

Correcting Errors in Turkington et al. (2014): Taking Criticism Seriously

Smits, T., Lakens, D., Ritchie, S.J., Laws, K. R.

Published April 2015 in

Journal of Nervous and Mental disease

Turkington et al. (2014) reported an exploratory trial in which they had case managers follow a 5-day training program in Cognitive Behavioral Therapy for Psychosis (CBTp), which they then applied in an intervention-only study to 38 patients diagnosed with schizophrenia. We (Smits et al., 2014) identified errors in Turkington et al.'s (2014) reporting of critical results. Turkington (2014) then wrote a rejoinder to our commentary in which he stated that he "[could] confirm that the findings have been accurately reported in [the] published article".

We were surprised to read this rejoinder, given that some of the errors we pointed out were unambiguous. Subsequent correspondence with the authors and the journal editor made clear that Turkington had checked the reported results against a summary file with statistics tables, but not, importantly, against the raw data. Turkington and colleagues then kindly provided us with the data so that we could calculate the correct statistics. Upon doing so, we confirmed that all the errors identified by Smits et al (2014) were indeed errors. In addition, we observed that the reported effect sizes in Turkington et al. (2014) were incorrect by a considerable margin. To correct these errors, Table 2 and all the figures in the paper need to be changed. The sentence in the Abstract where effect sizes are specified needs to be re-written.

The reported significant treatment effects in Turkington et al. (2014) remain statistically different from zero, but the Tables and Figures presented do not correctly communicate the data. This is especially problematic because Turkington et al. (2014) stated that their results could "now be used to calculate the power required for a larger randomized controlled trial" (p. 33). With incorrect effect sizes, this is not possible. Moreover, if this study were to be included in a future meta-analysis, it would lead to an overestimation of the true effect size.

When attempting to reproduce the erroneous effect size calculations by Turkington et al. (2014), our best guess is that the authors entered the wrong sample size (18 instead of 38) when calculating their Cohen's d values. We were not successful in reproducing the errors in the calculation of the confidence intervals, and they certainly do not conform to any standard

statistical methodology. In addition, as we stated in our original commentary (Smits et al., 2014), Turkington et al.'s Table 2 (included below) contains some data that are *statistically impossible* (e.g., the lower-bound value for QPR intrapersonal being the same as the point estimate), and several confidence intervals so wide they are *nearly impossible*. Turkington (2014) did not address these obvious issues in his rejoinder. Our re-analyses based on the raw data confirmed all numbers in Table 2 are incorrect.

TABLE 2. Cohen's *d* Effect Sizes With CIs for Primary and Secondary Outcomes

	Cohen's <i>d</i> Effect size (CI)
Auditory hallucinations	0.58 (-2.91 to 4.67)
Delusions	-0.07 (-1.36 to 1.46)
CPRS total	1.60 (-2.29 to 5.07)
CPRS negative symptoms	0.87 (-0.02 to 1.62)
CPRS schizophrenia	0.63 (-0.84 to 1.68)
CPRS anxiety	0.72 (-0.35 to 1.73)
CPRS depression	1.12 (- 0.35 to 1.73)
PSP total	-0.21 (-5.19 to 5.44)
QPR intrapersonal	-0.45 (-0.45 to 3.10)
QPR interpersonal	-0.40 (-1.40 to 0.67)

Correct Cohen's d_z values calculated from the *t*-tests in the raw data shared with us are reported below. We also report the means and standard deviations (which were not included in the original study), as these are essential for a proper interpretation of the data. Note that in addition to the errors we reported in Smits et al.(2014), it is also clear the reported signs of the effect sizes were often incorrect (e.g., positive effect sizes where there was in fact a score reduction).

