x_1 + \epsilon_1 - \mu_1)\big)$$

$$= E((x_1 - \mu_1)^2 + E(x_1 - \mu_1)E(\epsilon_1) = \sigma_x^2$$

$$\Rightarrow Cov(x_2, y_2) = \sigma_x^2$$

$$Cov(x_1, y_2) = E\big((x_1 - \mu_1)(x_2 + \epsilon_2 - \mu_2)\big)$$

$$= E(x_1 - \mu_1)(x_2 - \mu_2) + E(x_1 - \mu_1)E(\epsilon_2) = \sigma_{x_1x_2}$$

$$\Rightarrow Cov(x_2, y_1) = \sigma_{x_1x_2}$$

Result 2
By the preceding, if $|\Sigma_{xx} - \sigma_\epsilon^2 \mathrm{I}| > 0$ then the conditional distribution of the unobservable X_i given the observable Y_i is multivariate normal having parameters

$$E(X_i|Y_i) = [\mu_1, \mu_2]' + \Sigma_{xx}[\Sigma_{xx} + \sigma_\epsilon^2 \mathrm{I}]^{-1}(Y_i - [\mu_1, \mu_2]')$$
$$V(X_i|Y_i) = \Sigma_{xx} - \Sigma_{xx}[\Sigma_{xx} + \sigma_\epsilon^2 \mathrm{I}]^{-1}\Sigma_{xx}$$

This results follows from Result 1 and multivariate normal theory (see Result 4.6 on page 135 of Johnson[4]).
However, in practical applications, since x is unobservable, parameters of the joint distribution or conditional distribution cannot be directly estimated. Nonetheless, we can re-express the previous results in terms of the variance-covariance matrix of the observable y vector:
Equation 2

$$V([X_i, Y_i]) = \begin{bmatrix} \Sigma_{yy} - \sigma_\epsilon^2 \mathrm{I} & \Sigma_{yy} - \sigma_\epsilon^2 \mathrm{I} \\ \Sigma_{yy} - \sigma_\epsilon^2 \mathrm{I} & \Sigma_{yy} \end{bmatrix}$$

Where,
Equation 3

$$\Sigma_{yy} = \begin{bmatrix} \sigma_x^2 + \sigma_\epsilon^2 & \sigma_{x_1x_2} \\ \sigma_{x_1x_2} & \sigma_x^2 + \sigma_\epsilon^2 \end{bmatrix}$$

So that,
Equation 4

$$E(X_i|Y_i) = [\mu_1, \mu_2]' + \left(\Sigma_{yy} - \sigma_\epsilon^2 \mathrm{I}\right)\Sigma_{yy}{}^{-1}(Y_i - [\mu_1, \mu_2]')$$
$$V(X_i|Y_i) = \left(\Sigma_{yy} - \sigma_\epsilon^2 \mathrm{I}\right) - \left(\Sigma_{yy} - \sigma_\epsilon^2 \mathrm{I}\right)\Sigma_{yy}{}^{-1}\left(\Sigma_{yy} - \sigma_\epsilon^2 \mathrm{I}\right)$$

2. Evaluating Characteristics of Responder Definitions

Most published responder definitions represent conditions placed on the difference or ratio of two measurements separated in time. Such conditions can be expressed generally as a linear inequality in hyperspace, e.g. $\{(x_1, x_2, y_1, y_2):\ \ x_2 \geq a + bx_1\}$. A responder definition based on the difference of x_1, x_2 would set b=1 and a definition based on their ratio would set a=0.
Therefore, actual and apparent responder proportions can be evaluated by integrating the multivariate density function over the appropriate region in $\mathcal{R}^4$. Let ϕ denote the multivariate density function having parameters given by equations 1 and 2, then the characteristics of interest are defined by the following integrals:

Actual responder proportion=

Equation 5

$$\int_{\{(x_1, x_2, y_1, y_2):\ \ x_2 \geq a + bx_1\}} \phi \, d(x_1, x_2, y_1, y_2)$$

