



Rebuttal to Gravely et al. Validity of PTSD diagnoses in VA administrative data: Comparison of VA administrative PTSD diagnoses to self-reported PTSD Checklist scores. *J Rehabil Res Dev.* 2011; 48(1):21–30. Available from: <http://www.rehab.research.va.gov/jour/11/481/pdf/gravely.pdf>

Posttraumatic stress disorder (PTSD) is a major focal point for Department of Veterans Affairs (VA) research and policy. Spanning more than 10 years and covering over 7 million enrollees, the VA administrative data repository represents an irreplaceable resource to optimize care for veterans with PTSD. Yet to make use of this resource, we must understand how to appropriately identify individuals with PTSD. With this in mind, we read the recent article by Gravely and colleagues with great interest [1]. To briefly summarize, the authors examined the validity of two different administrative data-derived definitions of PTSD using a mailed, self-reported PTSD Checklist (PCL) score with 50 as the gold standard [2]. The authors compared the positive predictive value (PPV) of one or more administrative records versus two or more administrative records indicating PTSD. The principal finding was that PPV improved from 75 percent when requiring just one PTSD code to 82 percent when requiring at least two. From this, the authors concluded that, “To select a sample of veterans with more definitive PTSD from administrative data, researchers should select those veterans with at least two PTSD

diagnoses as opposed to at least one [1].” However, we are concerned that this conclusion, particularly the term “definitive PTSD,” could be misunderstood and misapplied.

Our primary concern is that the purported false positive cases are composed of two subgroups that address different validation issues. Without this distinction, the findings are more challenging to interpret. The first subgroup of false positives consists of veterans where PTSD was not diagnosed clinically, but an errant PTSD code was present in the administrative data. This subgroup addresses validity in terms of whether the administrative data accurately reflects the clinical care that was delivered. The second subgroup consists of veterans given a PTSD diagnosis by a VA clinician, and thus the administrative codes accurately reflect care, but the veteran does not actually have PTSD. This subgroup addresses validity in terms of the accuracy of VA clinicians in diagnosing PTSD. Ideally, two PPV estimates would be known: (1) the PPV of administrative PTSD codes for clinician diagnosed PTSD, and (2) the PPV of clinician diagnosed PTSD for PTSD diagnosed by gold standard evaluation. This is an important distinction because the target population for many VA investigators may be veterans clinically diagnosed with PTSD, and whether these individuals met formal diagnostic criteria is less relevant. Such investigators would only be interested in the first PPV estimate, and not a conflation of the two.

While we cannot be certain what proportion of the false positives fell

into these respective subgroups, some indirect evidence can be gleaned from the distribution of PCL scores in Table 1 of the original article. There was a small percentage of cases (2–4%) with scores below 30, suggesting a small base rate of errant PTSD codes. Therefore, the PPV for a single administrative PTSD code in identifying clinician diagnosed PTSD, which could be more relevant to many VA investigators, may be upwards of 95%. In contrast, the majority (>80%) of false positive cases had some appreciable level of PTSD symptoms (PCL scores from 30–49), suggesting that the PPV estimates were largely driven by disagreement between the PCL cutoff score and clinician coded PTSD. However, as fully acknowledged by the authors, the PCL screening questionnaire does not yield a gold standard diagnosis. The most direct interpretation of this study is that requiring two PTSD codes generated a marginally more symptomatic group, but in the absence of a gold-standard diagnostic evaluation, it remains unknown whether less symptomatic patients were less likely to have PTSD. There are probably remediable barriers to improving the accuracy of diagnosing PTSD, a vital area for VA research and practice. However, it is not clear whether more confidence should be placed in PCL scores than VA clinical practice for diagnosing PTSD, thus making the PPV estimates reported in this study difficult to interpret.

A related concern is that the authors did not discuss the potential impact of selection bias created by excluding veterans with only one

PTSD code. Table 3 of Gravely et al. shows that going from one to two diagnostic codes excluded 40 percent of the initial sample. As demonstrated by the authors, patients with two or more PTSD codes are different in many important ways from patients with only one code. Of particular concern, their PTSD symptoms could be more severe and chronic, and they probably have more complex medical comorbidities. Therefore, it is questionable whether results obtained from patient samples selected by this algorithm can be generalized to the broader PTSD population. For many research applications, a nominal increase in PPV may be offset by the loss in generalizability. Including this important trade-off in the abstract would have allowed the reader to make a more balanced evaluation of the benefits and consequences of applying the two algorithms.

We conclude that PTSD case definitions derived from administrative data should be selected based on project-specific objectives. While the authors acknowledge this point in their discussion, we are concerned by the omission of important information from the abstract. Many readers will stop at the abstract and may come away with the singular conclusion that two diagnostic codes should always be required to identify PTSD in VA administrative data, with limited understanding of the generalizability of these findings. Clearly, requiring two PTSD codes yields a more symptomatic PTSD cohort. However, it is unknown whether veterans with sub-threshold PCL scores truly did not have PTSD, and the potential consequences of excluding these individuals are substantial. We share common ground with the authors in stating that different PTSD case definitions pro-

duce different clinical populations and that characterizing these differences remains a crucial research need. Despite our differences in interpretation, Gravely and colleagues should be commended for providing the VA PTSD research community with this important contribution.

