

## **Scoring consistency checks for the Clinician Administered PTSD Scale (CAPS-IV & CAPS-V)**

**Jonathan Rabinowitz, Alon A. Rabinowitz, Sara Freedman**

**Bar Ilan University, Israel**

### **Abstract**

**Background:** The CAPS is regarded as the "gold standard" in PTSD assessment. It is a structured interview that yields a categorical diagnosis of PTSD and also a measure of the severity of PTSD symptoms. Since a PTSD diagnosis in some settings could be connected to getting benefits, scoring inconsistencies may be more abundant with this rating scale as they not only reflect raters carelessness but intentional inaccurate reporting by the subject.

**Objective:** The objective of the current effort was to derive rating consistency flags for the CAPS-IV.

**Methods:** We deconstructed CAPS scoring instructions and anchors to identify potential scoring inconsistency flags. These inconsistency flags were reviewed and confirmed by expert raters. To test the ability of the flags to identify careless responses the flags were applied to Monte Carlo simulated data of 100,000 CAPS administrations.

**Results:** Twelve flags were derived (presented in Table 1). Two flags applied to most of the 17 symptom items (Flag 1: Frequency=0 & Intensity>0 and Flag 2: Frequency>0 & Intensity=0). The remaining 10 flags pertained to individual items. Five flags were rated as "High" flags, representing very probably or definitely incorrect rating, one as medium, reflecting probably incorrect rating. Flags were raised for 95% of the Monte Carlo simulated CAPS administrations, 78% of the administrations had 4 or more flags and 60% 5 or more.

**Conclusions:** Scoring consistency flags for the CAPS may be useful in the quest to improve reliability and validity of clinical trials. Modified flags are currently being developed to cover the CAPS-V. Further testing of flags using clinical trial data is planned to examine their potential impact on signal detection.

## Introduction

The CAPS is regarded as the "gold standard" in PTSD assessment. It is a structured interview that yields a categorical diagnosis of PTSD and also a measure of the severity of PTSD symptoms. It can be administered in 30-60 minutes by a trained rater.

To improve the reliability and validity of measurement in clinical trials, we previously developed consistency checks "flags" for the Montgomery-Asberg Depression Rating Scale (MADRS)(Rabinowitz et al. ,2019, 2022), Positive and Negative Syndrome Scale (PANSS) (Rabinowitz et al., 2017; 2021), Personal and Social Performance scale (PSP) (Rabinowitz et al, 2021), the Young Mania Rating Scale (YMRS) (Rabinowitz et al, 2024).

Since a PTSD diagnosis in some settings could be connected to getting benefits, scoring inconsistencies may be more abundant with this rating scale as they not only reflect raters carelessness but intentional inaccurate reporting by the subject.

## Aim

The objective of the current effort was to derive consistency flags for the CAPS-IV.

## Methods

The first 17 items of CAPS IV elicits ratings on Frequency (0=never; 1=once or twice; 2=once or twice a week; 3=several times a week; 4-daily or almost every day) and intensity (0=no distress; 1=mild; 2=moderate; 3=severe; 4=extreme) of symptoms and are used to compute a total severity score. The next 4 items measure duration, subject distress and functional impairment. A scoring algorithm is applied to these 21 items to arrive at a diagnosis. The CAPS also includes 3 global ratings and ratings of 5 associated features.

We deconstructed CAPS scoring instructions and anchors to identify potential scoring inconsistency flags. These inconsistency flags were reviewed and confirmed by expert raters.

To test the ability of the flags to identify careless responses the flags were applied to Monte Carlo simulated data of 100,000 CAPS administrations.

## Results

Twelve flags were derived (presented in Table 1). Two flags applied to most of the 17 symptom items (Flag 1: Frequency=0 & Intensity>0 and Flag 2: Frequency>0 & Intensity=0). The remaining 10 flags pertained to individual items. Five flags were rated as "High" flags, representing very probably or definitely incorrect rating, one as medium, reflecting probably incorrect rating. Flags were raised for 95% of the Monte Carlo simulated CAPS administrations, 78% of the administrations had 4 or more flags and

60% 5 or more. Two high flags, Flag 1 and 2 were raised in more than 85% of the administrations. Table 2 presents flags for both the CAPS-IV and CAPS-V and a comparison of the items on each scale.

