HPTN
HIV Prevention
Trials Network

# Depression Measures Working Recommendations

**HPTN Socio-Behavioral and Structural Working Group**

## Introduction

The HIV Prevention Trials Network (HPTN) maintains network-wide Scientific Committees and Working Groups to inform the network's ongoing and proposed work in HIV prevention. In the case of the Socio-Behavioral and Structural Working Group (SBSWG), team members focus on identifying key socio-behavioral and structural priorities, questions and methodologies, including recommendations for network-wide use of key measures to capture important socio-behavioral constructs that are commonly used in HPTN studies. The Mental Health Subgroup (MHS) of the SBSWG was formed to aid the larger working group in developing responses to network-wide issues around mental health. With this document, the SBSWG MHS set out to provide recommendations for network-wide use of depression measures in HPTN studies.

## Mental Health Assessments in HIV Prevention Trials

Although most HPTN studies utilize some kind of mental health (MH) assessment, many studies include measures of mental health without a specific analysis plan and/or use varying measures. There is currently no network-wide guidance on what measures to use for the different MH challenges (potentially) experienced by participants who enroll in HIV prevention trials within the HPTN.

It is important to assess MH in HIV prevention trials because there is a large overlap where MH and HIV play out. MH assessments aid investigators and sites in providing an ethical response (i.e., referral to MH services) for acute individual responses and help raise awareness and appropriate action, if high overall proportions of MH problems are found in the study population. Additionally, MH challenges can disrupt models of health behavior change needed for uptake of biomedical or behavioral interventions and may moderate adherence and/or retention. Further, it is important to have a representative population in HIV prevention trials and MH assessments provide more evidence of how an intervention might play out in the "real world" (esp. with regards to adherence/adherence support). For this and other reasons, assessment of MH may help investigators interpret study results.

## Overall Considerations for Recommended Measures for Depression/Depressive Symptoms

Considerations for recommendations for depression measures were those which have high validity, have been used in diverse and global populations, and are in the public domain. Although specific nuanced depression measures do exist for certain populations (e.g., in pregnant woman, in older adults), the measures suggested below are arguably the most widely used, including in various specialty populations, and have been translated into many languages. The more nuanced measures mentioned above have not been shown to be more reliable or valid in the specific populations compared to the more widely used measures.

We selected two different measures. The PHQ is frequently used to screen for major depressive episode(s) and can be used with a continuous score. The CES-D, which is typically self-report, is frequently used to track symptoms/distress. Researchers should consider differences in use regarding goals for their study and why one of these two assessments might be better suited for the study than the other (e.g., the PHQ is continuous, but typically uses a cutoff for screening and the 9-item version includes an item that assesses suicidality; the CES-D is generally better for looking at a continuous score of symptom severity and does not assess suicidality).

**Patient Health Questionnaire (PHQ) Considerations**

The PHQ has been validated in various languages, races/ethnicities, and populations (adolescents, university students, adults, etc.). See https://www.phqscreeners.com/select-screener for copies of the scale in various languages. Each item is a symptom of a major depressive episode, rated on a 4-point scale from ‘not at all’ (0), to ‘nearly every day’ (3). For a very brief summary of the evidence behind the measure, click here: https://www.mdcalc.com/phq-9-patient-health-questionnaire-9#evidence.

The PHQ-2 (https://www.hiv.uw.edu/page/mental-health-screening/phq-2) is a brief screener and is just two items. Usually, a score of 2 on the PHQ-2 would indicate the need for further screening, though a score of 1 can also be used to increase sensitivity. These assess the first two symptoms of major depressive episode (sadness, loss of interest), and to meet criteria for a major depressive episode, a person needs to endorse at least one of those two symptoms with a total of at least 5 symptoms. *Generally, we would recommend that if a score of 1 or 2 is attained, one would follow up with a PHQ-9, though some studies with very limited capacity to assess depression may have circumstances to use the PHQ-2.*

The PHQ-9 assesses each of the 9 symptoms of major depressive disorder (MDD), mapping onto criteria for the Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnosis of MDD. A summary score of 10 is frequently used as a screener and has approximately 88% sensitivity and 88% specificity for MDD (against structured clinical interview). See APA (American Psychological Association) for more information: https://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/patient-health. (Examples of HPTN studies which have used the PHQ-9 are HPTN 074, 075, 083-01, 084-01, 091.)