Treatment effect (pre-measure minus post-measure)

	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>	d_z	d_{av}
Auditory hallucinations	-3.50	8.75	-2.47	37	0.018	-0.40	-0.29
Delusions	0.26	5.24	0.31	37	0.759	0.05	0.06
CPRS Total	-9.26	8.32	-6.86	37	<.001	-1.11	-0.78
CPRS negative sympt.	-1.47	2.42	-3.75	37	0.001	-0.61	-0.56
CPRS schizophrenia	-1.76	4.11	-2.65	37	0.012	-0.43	-0.44
CPRS anxiety	-1.61	3.16	-3.13	37	0.003	-0.51	-0.48
CPRS depression	-2.71	3.62	-4.62	37	<.001	-0.75	-0.76

PSP total	2.89	19.51	0.92	37	0.366	0.15	0.17
QPR interpersonal	.71	2.52	1.74	37	0.091	0.28	0.21
QPR intrapersonal	2.94	9.26	1.91	35	0.065	0.32	0.25

Cohen's d_z controls for intra-subject variability, and is therefore often seen as an overestimation of the 'true' effect size (compared to a between subject design, see Lakens, 2013). It is more typical to report Cohen's d_{av} which can be used to compare effect sizes from within- and between-subjects designs in a meta-analysis. Bootstrapped confidence intervals for within-subject designs should be calculated, and several free statistical software packages exist to perform such calculations (e.g., ESCI software by Cumming, 2012). For example, the correct (unbiased, using Hedges' g correction) Cohen's d_{av} and 95% CI for CPRS schizophrenia would be Cohen's $d_{av} = -0.43$, 95% CI [-0.78, -0.10]. We leave the calculations for the remaining confidence intervals to the authors. If the originally reported $d_z = 0.63$ for CPRS schizophrenia would be used in an a-priori power analysis, 80% power would be achieved with a sample of 22 individuals, whereas when the correct effect size of $d_z = -0.43$ is used, 45 individuals are needed.

At the very minimum, one would have expected Turkington's (2014) rejoinder to give details about how effect sizes were calculated (as we suggested in our commentary), instead of a denying that anything was wrong. Indeed, some of the statistical points we made are self-evidently incorrect and do not require any re-analysis of the data. For example, if all 95% confidence intervals around Cohen's d include zero (such as in Turkington et al., 2014), then all effects are non-significant. The fact that the editor accepted and published such a rejoinder to our commentary is troublesome. We expected more of an effort to make sure the published results were accurate, especially since the errors we pointed out were so egregious.

To conclude, our recommendation for the journal and the authors would now be to acknowledge there are clear errors in the original Turkington et al. (2014) manuscript, and either accept our corrections, or publish their own corrigendum. Moreover, we urge authors, editors, and reviewers to be rigorous in their research and reviewing, while at the same time being eager to reflect on and scrutinize their own research when colleagues point out potential errors. It is clear that the authors and editors should have taken more care when checking the validity of our criticisms. The fact that a rejoinder with the title 'A Response to Ritchie et al. [sic]' was accepted for publication in reply to a letter by Smits et al. (2014)

gives the impression our commentary did not receive the attention it deserved. If we want science to be self-correcting, it is important that we follow ethical guidelines when substantial errors in the published literature are identified.

References

- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. doi:10.3389/fpsyg.2013.00863.
- Smits, T., Lakens, D., Ritchie, S. J., & Laws, K. R. (2014). Statistical errors and omissions in a trial of cognitive behavior techniques for psychosis: Commentary on Turkington et al. *The Journal of Nervous and Mental Disease, 202*, 566. doi: 10.1097/NMD.0000000000000161
- Turkington, D., Munetz, M., Pelton, J., Montesano, V., Sivec, H., Nausheen, B., & Kingdon, D. (2014). High-Yield Cognitive Behavioral Techniques for Psychosis Delivered by Case Managers to Their Clients With Persistent Psychotic Symptoms: An Exploratory Trial. *The Journal of Nervous and Mental Disease, 202*, 30-34.
- Turkington, D. (2014). The reporting of confidence intervals in exploratory clinical trials and professional insecurity: a response to ritchie et Al. *The Journal of Nervous and Mental Disease, 202*, 567.