## **Conclusions**

Scoring consistency flags for the CAPS may be useful in the quest to improve reliability and validity of clinical trials. Modified flags are currently being developed to cover the CAPS-V. Further testing of flags using clinical trial data is planned to examine their potential impact on signal detection.

## **Acknowledgments**

The research leading to these results has received support from the Elie Wiesel Chair at Bar Ilan University.

Previous versions of this material have been presented at ISCTM 2021 Autumn Conference

Jonathan Rabinowitz has received research grant(s) support and/or travel support and/or speaker fees and/or consultant fees from Janssen (J&J), Eli Lilly, Pfizer, Lundbeck, Roche, Abraham Pharmaceuticals, Pierre Fabre, Intra-cellular Therapies, Minerva, Takeda and Amgen. Alon A. Rabinowitz and Sara Freedman have no disclosures.

## References

Rabinowitz J, Schooler NR, Anderson A, et al. Consistency checks to improve measurement with the Positive and Negative Syndrome Scale (PANSS). *Schizophr Res.* 2017;190:74-76.

Rabinowitz, J., Schooler, NR, Brown, B. et al, Consistency checks to improve measurement with the Montgomery-Asberg Depression Rating Scale (MADRS), *Journal of Affective Disorders* 2019; 256: 143-147

Rabinowitz, J., Opler, M., Rabinowitz, A. A., Negash, S., Anderson, A., Fu, D. J., Williamson, D., Kott, A., Davis, L. L., & Schooler, N. R. (2021). Consistency checks to improve measurement with the Personal and Social Performance Scale (PSP). *Schizophrenia research*, 228, 529–533.  
<https://doi.org/10.1016/j.schres.2020.11.040>

Rabinowitz, J., Williams, J. B. W., Anderson, A., Fu, D. J., Hefting, N., Kadriu, B., Kott, A., Mahableshwarkar, A., Sedway, J., Williamson, D., Yavorsky, C., & Schooler, N. R. (2022). Consistency checks to improve measurement with the Hamilton Rating Scale for Depression (HAM-D). *Journal of affective disorders*, 302, 273–279.  
<https://doi.org/10.1016/j.jad.2022.01.105>

Rabinowitz, J., Young, R. C., Yavorsky, C., Williams, J. B. W., Sedway, J., Marino, P., Matteo, C., Mahableshwarkar, A., Kott, A., Hefting, N., Engler, J., & Brady, C. (2024). Consistency checks to improve measurement with the Young Mania Rating Scale (YMRS). *Journal of affective disorders*, 345, 24–31.  
<https://doi.org/10.1016/j.jad.2023.10.098>

Rabinowitz, J. & Rabinowitz, A. (2021). Outlier-response pattern checks to improve measurement with the Positive and Negative Syndrome Scale (PANSS). *Psychiatry Research*; Sep;303

Rabinowitz, J., & Rabinowitz, A. A. (2022). Outlier-response pattern checks to improve measurement with the Montgomery-Asberg depression rating scale (MADRS). *Journal of affective disorders*, 299, 444–448.  
<https://doi.org/10.1016/j.jad.2021.12.076>

