**Evidence Based Assessment**

Materials to Help Organize Ideas and Illustrate Concepts

Version 1.0

|  |  |
| --- | --- |
| **Eric Youngstrom, Ph.D.**  Professor of Psychology and Neuroscience  University of North Carolina at Chapel Hill  [eay@unc.edu](mailto:eay@unc.edu) | http://musique-libre.org/wp-content/uploads/2014/06/logo_CC_BY_SA_300x100.png |

Introduction

This set of tables, figures, and notes is intended to pull together information to make it easier to teach, to clinically apply, and to do research on evidence based assessment (EBA) (Hunsley, 2007; Hunsley & Mash, 2007; E. A. Youngstrom, 2013).

These materials are licensed under the [CC-BY-SA 4.0 license](https://creativecommons.org/licenses/by-sa/4.0/legalcode). The goal is to have them continue to be available for people to use in their teaching, research, and clinical work, and to make them easy to update. EBA is a work in progress, and it is crucial that people be able to update the information quickly, too. The overarching goal is to help the most people possible, and inasmuch as evidence based assessment improves clinical decisions and outcomes, the more people that it reaches, the better.

I picked the CC-BY-SA license in part because it is fully compatible with Wikipedia, and it should make it easy to share and reuse this information in other formats online.

Let me know if you have suggestions or questions, and looking forward to sharing the ideas together!

Best regards,

Eric Youngstrom, PhD

Table B. Crossing psychometric parameters with the 3 Ps of Evidence Based Assessment (E A Youngstrom & Frazier, 2013)

|  |  |  |  |
| --- | --- | --- | --- |
| Criterion | Prediction | Prescription | Process |
| Norms | Important for accurate prediction (cf. effects of distilled samples or healthy controls when estimating performance in clinical settings) | Not important for diagnostic criteria; nor for treatment moderators | Not relevant for idiographic measurement; but central to outcome benchmarking |
| Internal consistency (Cronbach’s alpha, split half, etc.) | Not directly relevant | Not directly relevant | Could help improve precision for measures of change |
| Inter-rater reliability | For many assessments used at this stage, inter-rater reliability not a major facet | **Core feature.** Inter-rater reliability is pivotal here for assigning diagnosis and treatment selection | Nomothetic approaches using checklists de-emphasize inter-rater reliability; idiographic assessment de-emphasizes |
| Test-retest reliability (stability) | Not directly relevant; prediction equations (especially prediction over time) already model this (albeit not always isolating component due to stability) | Not clinically relevant | Evidence of stability when not treated may provide a benchmark, but evidence of sensitivity to change in context of treatment much more directly relevant (see below) |
| Content validity | Accurate prediction may be possible using correlates and risk factors; thus content validity may be sufficient but not necessary | Treatment moderators may include demographic and clinical features not normally subsumed in psychometric measures | Important. The factor measured should be a core part of the construct for progress and outcome measures. |
| Construct validity (e.g., predictive, concurrent, convergent, and discriminant validity) | Criterion validity more important: Good prediction could be achieved using other constructs and variables as predictors | **Core feature:** the diagnosis should have established construct validity (e.g., Robins & Guze framework) | **Core feature:** Treatment sensitivity, and possibly predictive validity, are key features |
| *Discriminative validity* | **Core feature**: showing ability to improve classification or discriminate between trajectories | Conceptually important, but rarely measured structured diagnostic interviews or moderators. High rates of apparent comorbidity could be inflated by poor discriminative validity | Not directly relevant, unless using process measure to discriminate responder versus non-responder |
| *Prescriptive validity* | Helpful, but diagnosis and formulation will need to be reviewed and finalized by clinician | **Core feature:** This is the reason for establishing a diagnosis or a formulation. The level of granularity should be dictated by what moderates treatment or changes the prescription. | Secondary; if the person is not responding, then prescribes a re-evaluation of formulation |
| *Convergent validity* | Indirectly connected. Could be a high correlation with a criterion. This could be converted to an AUC or other measure of discrimination. | Indirectly connected. Semi-structured interviews that use the diagnostic criteria are formative assessments. | Indirectly connected. |
| Validity generalization | Valuable addition. Important for identifying moderators that would change application to individual client. | Complicated. Diagnostic nosologies assume that same definitions mostly are invariant globally, and using consistent definition reveals differences in prevalence. | Rarely available, but valuable addition. |
| Treatment sensitivity | Not directly relevant. Risk factors, correlates, and predictors may or may not be mutable. However, large Cohen’s d separating clinical and non-clinical distributions would be desirable feature for ROC applications, as well as Jacobson benchmarks for clinically significant change | Not relevant. Treatment moderators may not be mutable. Loss of diagnosis could be a way of defining outcome, but alternate approaches will usually be more practical. | **Core feature**. Needs to be quantified different ways for progress and process measures (slope measures in mixed regressions, or generalizability facets of variance attributable to treatment) than for midterm and endpoint evaluations (where SEdifference and normative benchmarks are key to constructing client-level change measures) |
| Clinical Utility | Important if clinicians are going to use | Important if clinicians are going to use | Important if clinicians are going to use |

