RESPONSE TO COMMENT FROM T. D. STANLEY

(*NOTE: The responses below come from the corresponding author of the manuscript. The other coauthors may choose to add their own views in subsequent posts.)*

<u>COMMENT</u>: ***"Reed, Florax and Poot (2015) does not show what it claims. Mean squared errors (MSE) and type I errors are miscalculated, invalidating any generalization that one might wish to make from this study.***

<u>RESPONSE</u>: I thank Tom Stanley for his thoughtful, and very clear, comments. His comments make it very apparent where our point of disagreement lies. The remainder of this response will elaborate on this point of disagreement and explain why I believe our approach to estimating efficiency and size is correct, and why his is wrong. My response will (i) excerpt the relevant portion of Stanley's comment, (ii) show a multi-level diagram that summarizes the issue, and (iii) close with an explanation of why our approach is correct.

(i) ***<u>Excerpt from Stanley's comment explaining why he believes our approach is wrong (yellow-highlighted sections added for later reference)</u>:***

"Identifying the calculation error is relatively easy, once we understand how Reed, Florax and Poot (2015) define the population from which they are sampling. In several places, Reed, Florax and Poot (2015) identify their population clearly. First, they describe how their "population of M studies" (p.4) is taken from OLS estimates of a known true effect (a regression coefficient), α, after data has been generated from a specified DPG [sic]. Note the typo under equation (1); a is not the "true" effect. On page 5, they give further details about how their population is generated, "we repeat the above process M times, so that there is a sample of M studies . . .*This sample constitutes the population of all studies*. . . . M=300" (emphasis added). And, they confirm that these 300 studies is their population, "the population of 300 studies" (p. 9). This establishes their population, the mean of which is the population mean. These population means will be differ by random sampling error as a new set of 300 estimates is drawn and estimated from the DPG [sic] described on pp. 4-5. To these population, Reed, Florax and Poot (2015) apply multiple filters to sample selectively, pp.5-6. Lastly, alternative meta-analyses are performed upon these selected samples from their population of 300 estimates.

The statistical issue here is that the "true" population mean will differ in all cases from the "true" effect, α, listed in the left column of their tables (e.g., Table 2, 3, 5, 6). This difference is the result of the way Reed, Florax and Poot (2015) choose to generate their populations by random sampling 300 estimates from their estimation model and DPG (p.4). The mean for each population of 300 is the "true" population effect that all meta-analysis methods seek to estimate, not α. In statistics, sample estimates are only meant to represent the population from which they are drawn (the population mean of these 300 estimates) and can only be judged relative to this population mean. Comparing them to some remote 'true' effect, α, is not a valid way to calculate MSE or type I errors."

(ii) ***Multi-level diagram***. The diagram below conceptualizes the different levels the data pass through on the way to showing up in the meta-analyst's data set.

LEVEL	***DGP/Estimates***	***Comments***
LEVEL 1: Data-generating process (DGP)	$y_t = 1 + \alpha\, x_t + e_t,$	α is the "true" effect of x on y
	↓	
Individual observations	$(y_{it}, x_{it}), t=1,2,...T_i$	The *i* subscript identifies the observations as belonging to the *ith* study. *Ti* is the number of observations in the *ith* study.
	↓	
LEVEL 2: Individual studies	Estimate: $y_{it} = \beta_0 + \beta_1\, x_{it} + e_{it},$	NOTE: Individual studies draw observations from the same DGP
	↓	
LEVEL 3 (unobserved): Pre-Publication Bias Sample	$\hat{\beta}_{11}, \hat{\beta}_{12}, \dots, \hat{\beta}_{1N}$	The different $\hat{\beta}_{1i}$ are the estimates of the effect of x on y from the different studies. Each estimate is measured with standard error, SE$(\hat{\beta}_1)_i$
	↓	
LEVEL 4: Published studies = Post-Publication Bias Sample	$\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_M, M \leq N$	It is these estimates that the meta-analyst uses to obtain estimates of *a*

LEVEL 1 is the DGP that produces the individual estimates. The conceptual diagram assumes a “fixed effect” framework, where the effect of x on y is given by a constant, α. (Allowing more complex cases such as Random Effects or Panel Random Effects does not substantively change the multi-level representation above.)

