Pitfalls With the Unquestioning Use of Statistics

To the Editor: With regard to the recent article by O'Regan and colleagues,[1] I share the authors' concern that patients with Alzheimer's disease are maintained on cholinesterase inhibitors for much longer periods than what the conclusions of randomized controlled trials (RCTs) allow and that clinical guidelines have been vague on this issue. While I welcome their research initiative to summarize the effects of cholinesterase inhibitor discontinuation, I find their synthesis of the summary estimates flawed.

In any meta-analysis, heterogeneity of the effect sizes of individual RCTs is assessed, followed by the calculation of the average effect size using an appropriate model (fixed or random effects), based on the extent of heterogeneity. This was precisely what O'Regan et al did, and at first glance, this is all well and good. However, on closer scrutiny, the unquestioning use of the result heterogeneity $I^2 = 0\%$ raises concerns.

To better understand the issues with the authors' erroneous interpretation of I^2, we need to revisit the evolution of the 2 most commonly used measures for heterogeneity: Q statistic and I^2. The Q test was established in 1954 to evaluate heterogeneity. Its shortcoming is its poor power to detect heterogeneity when the meta-analysis has few studies.[2] The I^2 was subsequently developed[3] to overcome this issue. However, new evidence indicates that both tests perform similarly—just like the Q test, the I^2 index has low power with a small number of studies.[4]

Applying these findings to this meta-analysis, the number of studies used is 3 (for outcome measure of Neuropsychiatric Inventory) or 5 (Mini-Mental State Examination). Hence, the I^2 estimate here is likely underpowered to detect heterogeneity. If one scrutinizes the characteristics of the RCTs used for the meta-analysis, this conclusion makes sense, intuitively. Indeed, the authors have astutely pointed out that "the lack of heterogeneity…is surprising considering the variation between study designs."[1(p e1429)] It is thus a pity that they did not follow through with their observation, but based their choice of a fixed effects model on the finding of $I^2 = 0\%$. This highlights the perils of the unquestioning use of statistics.

Given the current state of affairs in meta-analyses, whereby the median number of studies is 7 in Cochrane Reviews and 12 in published journals,[5,6] the unsavory ingredients for an underpowered I^2 estimate are likely to feature prominently in the synthesis of summary estimates. The unfortunate example seen in this article should hence not be viewed in isolation.

Moving forward, what could be done for better reporting and interpretation of I^2 in meta-analyses? Suggestions include the reporting of I^2 with its 95% confidence intervals, the routine use of the random effects model[4] regardless of the point estimate of I^2, and the use of sensitivity analyses based on a plausible spectrum of degrees of heterogeneity.

While no statistical maneuver is perfect, it is pertinent that measures are accurately presented to highlight the existing vagaries of biostatistics. Otherwise, the undiscerning use of statistics may turn into a mere number-crunching exercise that could ultimately misinform.

References

1. O'Regan J, Lanctôt KL, Mazereeuw G, et al. Cholinesterase inhibitor discontinuation in patients with Alzheimer's disease: a meta-analysis of randomized controlled trials. *J Clin Psychiatry*. 2015;76(11):e1424–e1431.
2. Alexander RA, Scozzaro MJ, Borodkin LJ. Statistical and empirical examination of the chi-square test for homogeneity of correlations in meta-analysis. *Psychol Bull*. 1989;106(2):329–331.
3. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539–1558.
4. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, et al. Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychol Methods*. 2006;11(2):193–206.
5. Moher D, Tetzlaff J, Tricco AC, et al. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med*. 2007;4(3):e78.
6. Tricco AC, Tetzlaff J, Pham B, et al. Non-Cochrane vs Cochrane reviews were twice as likely to have positive conclusion statements: cross-sectional study. *J Clin Epidemiol*. 2009;62(4):380–386, e1.

Shian Ming Tan, MMed[a]
tan.shian.ming@singhealth.com.sg

[a]Department of Psychiatry, Singapore General Hospital, Singapore

Potential conflicts of interest: None.

Funding/support: None.