Table 1 List of CAPS IV items and inconsistency FLAGS

<b>A. Identifiable traumatic event (required for diagnosis)</b>
Items in subsequent sections B, C & D rated on Frequency and Intensity <b>F=Frequency</b> (0=never; 1=once or twice; 2=once or twice a week; 3=several times a week; 4-daily or almost every day) <b>I=Intensity</b> (0=no distress; 1=mild; 2=moderate; 3=severe; 4=extreme)
Flag 1 (High): F=0 & I>0 applies to all 17 items in sections B,C & D Flag 2 (High): F>0 & I=0 applies to all 17 items except B3, C1, C2
<b>B. Reexperiencing symptoms (1 needed for diagnosis)</b> (1) B1- Intrusive recollections, (2) B2- Distressing dreams, (3) B3- Acting or feeling as if event were recurring (Flag 1 only); (4) B4-Cued psychological distress; (5) B5-Cued physiological reactions
<b>C. Avoidance and numbing symptoms (3 needed for diagnosis)</b> (6) C1- Avoidance of thoughts or feelings; (7) C2- Avoidance of activities, places, or people (8) C3-Inability to recall important aspect of trauma Flag 3: F>2 & I<3; F=4 & I<3; F=3 & I=1 (9) C4- Diminished interest in activities Flag 4: F>2 & I<2; F=4 & I<3; F=3 & I=1 (10) C5- Detachment or estrangement; (11) C6- Restricted range of affect; (12) C7- Sense of a foreshortened future
<b>D. Hyperarousal symptoms (2 needed for diagnosis)</b> (13) D1- Difficulty falling or staying asleep; (14) D2-Irritable behavior and angry outbursts; (15) D3-Difficulty concentrating; (16) D4-Hypervigilance; (17) D-5 Exaggerated startle response
<b>E. Duration of disturbance is more than 1 month (required for diagnosis)</b> (18) Onset of symptoms; (19) Duration of symptoms
<b>F. Significant distress or impairment in functioning</b> (20) Subjective distress; (21) Impairment in social functioning; (22) Impairment in occupational functioning
<b>Global Ratings</b> (23) Global validity; (24) Global severity; (25) Global improvement
<b>Associated features</b> (26) Guilt over acts of commission or omission; (27) Survivor guilt; (28) Reduction in awareness of surroundings; (29) Derealization; (30) Depersonalization
<b>Composite flags</b>
<b>Flag 5 (High)</b> Items 20- Subjective distress and 24- Global severity need to be similar. If they are not, we would expect it to have expression in 23- Global validity. For example, If 20 is 2 or more and 24 is 0 then 23 should be 1 or more.
<b>Flag 6</b> 20-Subjective distress I=0 & at least one item in sections B,C,D I>0 20-Subjective distress I=4 & no items in sections B,C,D I>2
<b>Flag 7 (Medium)</b> 20-Subjective distress>1 & 21- Impairment in social functioning OR 22- Impairment in occupational functioning=0 and 23- Global validity=0
<b>Flag 8</b> 26-Guilt over acts of commission or omission is F>1 and I>2 & 20- Subjective distress I<2
<b>Flag 9</b> Unrealistically large differences between visits on a given item <b>not during treatment</b> . For example, at visit 2 a score item “x” has a rating of F>2 & I>2 and at visit 3 a rating of F<2 and I<2
<b>Flag 10 (High)</b> 2 (moderate) or worse on 24- global severity & 0-asymptomatic 1-considerable improvement on 25- Global improvement
<b>Flag 11</b>

4- Psychological distress at exposure to cues F>1 and I>1 & 20- Subjective distress<2
Flag 12 (High)
Flag 11 raised and Global validity<2