The DGP “produces” a sample of observations that are collected by the *ith* study.

In LEVEL 2, the *ith* study estimates the effect of x on y and obtains an estimate, $\hat{\beta}_{1i}$, with standard error, SE$(\hat{\beta}_1)_i$.

LEVEL 3 is unobserved. It is the collection of all estimates that have been made. Consistent with the literature, Reed, Florax, and Poot (2015) – henceforth RFP -- call this a “population.“ But it is not a population in the statistical sense. It is a sample. It is more accurately called the “Pre-Population Bias sample“ of estimates.

LEVEL 4 is what is observed by the meta-analyst post-Publication Bias. This is the sample of estimates that the meta-analyst analyses, recognizing that it is possibly subject to sample selection due to Publication Bias.

(iii) ***<u>Why I believe our approach is correct</u>.*** The yellow-highlighted sections in the excerpt of Part (i) identify the point of disagreement between Stanley and RFP. Stanley believes that the relevant population parameter to use for MSE calculations and hypothesis testing is $\bar{\hat{\beta}}_1 = \frac{\sum_{i=1}^{N}\hat{\beta}_{1i}}{N}$, the mean of estimated effects in the Pre-Publication Bias sample. RFP asserts that the relevant population parameter is α, the true effect of x on y.

Frankly, it is difficult to understand why any researcher would be interested in estimating $\bar{\hat{\beta}}_1$, the mean of estimated effects in the Pre-Publication Bias sample. Meta-analysts aggregate estimates from individual studies, *<u>each of which is concerned with estimating α, the true effect of x on y</u>*. The idea is that by aggregating individual estimates of α, the meta-analyst can get an overall, “better” estimate of α.

This should be obvious. But just in case it’s not, consider how meta-analytic estimators such as Fixed Effects are constructed. They are typically weighted by SE$(\hat{\beta}_1)_i$, the standard error of the estimate of $\hat{\beta}_1$ from study *i*. The idea being, that studies with more precise estimates of $\hat{\beta}_1$ *<u>provide more precise estimates of α, the true effect of x on y</u>.* There would be no need to employ this weighting scheme if one was interested in estimating the mean of the “Pre-Publication Bias sample” of estimates.

RFP is designed to address the question, which meta-analytic estimation procedure performs "best" in identifying the true effect of x on y. This true effect is the population parameter α that meta-analysts are ultimately interested in knowing.

This is not just RFP's interpretation. One of the best references on meta-analytic procedures is Stanley and Doucouliagos' 2012 book, *Meta-Regression Analysis in Economics and Business*. This is exactly how they interpret the estimates they derive from applying the procedures analysed in our paper.

For example, on page 62 of their text, they write: "Notice the coefficients on $1/SE_i$ in TABLE 4.1. Testing H_0: β_0 serves as *a test of whether or not there is a genuine underlying empirical effect*..." [emphasis added]. By "genuine underlying empirical effect," they are referring to the "true effect" of x on y (cf. Equation 5 in RFP).

Further on, in speaking of the estimate they obtain from a meta-analysis of the price elasticity of water, they write (page 63): "To achieve, say, a 50 percent reduction in residential water consumption, our corrected elasticity implies that prices would need to be raised by 610 percent."

These statements make clear that Stanley and Doucouliagos (2012) are not interested in using meta-analysis to obtain an estimate of the mean of the "Pre-Publication Bias sample" of estimates, but of the true effect of x on y -- the population parameter α.

I agree with Stanley that referring to the collection of pre-Publication Bias studies as a "population" is misleading. The revised version of the paper will omit all references to the Pre-Publication Bias sample as a "population" in order to avoid confusion.