



August 30, 2018

Robert A. Gross, MD, PhD
Editor-in-Chief
Neurology
201 Chicago Avenue
Minneapolis
MN 55415

Dear Dr Gross,

Neurology recently published a research article by Schultz *et al.* reporting on the safety of tetrabenazine (TBZ) use and incident depression and suicidality in Huntington's disease (HD) [1]. Schultz *et al.* sought to determine whether TBZ use is associated with an increased incidence of depression and/or suicidal ideation. On the basis of observational evidence, they conclude that TBZ is *not* associated with an increased incidence of depression or suicidality, and that TBZ may be *safe* to use in patients with HD who have a history of depression.

We believe the results and conclusions of Schultz *et al.* are flawed and may pose a risk to patient safety.

We have submitted a Comment in response to this article outlining our headline concerns. However, given the space restrictions imposed for this submission format and the gravity of our concerns, we here outline the identified issues in detail and provide relevant context to assist in your editorial assessment.

The conclusions drawn by the authors conflict with evidence from a randomized, double-blind, placebo-controlled trial [2] – the gold standard for establishing causal inference. It is worth noting that the FDA has ascribed a **black box warning to TBZ for depression and suicidality** [3].

Aside from this conflict with experimental evidence, we identify several issues with the Schultz *et al.* paper that severely undermine confidence in the conclusions drawn.

These include problems relating to study design and analytical strategy, quality of reporting, interpretation of reported results, and internal inconsistencies in reporting, as outlined in detail below.

DESIGN AND ANALYSIS

The authors' decision to focus exclusively on data gathered at a single baseline visit, despite the availability of prospectively collected repeated measures data within Enroll-HD, introduces several problems relating to accurate and unbiased characterization of exposure status, outcome status, and the temporal association between exposure and outcome. These are described below, alongside other design and analytical flaws.

TBZ user definition and potential survival bias:

The basis for ascertaining exposure to TBZ introduces potential survival bias. The authors consider only TBZ use at time of baseline assessment. Only individuals that have been exposed and who have tolerated TBZ are captured as "users." All users that were exposed and abandoned treatment eventually (for example due to depression or suicidality) are not counted. This bias would serve to induce a spurious negative association between TBZ use and depression/suicidality (i.e., bias estimates towards the null or in the direction of an apparent protective effect). We might expect this bias to be strongest in those with a history of depression (who might be at greatest risk of the adverse effects of TBZ), which is consistent with the results stratified on history of depression presented by Schultz et al.

Characterization of depression/suicidality incidence:

To address the research question at hand, only incident cases of depression/suicidality occurring after exposure to TBZ should be considered. The authors could not adequately establish such temporality given their decision to focus exclusively on data gathered at a single baseline visit, despite availability of prospectively-collected data. We do not know if the depressive or suicidality symptoms are truly incident (detected first time in that visit) or prevalent (carried forward from assessments prior to study entry).

Separately, the number of individuals characterized as experiencing incident depression/suicidality may have been underestimated. The instrument used by the authors to establish incident depression was the PBA-s (not the PBA as the authors erroneously report). The PBA-s is a clinician-rated scale that uses a structured interview to elicit information regarding the prior four weeks. The authors relied on items assessing the current severity and frequency of symptoms. This would not have captured individuals who were currently using antidepressants (assuming treatment was effective). This is particularly pertinent given observed differences in antidepressant use between exposure groups (antidepressant use was considerably higher in the full sample TBZ user group relative to the TBZ nonuser group; 55.8% vs. 47.9%, $p = 0.001$).

Adjustment for antidepressant use and potential collider bias:

Antidepressant use may result from TBZ use (given the link between TBZ and depression), or from other causes of depression. Collider bias occurs when a variable caused by two other variables is adjusted for or stratified on. Adjusting for antidepressant use, as the authors have done, may therefore introduce collider bias, which would induce a false negative association between TBZ use and depression (i.e., spuriously indicating a potential *protective* effect of TBZ on depression). This is compounded by the cross-sectional nature of the analysis, which precludes the definition of incident depression and suicidality, despite prospective data being available within Enroll-HD. Notably, the negative associations are seen in the history positive group, where any collider bias may be expected to be greatest, whereas positive associations (i.e., reflecting increased risk of depression in TBZ users, consistent with experimental evidence) are seen in the history negative group. This bias potentially compounds the survival bias described above.

QUALITY OF REPORTING AND INTERPRETATION OF RESULTS

The design and analysis flaws described above are sufficient to question the validity of the results and conclusions. The quality of reporting and interpretation of results further exacerbates the problem.