Table C

*Lifetime Prevalence Benchmarks from Epidemiological Studies and Medicaid Surveillance*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | NCS-R |  |  |  |  |  |
| Diagnosis or Target Condition | NHANESa 8-15 Years | NCS-Ab 13-18 Years | 18-29 Years\* | 30-44 Years | 45-49 Years | 60+ Years | Rettew et al. (2009) SDI | Rettew et al. (2009) DAU |
| **Any Disorder** | 13% | -- | 52% | 55% | 47% | 26% | -- | -- |
| **Any Anxiety** | **1%** | **32%** | **30%** | **35%** | **31%** | **15%** | -- | -- |
| PTSD | <1% | 5% | 6% | 8% | 9% | 3% | 9% | 3% |
| GAD | <1% | 2% | 4% | 7% | 8% | 4% | 10% | 5% |
| Panic Disorder | <1% | 2% | 4% | 6% | 6% | 2% | 11% | 12% |
| Social Phobia | -- | 9% | 14% | 14% | 12% | 7% | 20% | 6% |
| Separation Anxiety | -- | 8% | 2% | 2% | 1% | 1% | 18% | 8% |
| **Any Impulse Control Disorder** | **--** | **20%** | **27%** | **23%** | **--** | **--** | **--** | **--** |
| ODD | -- | 13% | 10% | 8% | -- | -- | 38% | 37% |
| Conduct Disorder | 2% | 7% | 11% | 8% | -- | -- | 25% | 17% |
| ADHD | 9% | 9% | 8% | 8% | -- | -- | 38% | 23% |
| **Any Mood Disorder** | **4%** | **14%** | **21%** | **25%** | **23%** | **12%** | **--** | **--** |
| MDD | 3% | 12% | 15% | 20% | 19% | 11% | 26% | 17% |
| Bipolar I & II | -- | 3% | 6% | 5% | 4% | 1% | -- | -- |
| Dysthymia | 1% | -- | 2% | 3% | 4% | 1% | 8% | 10% |
| **Any Substance Abuse Disorder** | **--** | **11%** | **17%** | **18%** | **15%** | **6%** | **30%** | **20%** |

*Note.* Statistics adapted from (Kessler, Berglund, Demler, Jin, & Walters, 2005; Merikangas, He, Brody, et al., 2010; Merikangas, He, Burstein, et al., 2010; Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009). NCS-A = National Comorbidity Survey Replication – Adolescent Supplement; NCS-R = National Comorbidity Survey – Replication; SDI = structured diagnostic interview; DAU = diagnosis as usual; PTSD = post-traumatic stress disorder; GAD = generalized anxiety disorder; ODD = oppositional defiant disorder; ADHD = attention-deficit/hyperactivity disorder; MDD = major depressive disorder. NHANES reports 12 month prevalence instead of lifetime.