**Brian C. Lund, PharmD, MS;^{1*}
Thad E. Abrams, MD, MS¹⁻²**

¹Center for Comprehensive Access & Delivery Research and Evaluation at the Iowa City VA Medical Center, Iowa City, IA; ²Division of General Internal Medicine, Department of Internal Medicine, University of Iowa Carver College of Medicine, Iowa City, IA

*Email: Brian.Lund@va.gov

REFERENCES

1. Gravely AA, Cutting A, Nugent S, Grill J, Carlson K, Spont M. Validity of PTSD diagnoses in VA administrative data: Comparison of VA administrative PTSD diagnoses to self-reported PTSD Checklist scores. *J Rehabil Res Dev.* 2011;48(1): 21–30. [PMID: 21328160] DOI:10.1682/JRRD.2009.08.0116
2. Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA. Psychometric properties of the PTSD Checklist (PCL). *Behav Res Ther.* 1996;34(8):669–73. [PMID: 8870294] DOI:10.1016/0005-7967(96)00033-2

RESPONSE

Dear Editor:

We would like to thank Drs. Lund and Abrams for their careful review and thoughtful comments about our article. Hopefully, we will be able to clarify the points they raise in their letter.

There were two main concerns raised in the letter. Their first concern is about potential error in the PTSD

administrative diagnoses, either by clerical error or clinician error. We agree that there is error and that these are two distinct sources of diagnostic error that we did not specify in our article. In our study, specific efforts were made to avoid clerical errors. Cases were not included in which a PTSD diagnosis was recorded as part of a “compensation and pension” evaluation or research visit. We also did not include cases that were given in ancillary medical clinics such as audiology clinics. We estimate that less than 5 percent of PTSD diagnoses in the larger primary sample were because of clerical error given the number and type of appointments that were changed after quarterly review (2.6%). This information was recently gathered as part of the main study on which this article was based.

The potential for clinician diagnostic error is an unstated assumption of our article, and is part of the rationale for the study. Although we may not have been sufficiently clear, the goal of the study was to identify an algorithm that could be used to identify a sample of veterans with probable PTSD in administrative data given that clinical diagnoses are not always accurate (particularly if one is interested in a current diagnosis vs a lifetime diagnosis). If clinical diagnoses were always accurate, no algorithm would be needed. The authors’ concern rests on our use of a PTSD Checklist (PCL) score over 50 as our confirmatory criterion. In part, the disagreement is semantic in nature. While we explicitly state that the PCL is not a gold standard, we also use language suggesting that the PCL provides diagnostic validation. It is understandable, therefore, that our intent might not have been clear. However, Lund and Abrams take issue with our use of the

PCL at all since we risk a potentially high false-negative rate using a cutoff of 50. If we are using the PCL as a diagnostic tool, their concern would be well placed—it is a strict criterion even when the base rate of PTSD is high, as would be true in our sample [1–2]. The study goal was not to identify all possible cases of “true” PTSD, but to demonstrate the effects of differential selection criteria depending on how much diagnostic certainty was required by the study in question and showing variations by subgroups of interest. For example, a study designed to test a new pharmacologic agent to treat symptoms of PTSD would be better served by a sample that is more symptomatic and certain to meet current criteria for PTSD. In contrast, a study examining the relationship between psychiatric diagnosis and adherence to a diabetes treatment in a large sample would be better served using a single PTSD-related appointment criterion. Importantly, we agree

with the authors that sampling from administrative data should be based on “project-specific objectives.”

The second point raised by the authors rests on whether our algorithm is generalizable to the broader PTSD population. Specifically, they state, “. . . patients with two or more PTSD codes are different in many important ways from patients with only one code. Of particular concern, their PTSD symptoms could be more severe and chronic, and they probably have more complex medical comorbidities. Therefore, it is questionable whether results obtained from patient samples selected by this algorithm can be generalized to the broader PTSD population.” We do not disagree with this assertion, and note that whether this is a problem depends on the goal of the study. Our study helps researchers understand in what ways their sample might be biased by using one versus two PTSD-related appointments for their sample selection.

Again, we would like to thank Lund and Abrams for their insightful comments. It is our hope that this discussion spurs more research on this very important topic.

Amy A. Gravely, MA

Minneapolis VAMC, Center for Chronic Disease Outcomes Research, Minneapolis, MN

Email: amy.gravely@va.gov

REFERENCES

1. Terhakopian A, Sinaii N, Engel CC, Schnurr PP, Hoge CW. Estimating population prevalence of posttraumatic stress disorder: An example using the PTSD checklist. *J Trauma Stress*. 2008;21(3):290–300. [PMID: 18553416] [DOI:10.1002/jts.20341](https://doi.org/10.1002/jts.20341)
2. McDonald SD, Calhoun PS. The diagnostic accuracy of the PTSD checklist: A critical review. *Clin Psychol Rev*. 2010;30(8):976–87. [PMID: 20705376] [DOI:10.1016/j.cpr.2010.06.012](https://doi.org/10.1016/j.cpr.2010.06.012)