Conflicting internal results:

The statistical results presented for risk of depression in individuals with no history of depression are reported inconsistently within the paper; in Table 2, $OR = 1.59$, $p = 0.182$, in the text of the results section, $OR = 1.39$, $p = 0.334$.

Failure to report unadjusted results:

Only results from fully-adjusted models are reported in the paper. Results from unadjusted and partially-adjusted models should be reported as best practice. Here this would have enabled assessment of potential bias introduced by adjusting for antidepressant use (see above).

Lack of correction for (or consideration of) multiple testing:

The p -value criterion of 0.05 selected to indicate statistical significance affords a very low level of stringency when multiple comparisons are made. The results of (>) six analyses are reported. If a simple Bonferroni adjustment for six tests is made to the alpha level ($.05 / 6 = .008$) then none of the results reported reach statistical significance.

Failure to consider observed effect and sample size in comparing and discussing results:

The authors fail to consider sample size when discussing results, and ignore effect magnitude, commenting only on the direction of effect when accompanied by a 'significant' result. This is problematic for many reasons, not least the differing sample sizes on which the results are based. The results generated for participants with a personal history of depression are based on $n=2,757$. The results generated for participants *without* such a history are based on a sample of less than half that size ($n=1,334$). The direction of effect differs between groups (OR = 0.71 vs. OR = 1.59 respectively), but the magnitude of effect is similar (actually larger in the latter group). Critically, the association observed in the latter group (i.e., TBZ *increases* incidence of depression) is ignored as the p -value is > 0.05 . Should the sample sizes have been equivalent, the deleterious effect noted may have exceeded the alpha level selected to claim statistical significance.

Ascribing causality (or lack of causality) in observational epidemiology is notoriously challenging: TBZ users will differ from non-users in many ways, and statistical adjustment will not be sufficient to account for this (and may introduce other biases as described above). For the various reasons we describe, we consider the conclusions presented by Schultz *et al.* regarding the risk of depression and suicidality associated with TBZ use as flawed and potentially dangerous to patients.

It is critically important to correct this view given the clear implications for patient safety.

CHDI would like to foster the use of Enroll-HD and other large data sets to promote the public health good. Utilization of available resources to this end has also been recently encouraged both by the FDA and EMA. Indeed, pharmacoepidemiological studies can be informative in evaluating/establishing the magnitude of safety risks identified in clinical trials. Enroll-HD, a high value resource of rich, longitudinal data, may serve this purpose. However, **it is critical that appropriate methodologies are adopted**. To do otherwise not only risks misleading research but, in extreme cases such as this, could also compromise patient safety.

CHDI Foundation funds an Independent Statistical Standing Committee (ISSC; <https://chdifoundation.org/independent-statistical-standing-committee/>) to provide expert advice regarding the design and analysis of HD research studies, including observational studies based on the Enroll-HD data set.

We hope that you will publish our Comment in order to highlight these issues so that the readership of *Neurology* can evaluate these results appropriately. We would welcome a response from Schultz *et al.*, preferably with the inclusion of results presented with and without statistical adjustment, so that the impact of these can be better evaluated. However, ultimately, we believe the methodological approach to be fundamentally flawed and we are disappointed that the manuscript reviewers did not identify these problems (while acknowledging that peer review is imperfect and imprecise). We would be happy to work

with *Neurology* on potential solutions that would help guard against the publication of similarly questionable studies in future.

Sincerely,

Cristina Sampaio¹, Jennifer Ware¹, Malcolm Macleod^{2,3}, E. J. Wagenmakers^{3,4}, Marcus Munafo^{3,5,6}

1. CHDI Management/CHDI Foundation, Princeton, NJ, United States.
2. University of Edinburgh, Edinburgh, United Kingdom.
3. Independent Statistical Standing Committee (<https://chdifoundation.org/independent-statistical-standing-committee/>)
4. University of Amsterdam, Amsterdam, The Netherlands.
5. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom.
6. School of Experimental Psychology, University of Bristol, Bristol, United Kingdom.

References

1. Schultz, J.L., Killoran, A., Nopoulos, P.C., Chabal, C.C., Moser, D.J., and Kamholz, J.A. Evaluating depression and suicidality in tetrabenazine users with Huntington disease. *Neurology* 91, 3 (2018), e202–e207
2. Huntington Study Group. Tetrabenazine as antichorea therapy in Huntington disease: a randomized controlled trial. *Neurology* 66, 3 (2006), 366–372
3. https://www.accessdata.fda.gov/drugsatfda_docs/label/2011/021894s005lbl.pdf (accessed August 2018)