Table 2 CAPS IV and CAPS V Inconsistency FLAGS

CAPS IV ratings on Frequency and Intensity	CAPS V differences
<p><b>F=Frequency</b> (0=never; 1=once or twice; 2=once or twice a week; 3=several times a week; 4-daily or almost every day); <b>I=Intensity</b> (0=no distress; 1=mild; 2=moderate; 3=severe; 4=extreme)</p> <p><u>Basic item Flags relevant where indicated</u></p> <p>* {F=0 &amp; I&gt;0} – High (Flag 1)</p> <p># {F&gt;0 &amp; I=0} -High (Flag 2)</p>	<p>TR=Trauma relatedness: started or got worse as a result of the event (relevant for 9 (D2), 11 (D4), 12 (D5), 13 (D6), 14 (D7), 15 (E1), 16 (E2), 17 (E3), 18 (E4), 19 (E5), 20 (E6), 29 (1), 30(2), (0=Unlikely, 1=Probable, 2=Definite)</p> <p>TR=<i>Unlikely</i> should not be counted toward a PTSD diagnosis or included in the total CAPS-5 symptom severity score.</p> <p>Based on Frequency and Distress rater combines and makes a single severity score.</p> <p>F=Frequency (e.g., how often in the past week or how much of the time in the past week). Never, 1 week only,</p> <p>0=Absent; 1=1x/week only (1-19%); 2=1x/week or more (20-100%); 3=2x/week or more (50-100%); 4=At least 2x/week but not daily/almost every day (20-79%); 5=Daily/almost every day (80-100%)</p> <p>I=Intensity (aka Distress) (Absent=0, Minimal (Mild)=1, Clearly present (Moderate)=2, Pronounced (Severe)=3, Extreme=4)</p> <p><u>Severity score based on F &amp; I</u></p> <p>0=Absent (F=0 &amp; I=0); 1=Mild subthreshold (F&gt;0 &amp; I=1); 2=Moderate/Threshold (F&gt;0 &amp; I=1 or (F&gt;1 &amp; I=2); 3=Severe/markedly elevated (F&gt;2 &amp; I=3); F=1 &amp; I=4 4=Extreme/incapacitating (F=5 &amp; I=4)</p> <p><u>Basic item Flags relevant where indicated</u></p> <p>* {F=0 &amp; I&gt;0} – High (Flag 1)</p> <p># {F&gt;0 &amp; I=0} --High (Flag 2)</p> <p>! F&gt;0 &amp; I&gt;1 &amp; S&lt;2 (Flag 3)</p> <p>!! TR=0 &amp; S&gt;0 (Flag 4)</p> <p>!!! TR&gt;0 &amp; S=0</p> <p>(Best practice: Capture in data, Int, Freq, and Sev and TR ratings.)</p> <p>* if frequency is somewhat lower than required, but intensity is higher (e.g., if severity= <i>Moderate / threshold</i> if a symptom occurs 1 X month (instead of the required 2 X month) as long as intensity is rated <i>Pronounced or Extreme</i> (instead of the required <i>Clearly Present</i>). Similarly, you may make a severity rating of <i>Severe / markedly elevated</i> if a symptom occurs 1 X week</p>





Global ratings	->
(23) Global validity	(26) ->
(24) Global severity	(27) ->
(25) Global improvement	(28) ->
Associated features	
(26) Guilt over acts of commission or omission	
(27) Survivor guilt	
(28) Reduction in awareness of surroundings	
	Dissociative symptoms (need 1 for subtype)
(29) Derealization	(30) 2 !!
(30) Depersonalization	(29) 1 !!
<b>Composite flags</b>	
<u>Flag 5 (level high)</u> Items 20- subjective distress and 24- Global severity need to be similar. If they are not, we would expect it to have expression in 23- Global validity. For example, If 20 is 2 or more and 24 is 0 then 23 should be 1 or more.	
<u>Flag 6</u> 20-subjective distress I=0 & at least one item I>0 20-Subjective distress I=4 & no items I>2	23-Subjective distress S=0 & at least one item S>0 23-Subjective distress S=4 & no items S>2
<u>Flag 7 level medium</u> If, 20- Subjective distress and 21- impairment in social functioning or 22- impairment in occupational functioning are not similar. (If 20- subjective distress, is 2 or more then we would expect on item 21 or 22 a rating of at least 1 or more or a 1 or more on item 23- Global validity).	If 24 is C1 and C2 item 24
<u>Flag 8</u> 26-Guilt over acts of commission or omission is F>1 and I>2 & 20- subjective distress I<2	
<u>Flag 9</u> Unrealistically large differences between visits on a given item <b>not during treatment</b> . For example, at visit 2 a score item "x" has a rating of F & I>2 and at visit 3 a rating of F<2 and I<2.	
<u>Flag 10</u> 2 (moderate) or worse on 24- global severity and 0-asymptomatic 1-considerable improvement on 25- global improvement (High level flag)	
<u>Flag 11</u> 4- Psychological distress at exposure to cues and 20- Subjective distress should be similar, if rating on 4 is F>1 and I>1 then we would expect that the rating on 20 should be at least 2 or more or the rating on 24- global severity should be 2 or more. If not a flag is raised. (all of these items refer to distress of the subject at a given time point. Thus, the expectation is that there should be a similarity between the three items. If they are not similar then we would expect low validity (item 23) (If not then we would expect low 23-validity) (High level flag)	