Apparent responder proportion=

SUPPLEMENTARY DATA

Equation 6

$$\int_{\{(x_1,x_2,y_1,y_2):\ \ y_2 \geq a+by_1\}} \phi\, d(x_1,x_2,y_1,y_2)$$

Misclassification error rates can be evaluated by integrating the conditional distribution defined in Equation 3: Let $\phi_{x|y}$ denote the density function of X given Y and let ϕ_y denote the marginal density of Y. $\phi_{x|y}$ is bivariate normal with parameters given by Equation 4. ϕ_y is also bivariate normal with parameters $E(Y) = E(X)$ and $V(Y) = \Sigma_{yy}$ as defined in Equation 3. Recalling that $\phi_{x|y}$ is a function of y, we can then define the following quantities;

False responder proportion=

Equation 7

$$\int_{\{(y_1,y_2):y_2 \geq a+by_1\}} \left\{ \int_{\{(x_1,x_2):\ \ x_2 < a+bx_1\}} \phi_{(x|y)}\, d(x_1,x_2) \right\} \phi_y\, d(y_1,y_2)$$

False non-responder proportion=

Equation 8

$$\int_{\{(y_1,y_2):y_2 < a+by_1\}} \left\{ \int_{\{(x_1,x_2):\ \ x_2 \geq a+bx_1\}} \phi_{(x|y)}\, d(x_1,x_2) \right\} \phi_y\, d(y_1,y_2)$$

3. Practical Implementation

Although the computation of multivariate normal integrals can be done with functions such as “sadmvn” in the R package “mnormt”, these functions require limits of integration that can be expressed as fixed intervals. In our situation, in which the lower limit of integration is a function of one of the integrated variables (e.g. $x_2 \geq a + bx_1$ when computing the actual responder proportion), the lower limit of integration is constantly changing. To address this problem, we randomly sampled 10,000 multivariate observations from the appropriate distribution and used the proportion falling into the region of interest as the value of the responder proportion. Random sampling of multivariate observations was accomplished with the R function “mvrnorm”. We chose 10,000 observations in order to give adequate precision in estimation (standard error ≤ 0.005) while reducing computational time to an acceptable level.

Computation of the conditional probabilities defining misclassification is made additionally difficult by the need to conduct nested integration with varying limits of integration. We used the previous method of estimating proportions via sampling from multivariate normal populations to achieve a solution to this problem as well.

Define the following unconditional probabilities:

False Positive Proportion≡

$$\int_{\{(x_1,x_2,y_1,y_2):\, x_2 < a+bx_1\, \cap\ \ y_2 \geq a+by_1\}} \phi\, d(x_1,x_2,y_1,y_2)$$

and, False Negative Proportion ≡

$$\int_{\{(x_1,x_2,y_1,y_2):\, x_2 \geq a+bx_1 \cap\, y_2 < a+by_1\}} \phi \, d(x_1,x_2,y_1,y_2)$$

Then, by the definition of conditional probabilities, it follows that the probabilities of interest can be computed by the following ratios:
False Responder Proportion=False Positive Proportion/Apparent Responder Proportion,
and,
False Non-responder Proportion=False Negative Proportion/(1-Apparent Responder Proportion).

References

1. Oppenheimer L, Kher U. The impact of measurement error on the comparison of two treatments using a responder analysis. *Statistics in Medicine.* Aug 30 1999;18(16):2177-2188.

2. Kunz M. On responder analyses when a continuous variable is dichotomized and measurement error is present. *Biometrical J.* Feb 2011;53(1):137-155.

3. Kieser M, Röhmel J, Friede T. Power and sample size determination when assessing the clinical relevance of trial results by 'responder analyses'. *Statistics in Medicine.* 2004;23(21):3287-3305.

4. Johnson RA, Wichern, D.W. *Applied Multivariate Statistical Analysis*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.; 1982.