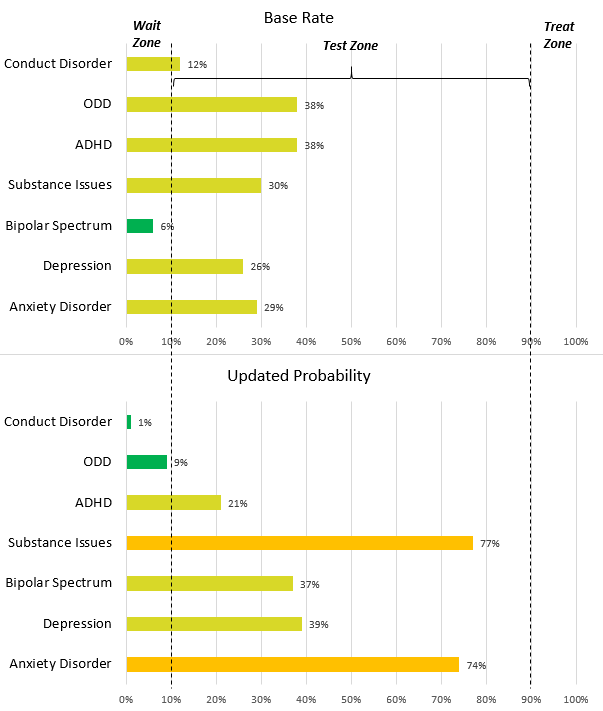
Table A

*Twelve Steps in Implementing Evidence-Based Assessment in Clinical Evaluation of Children and Adolescents*

|  |  |  |
| --- | --- | --- |
| **Assessment Step** | **Rationale** | **Steps to Put in Practice** |
| *Preparatory Work Before Seeing Client* | | |
| 1. Plan for most common issues in our setting | Think about what common clinical hypotheses are; have assessment methods ready that could identify each | Review our practice; select a sample of cases (six months, random draw from past year); make list of most common diagnoses and clinical issues |
| 1. Benchmark base rates for issues | Know the “top ten” issues for our practice; have plan for detecting | Compare local base rate to external benchmarks—other practices and published rates; see Table 4 |
| *Prediction Phase* | | |
| 1. Evaluate risk and protective factors and moderators | Risk factors can trigger more detailed assessment; protective factors (e.g., coping skills, parental warmth) and moderators may inform treatment planning | Make short checklist of key risk factors; make second list of factors that might change treatment selection or moderate outcome; develop plan for how to routinely assess them |
| 1. Revise probabilities based on intake assessments | EBA approach gives appropriate weight to each finding; eliminates geographic and training variations in interpretation (Mulley & Wennberg, 2011); provides clear guidance about next action | Make a table matching assessment tools with common presenting problems. Identify gaps in coverage. Make cheat sheet with key information about assessment for each application. |
| 1. Gather collateral, cross-informant perspectives | Impairment across settings is core to establishing some diagnoses. Discriminative validity of caregiver, teacher, or youth report may differ for specific conditions. Collateral may be crucial if client is young, in denial, or using substances | Gather collateral information to revise case formulation; consider parent, spouse, roommate; also behavioral traces such as Facebook postings. Anticipate typical level of agreement. |
| *Prescription Phase* | | |
| 1. Add focused, incremental assessments | Broad measures will detect problems; focused measures helpful for differentiation among competing hypotheses. | Have follow-up tests available and criteria for when they should be used. Organize so that key information is easy to integrate |
| 1. Add more intensive methods to finalize diagnoses and case formulation | If revised probability falls in the “assessment zone,” what are evidence-based methods to confirm or rule out the diagnosis in question? | Do (semi-)structured interview or review checklist with client to confirm sufficient criteria; supplement with other methods as needed to cross treatment threshold. Cognitive ability, achievement, and diagnostic reading tests might be using here for addressing learning disability hypothesis. |
| 1. Assess for treatment plan and goal setting | Rule out general medical conditions, other medications; family functioning, quality of life, personality, work/school adjustment; comorbidities. Rating scales also can establish baseline severity benchmarks against which to measure treatment response. | Develop systematic ways of screening for medical conditions and medication use. Assess family functioning, personality, comorbidity, SES and other potential treatment moderators. Verbal ability might also be a treatment moderator, changing the appropriateness of cognitive versus behavioral components. |
| X. Learn and use client preferences | Client and caregiver beliefs and attitudes affect treatment seeking and engagement, and are vital for discussing balance of risks and benefits of treatment. | Assess client concordance with treatment plan; ask about cultural factors that might affect treatment plan and engagement; shared decision-making programs |
| *Process Phase* | | |
| 1. Goal setting: Milestones and outcomes (“midterm and final exams”) | Select key targets for treatment; define interim and outcome goals. Periodically repeat main severity measures; if poor response, revisit diagnoses | Make cheat sheet with Jacobson & Truax (1991) benchmarks for measures routinely used; track homework, progress on skills |
| 1. Progress and process measures (“dashboards, quizzes and homework”) | Check learning and use of therapy skills; benchmark early treatment response – progress, or need to tweak treatment? | Track homework, session attendance, life charts, symptom check-ins at each visit, medication monitoring, therapy assignments, daily report cards |
| 1. Wrapping up and maintaining gains | Excellent termination planning celebrates and consolidates gains, and plans for maintenance (e.g., knowing and managing anniversaries and trigger situations) | Develop list of key predictors, recommendations about next action if starting to worsen. |