J Clin Psychiatry 2016;77(8):e1005
dx.doi.org/10.4088/JCP.16lr10666

Dr Mazereeuw and Colleagues Reply

To the Editor: We thank the Editor for the opportunity to respond to Dr Tan's letter, which used our recently published meta-analysis on cholinesterase inhibitor (ChEI) discontinuation in patients with Alzheimer's disease[1] to highlight the limitations of statistical indicators of heterogeneity.

We accept Dr Tan's view that improving statistical management of heterogeneity in meta-analyses with only a small number of studies may be an important area of future research. Indeed, we acknowledged the small number of studies, as well as the small sample sizes in the included studies, as limitations of our meta-analysis. However, Dr Tan's suggestion that we used inappropriate methodology and unquestioningly presented our findings is not well supported.

The *Cochrane Handbook for Systematic Reviews of Interventions*[2] recommends that authors compare the effects of their intervention using both fixed- and random-effects models in order to identify the potential influence of small study effects. The *Handbook* suggests that if no differences in the intervention effect are detected between the fixed- and random-effects models, then authors may conclude that small study effects had little influence on the outcome. As such, we investigated the effect of ChEI discontinuation using both fixed- and random-effects models to determine which model was more appropriate to report in the article. Although we did not publish the results of the random-effects models in that article, we present them here to demonstrate the consistency of our findings.

Using a random-effects model, our analysis shows that patients in the ChEI discontinuation group experienced significantly greater cognitive decline than those remaining on ChEIs (standardized mean difference [SMD] = −0.29 [95% CI, −0.45 to −0.13] $P < .001$; heterogeneity: $\chi^2 = 2.62$, $I^2 = 0$, $P = .62$). Similarly, using a random-effects model, the ChEI discontinuation group experienced significantly greater neuropsychiatric symptoms than those remaining on ChEIs (SMD = −0.32 [95% CI, −0.51 to −0.12] $P = .001$; heterogeneity: $\chi^2 = 0.94$, $I^2 = 0$, $P = .63$). The results seen when using random-effects models are identical to those when using fixed-effects models, supporting our findings and reinforcing the lack of heterogeneity. This consistency is not at all surprising given the lack of heterogeneity reported in the article.

Contrary to Dr Tan's claim that we did not follow through on our observation of heterogeneity (despite none being detected using either model), we did explore potential modifiers of the observed treatment effects and published those findings as part of our original article. Specifically, we showed that the effect of ChEI discontinuation on cognitive changes and neuropsychiatric symptoms was not influenced by the mean patient age, sex, previous duration of ChEI treatment, duration of follow-up in each study, or the baseline severity of dementia. As such, we fully explored potential sources of heterogeneity to increase the transparency of our findings in light of the variability in study designs.

In summary, we conducted our meta-analysis according to established guidelines, using established statistical tests. We investigated heterogeneity appropriately and reported our findings accordingly. Therefore, we suggest that Dr Tan's criticism of our study is not well supported and brings unnecessary skepticism to our findings. Our methods were rigorous, and we achieved our goal of summarizing the available literature and bringing attention to the need for more studies in this area.

References

1. O'Regan J, Lanctôt KL, Mazereeuw G, et al. Cholinesterase inhibitor discontinuation in patients with Alzheimer's disease: a meta-analysis of randomized controlled trials. *J Clin Psychiatry*. 2015;76(11):e1424–e1431.
2. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0*. The Cochrane Collaboration; 2011. http://handbook.cochrane.org/. Updated March 2011.

Graham Mazereeuw, PhD[a,b]
Krista L. Lanctôt, PhD[a,b,c]
Nathan Herrmann, MD, FRCPC[a,c]
nathan.herrmann@sunnybrook.ca

[a]Neuropsychopharmacology Research Group, Hurvitz Brain Sciences Research Program, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada
[b]Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, Canada
[c]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada

Potential conflicts of interest: None.

Funding/support: None.

J Clin Psychiatry 2016;77(8):e1006
dx.doi.org/10.4088/JCP.16lr10666a