|  |  |  |  |
| --- | --- | --- | --- |
| Question | Analytic Approach | Significance Testing | Action Steps |
| Does the measure predict depression onset? | Discrete-Time Survival Analyses | If *p<*.05 for the regression models. | If significant, continue to next step. If non-significant, predictor should not be considered in the screening battery. |
| Can the measure differentiate between those who go on to develop depression and those who do not (*discrimination*)? | Receiver Operating Characteristics | If the 95% confidence interval does not include 0.50 for the AUC statistic | If significant, continue to next step. If non-significant, predictor should not be considered in the screening battery. |
| Is one measure better at differentiating between those at-risk and those not at-risk? | DeLong test for paired ROC curves | If *p*<.05 for the *U* statistic. | If significant, certain index tests should be prioritized and/or differentially weighted within an algorithm. |
| What is the likelihood of developing depression? | Diagnostic Likelihood Ratios | N/A | Can use the likelihood estimates to determine different levels of risk (i.e., cutoffs) based on the estimated odds of developing depression? |
| Is the estimated risk for developing depression valid across risk profiles (*calibration*)? | Expected-Observed (E/O) Index | If the confidence interval for the E/O Index includes 1.00 | If significant, the measure can be included within a screening battery. If non-significant for a certain score (e.g., elevated risk) those scores should not be included in the decision algorithm. |
| Is the estimated risk for developing depression valid across risk profiles (*calibration*)? | Calibration Plots based on Predicted Probabilities and Observed Incidents | N/A | If well-calibrated, the data points should conform to a 45 degree angle. |
| Does the risk algorithm (i.e., the combination of significant predictors) outperform current screening methods (or other individual measures)? | Multivariate Discrete-Time Survival Analyses | If *p*<.05 for the Risk Algorithm score, it provides incremental validity. | Can be used to directly compare which method is more effective. |
| Is the risk algorithm better at differentiating between those at-risk and those not at-risk compared to other screening methods? | DeLong test for paired ROC curves | If *p*<.05 for the *U*-statistic | If significant, it can be determined that one method better differentiates compared to the other. |

*Supplemental Table 1. Summary of Analytic Approach.*

*Note:* All steps above were only considered significant if the findings replicated across both the GEM and MTL-CHI with the exception of tests of calibration, which requires the use of both samples simultaneously (GEM as the development dataset and MTL-CHI as the validation dataset).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 6-MO | 12-MO | 18-MO | 24-MO |
| **GEM Study** (*N*=591) |  |  |  |  |
| Number Entering Follow-Up | 591 | 579 | 555 | 534 |
| Withdrawn from Study | 3 | 21 | 9 | 24 |
| Number with Depression Onset | 9 | 3 | 12 | 7 |
| **MTL-CHI Study** (*N*=348) |  |  |  |  |
| Number Entering Follow-Up | 348 | 335 | 329 | 326 |
| Withdrawn from Study | 0 | 0 | 11 | 1 |
| Number of Episodes | 13 | 6 | 3 | 10 |

*Supplemental Figure 1. Life Table and Kaplan-Meier Survival Curves.* The above table indicates participant flow at each follow up, with the accompanying survival curves displayed below. Youth were censored after a depressive episode or at the time of their lastcompleted follow-up.



*Supplemental Figure 2*: E/O Index=Expected/Observed Index (Hanson 2016), which is created by taking the number of expected depressive episodes for a given risk score in the MTL-CHI study, based on the estimated probability from the GEM study algorithm, and dividing it by the observed number of cases in the MTL-CHI study for that score range. Each data point represents the score on the Cumulative Risk score. Confidence Intervals (CIs) are displayed for those values that were not exactly 1 (indicating perfect calibration).

Observed

Predicted

Categorical Approach

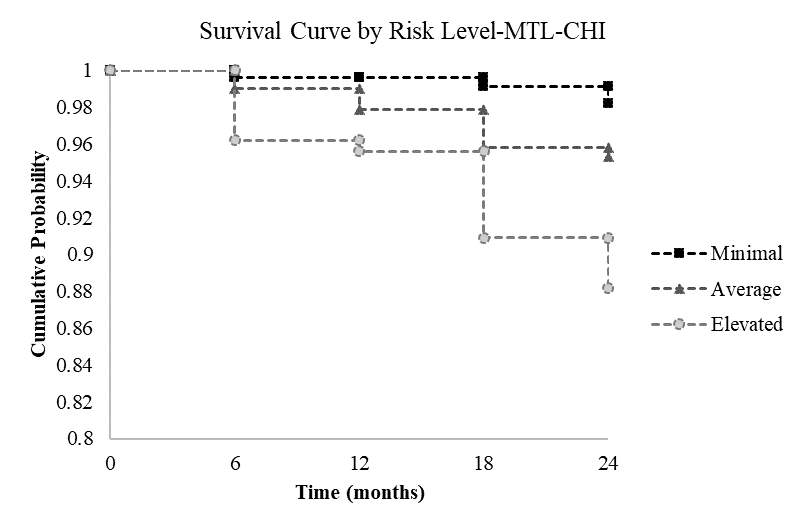
Dimensional Approach

*Note:* Numbers reflect the number of people in each group (e.g., 117 in MTL-CHI means that 117 individuals had a threshold score)

Observed

Predicted

*Supplemental Figure 3. Calibration for Dimensional and Categorical Cumulative Risk Scores. Note:* Calibration plots compared predicted and observed risks of onset of depressive episodes in the test and validation samples. GEM represents the test sample (e.g., predicted values from the GEM model were compared to observed values in the MTL-CHI sample) and MTL-CHI represents the validation sample. Individuals are grouped by predicted probability and points are labelled with the number of individuals in each group. Circles capture grouping by dimensional algorithm scores and triangles capture grouping by Risk Scores (Minimal, Average, Elevated). Numbers that are bolded and italicized represent individuals in each Risk Score group using the categorical approach.



*Supplemental Figure 4. Survival Curves for Risk Score Algorithms. Note:* The above Discrete-Time Survival Curves are for our three categories of Risk Scores: Minimal (0-2), Average (3-4), and Elevated (5-6). Survival curves reflect that those at elevated risk were most likely to experience depression onset by the end of the study, and those at minimal risk were less likely to experience depression onset compared to the sample average.