*Note.* Steps use letters instead of numbers to reinforce the idea that there is not a strict order. Considering patient preferences and cultural factors is ideally infused throughout the assessment and treatment process. It is an x-factor for enhancing rapport and engagement; hence we label it step “X” even though it is mentioned in the middle of the table.

*Figure A.* Using base rates and Bayesian updated probabilities to create a dashboard of probable clinical hypotheses and next actions. The Wait-Test and Test-Treat thresholds are concepts from Evidence Based Medicine (Straus, Glasziou, Richardson, & Haynes, 2011). The exact location of the threshold depends on the risks and benefits involved, as well as patient preferences (Step X in the EBA model). The updated probabilities integrate the information from test results and risk factors, expressed as diagnostic likelihood ratios.

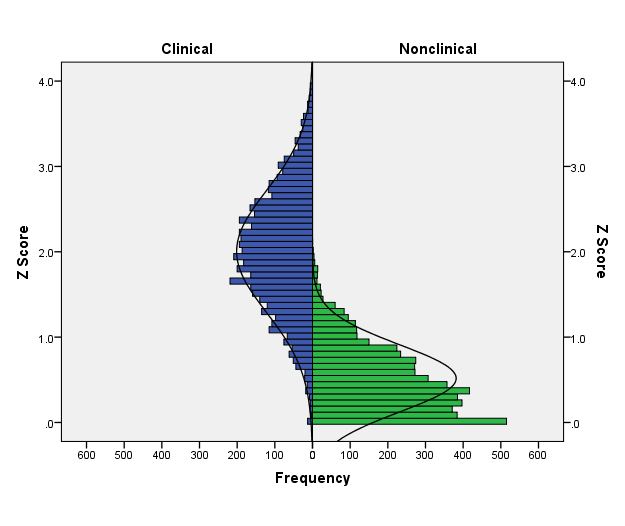


Test-Treat Threshold

Wait-Test Threshold

*Figure B.* Probability nomogram used to combine prior probability with likelihood ratios to estimate revised, posterior probability. Straus et al. (2011) provide the rationale and medical examples; Frazier & Youngstrom (2006) illustrate applying the nomogram to a case with possible ADHD.

*Figure C.* Jacobson’s benchmarks for clinically significant change, based on score distributions in clinical and nonclinical reference groups. Simulated distribution of 10,000 people (5000 in each group). Nonclinical scores bounded at zero; clinical average is +2 SD. Area Under Curve would be .96 (Cohen’s *d* = 2.6). Scores of *z* ≥.65 have 97.5% of the clinical group scoring at or above (Away), *z* ≤1.40 have 97.5% of the nonclinical scoring at or below (Back), and *z* = 1.25 is the threshold for Crossing Closer.

