

# GBV Measurement Error Calculations

Evan Rosenman

December 2019

We denote as  $Y$  our treated mean and  $X$  our control mean. We suppose there is randomness in  $Y$  and  $X$  due to measurement error. Without loss of generality, suppose  $Y$  has mean  $\mu > 0$  and  $X$  mean 0, such that there is a true treatment effect. We will compute  $P(Y \leq X)$  under two scenarios.

First, we suppose we have symmetric error modeled as a normal distribution. We suppose the measurement error is the same in both treatment states, captured via a variance  $\sigma^2$ . Then  $Y \sim N(\mu, \sigma^2)$ ,  $X \sim N(0, \sigma^2)$ . Since they are assumed independent, we have  $Y - X \sim N(\mu, 2\sigma^2)$ . Then:

$$\begin{aligned} P(Y \leq X) &= P(Y - X \leq 0) \\ &= P\left(\frac{Y - X - \mu}{\sqrt{2\sigma^2}} \leq \frac{-\mu}{\sqrt{2\sigma^2}}\right) \\ &= \Phi\left(-\frac{\mu}{\sqrt{2\sigma^2}}\right) \end{aligned}$$

where  $\Phi(\cdot)$  represents the standard normal CDF.

Next, suppose the error is not symmetric but is instead modeled as a half-normal distribution whose minimum is given by 0 for  $X$  and by  $\mu$  for  $Y$ . In this case, we are modeling the error as solely biasing measurements upward, though by symmetry our results will also hold for error which biases downward. For now, let's define the scale parameter of this half-normal distribution to be  $\phi$ . Now, we seek to compute:

$$\begin{aligned} P(Y \leq X) &= P(Y - X \leq 0) \\ &= P((Y - \mu) - X \leq -\mu) \end{aligned}$$

where we recognize the LHS of inequality on the second line as the difference

between two independent half-normals. Define  $Z = Y - \mu$ . Then, we have:

$$\begin{aligned}
P(Y \leq X) &= P(Z - X \leq -\mu) \\
&= P(Z \leq X - \mu) \\
&= \int_{\mu}^{\infty} \operatorname{erf}\left(\frac{x - \mu}{\phi\sqrt{2}}\right) \frac{\sqrt{2}}{\phi\sqrt{\pi}} e^{-\frac{x^2}{2\phi^2}} dx \\
&= \frac{\sqrt{2}}{\phi\sqrt{\pi}} \int_{\mu}^{\infty} \operatorname{erf}\left(\frac{x - \mu}{\phi\sqrt{2}}\right) e^{-\frac{x^2}{2\phi^2}} dx
\end{aligned}$$

This quantity can be evaluated numerically. Hence, we make some explicit comparisons below.

In the first case, we equalize the variances across the symmetric and asymmetric case. To keep the variances the same, we set the scale parameter  $\phi = \sigma \frac{1}{\sqrt{1-2/\pi}}$ . Then, we can compute a few cases of the symmetric probability of  $P(Y \leq X)$  vs. the asymmetric probability of  $P(Y \leq X)$ .

$\mu$	$\sigma$	Symmetric Probability	Asymmetric Probability
0	1	0.50	0.50
1	1	0.24	0.22
2	1	0.08	0.08
3	1	0.017	0.020

In this case, it appears that the probabilities between the two cases track pretty closely together.

We might, instead, assume that we should set  $\phi = \sigma$  – essentially modeling the measurement error in the asymmetric case as a truncated version of the error in the symmetric case. In this case, we do see significant improvements.

$\mu$	$\sigma$	Symmetric Probability	Asymmetric Probability
0	1	0.50	0.50
1	1	0.24	0.22
2	1	0.08	0.08
3	1	0.017	0.0006

## 1 Potential outcomes model – reporting classes

Let  $y_i^*$  be the actual outcome of interest for the  $i$  member of the study. Let  $y_i$  be the observed outcome, described by the model  $y_i = y_i^* + \varepsilon$ , where  $\varepsilon$  is the error term which, for some of the observational units, causes  $y_i^* \neq y_i$ .

In non-differential error the model has four types:

This table becomes more complicated if differential error.

type	actual outcome	observed outcome	frequency
truthful	$y_i^*$	$y_i^*$	$p_t$
under-report	$y_i^*$	0	$p_u$
over-report	$y_i^*$	1	$p_o$
liar	$y_i^*$	$1 - y_i^*$	$p_l$

If these reporting classes are fixed (non-differential by arm, outcome-invariant, and time-invariant) then the distribution is assigned by the randomization.

Arguments about measurement error hinge on the distribution of the frequencies – the  $p$ . Arguments increase/decrease the plausibility of the frequencies of  $p$ . In one-sided measurement error, we have  $p_o = 0$  and  $p_l = 0$ .

This is not trivial because  $y_i^*$  can be modified by treatment assignment. But it's not that challenging.