Note that the 9th item on the PHQ-9 assesses suicidality. We therefore recommend a full examination of the pros and cons of administering the PHQ-9 as an interviewer-administered measure. A major advantage of doing so is that the study team contemporaneously knows if the participant endorses suicidality and can follow local protocols for safety and ensuring documentation of safety. This may involve immediate action to prevent harm to self, such as observed care or hospitalization, or may involve generating a set of maintenance plans which identify sources of social support, providing resources for mental health treatment, emergency planning, and coping strategies. If used via self-report without an interviewer, procedures would need to be put in place to have real time monitoring of

symptoms endorsed before the participant leaves the visit. This can sometimes be difficult to consistently maintain in an error-proof way. Also, participants would need to be informed that, unlike most A/CASI type assessments, these data would not be anonymous/confidential. Please see Appendix I for a decision tree which investigators can reference when deciding how to assess suicidality.

The PHQ-8 does not ask the suicidality question. However, there is no evidence to suggest that asking about suicide causes or exacerbates suicidal action, so assessing suicide risk (with item 9 on PHQ-9) and then providing help to a participant may be of most interest to a participant who is struggling.

### Center for Epidemiologic Studies Depression Scale (CES-D) Considerations

The CES-D is likely the most widely used measure of depressive symptoms, is in the public domain, and yields a continuous score. The original CES-D was developed in 1977 and revised in 2004. The revision was meant to map onto the symptoms of MDD more closely and also removed items assessing positive affect in order to increase the psychometric reliability. At present, it appears that the original version of the CES-D is still most widely used and has more adaptations/translations available. Additionally, some argue that the absence of positive affect is also important in the assessment of depression and helps differentiate depression from anxiety. Therefore, at present, for HPTN studies that will be working across multiple languages and settings, we currently recommend the *original* CES-D (vs. CES-D Revised), primarily due to the larger number of language translated versions. Information about the revised version can be found here: https://cesd-r.com/. Many HPTN studies have used the CES-D and CESD-10, though some more recent studies have used the PHQ-9. Note that the CES-D does not assess suicidality and hence would screen for elevated depressive symptoms, but not necessarily self-harm.

The original CES-D (20 items) can be found here: http://www.chcr.brown.edu/pcoc/cesdscale.pdf. As mentioned, it yields a continuous score, and a score of 16 can be used as a screening for MDD. The CES-D-R has the same cutoff.

There is a shorter version, the CES-D-10, a 10-item version. This may be used in studies where depression is not central to the study (see: https://www.brandeis.edu/roybal/docs/CESD-10_website_PDF.pdf). For this measure, a score of 10 is used as a cutoff to screen for clinically significant depressive symptoms. Although this measure is validated, sometimes journal reviewers criticize its use over the 20-item version. If more interested in sensitivity to change over time, consider the 20-item version of the CES-D.

### Conclusions

Overall, the most widely globally used measures of depression are the PHQ and the CES-D. Advantages and disadvantages are reviewed above. The biggest differences are that the PHQ tends to be used for a clinical cutoff when wanting a dichotomous screening for clinical depression (but does also provide a continuous score) and the 9$^{th}$ item assesses suicidality. The CES-D is typically used as a continuous measure, though it can also be used to estimate a clinical severity cutoff.

The aforementioned are widely recommended measures across broad populations, age groups, geographies; however, depending on the objectives of the study and specifics of the population, there are other measures that could add more information.

This document was drafted in collaboration with the following HPTN Socio-behavioral and Structural Working Group Mental Health Subgroup members (in alphabetical order):

Lynda Marie Emel, erica l. hamilton, Sybil Hosek, Nyaradzo Mavis Mgodi, Kenneth Ngure, Julie Pulerwitz, Steve Safren, and Teri Senn

If you have any questions regarding this document, feel free to contact Dr. Steve Safren at ssafren@miami.edu.

**APPENDIX I**