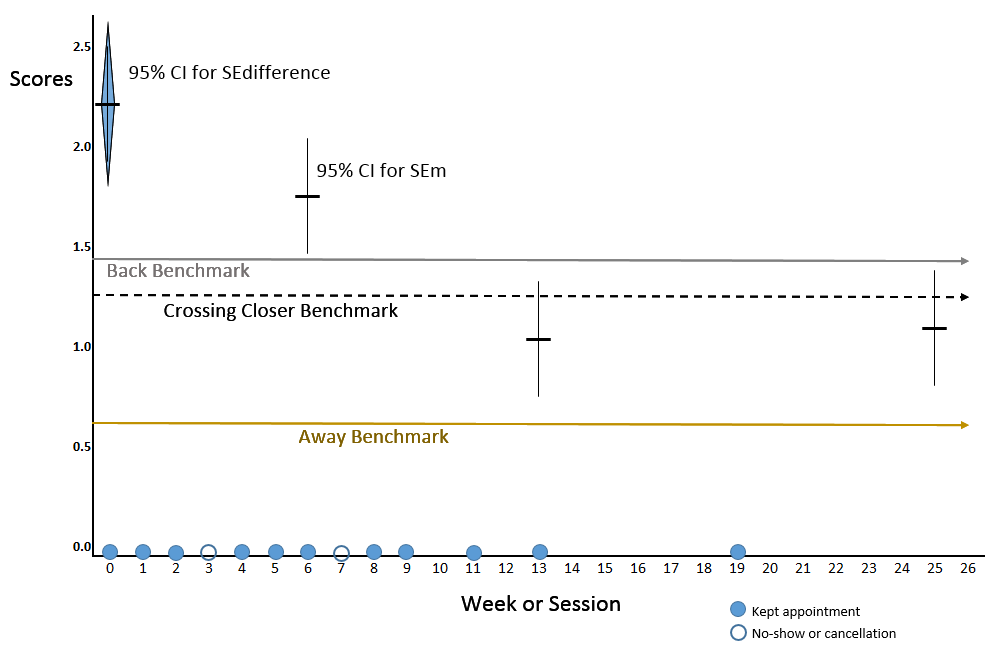


Crossing Closer to Nonclinical

**Away from Clinical**

Back into Nonclinical

*Figure D.* Using the reliable change index and normative benchmarks to look for clinically significant change at the individual patient level, using repeated “midterm” and final assessments. SEdifference = standard error of the difference score; SEm = standard error of the measure. Away, Back, and Closer thresholds based on Jacobson & Truax (1991) definitions. Juxtaposing the session attendance may reveal associations between engagement and progress (Powsner & Tufte, 1994).



Week 25

Maintenance

*z* = 1.2

Week 13

Termination session score

*z* = 1.1

Week 6

Midterm score

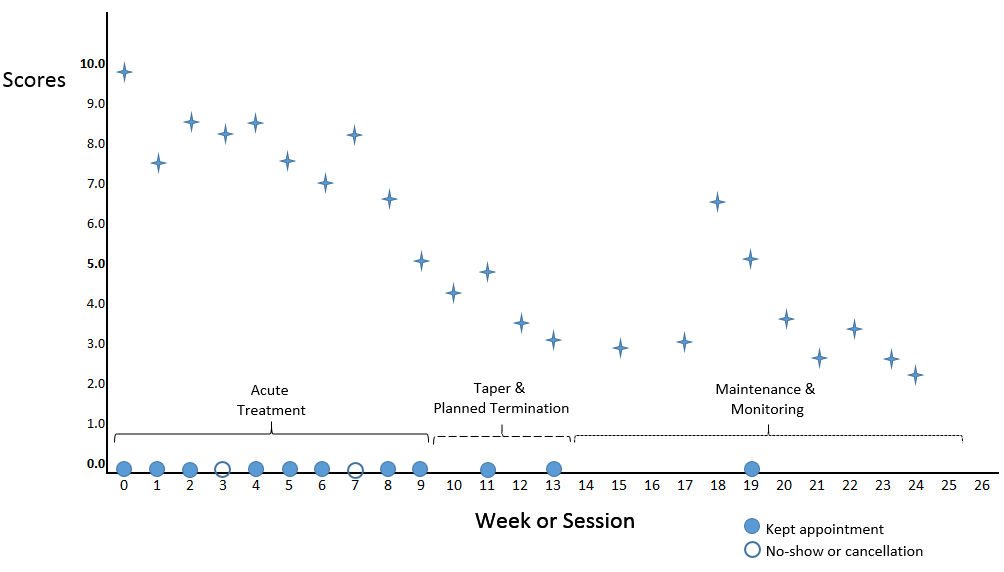
*z* = 1.7

Week 0

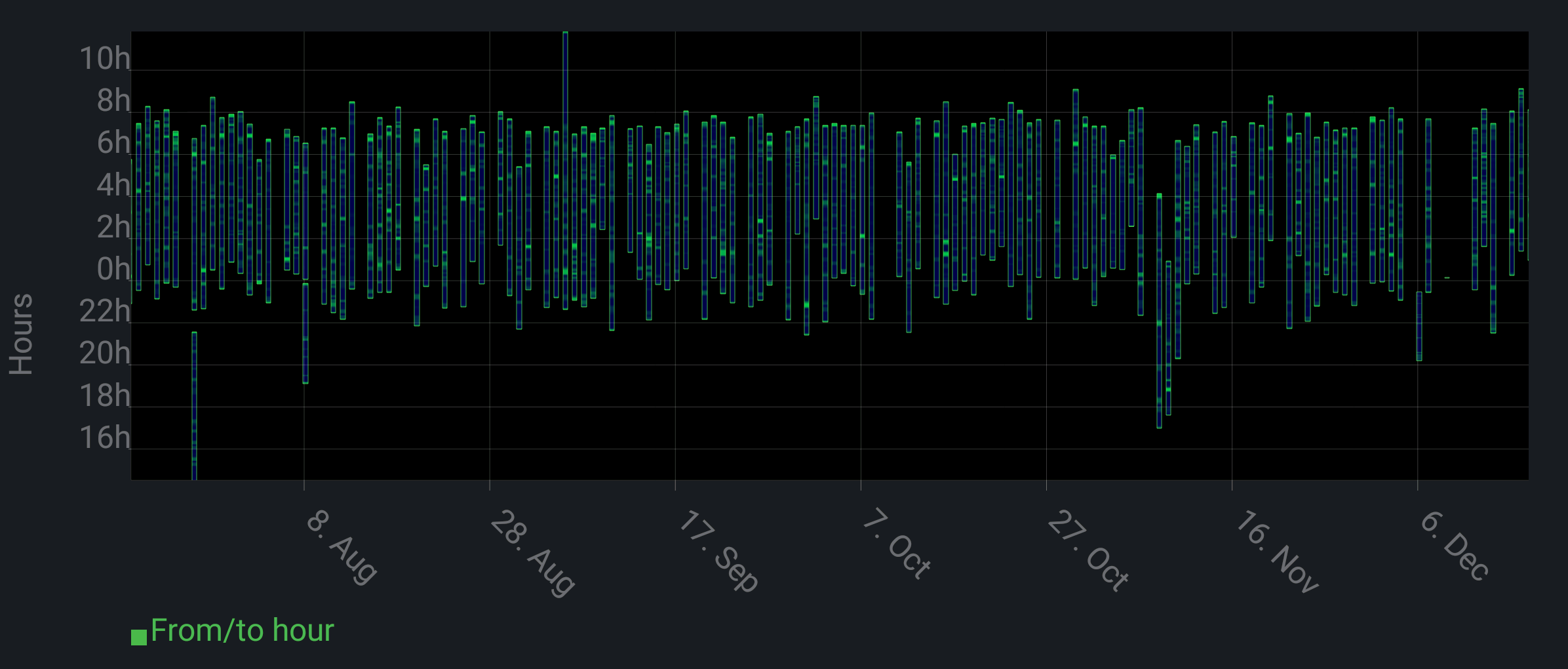
Baseline score

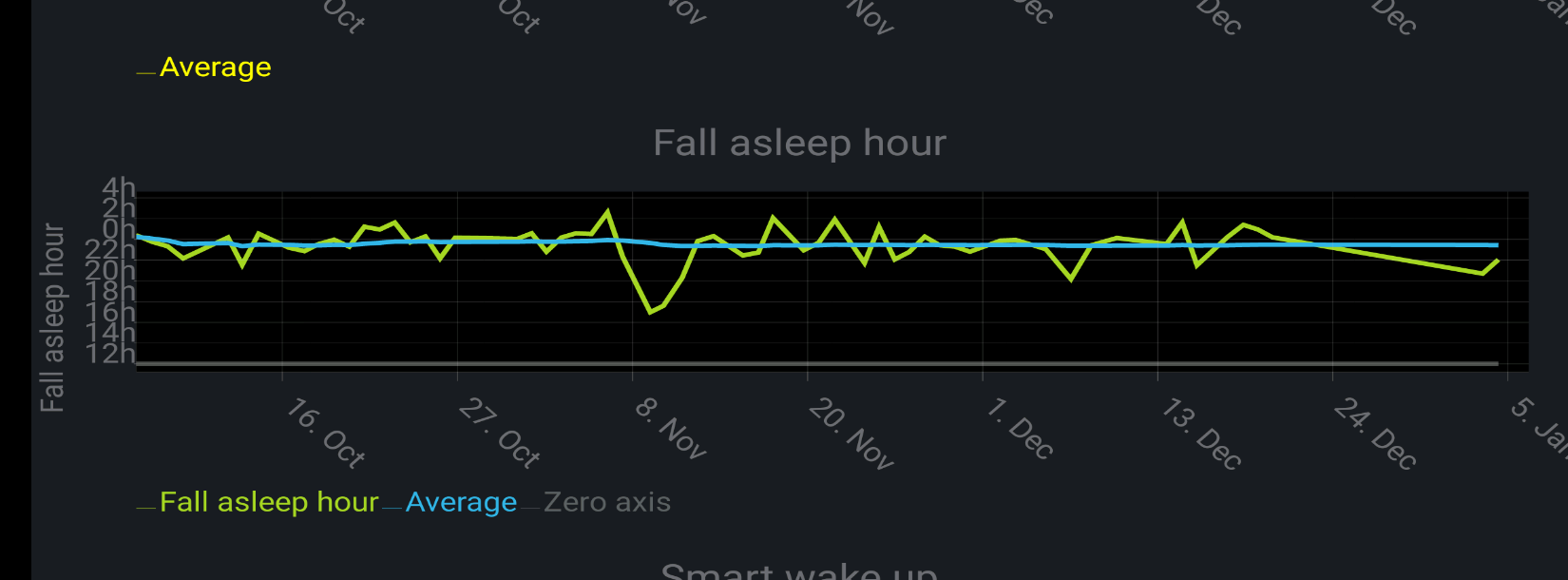
*z* = 2.2

*Figure E.* Tracking session-by-session progress during acute treatment, tapering, and post-termination maintenance. The client can rate progress in between sessions or during session. The frequency and response format are similar to the “Top Problems” methodology (Weisz et al., 2011).



*Figure F.* Example of daily tracking of sleep start, duration, and activity (shown as bar shading color) using a [smartphone application](https://play.google.com/store/apps/details?id=com.urbandroid.sleep&hl=en). The client launches the application and puts the phone on the bed next to them while they sleep. The phone’s gyroscope tracks the client’s motion to estimate time that they fall asleep, wake, and sleep stage (REM, deep sleep, or light sleep) (Lee, Kim, & Welk, 2014). Overlaid red circles indicate a period of major sleep disruption.





**References**

Frazier, T. W., & Youngstrom, E. A. (2006). Evidence-Based Assessment of Attention-Deficit/Hyperactivity Disorder: Using Multiple Sources of Information. *Journal of the American Academy of Child & Adolescent Psychiatry, 45*, 614-620. doi:10.1097/01.chi.0000196597.09103.25

Hunsley, J. (2007). Training psychologists for evidence-based practice. *Canadian Psychology, 38*, 32-42.

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual review of clinical psychology, 3*, 29-51. doi:10.1146/annurev.clinpsy.3.022806.091419

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12-19. doi:10.1037/0022-006X.59.1.12

Kessler, R. C., Berglund, P., Demler, O., Jin, R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*, 593-602.

Lee, J. M., Kim, Y., & Welk, G. J. (2014). Validity of consumer-based physical activity monitors. *Med Sci Sports Exerc, 46*, 1840-1848. doi:10.1249/MSS.0000000000000287

Merikangas, K. R., He, J. P., Brody, D., Fisher, P. W., Bourdon, K., & Koretz, D. S. (2010). Prevalence and treatment of mental disorders among US children in the 2001-2004 NHANES. *Pediatrics, 125*, 75-81. doi:10.1542/peds.2008-2598

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., . . . Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). *Journal of American Academy of Child and Adolescent Psychiatry, 49*, 980-989. doi:10.1016/j.jaac.2010.05.017

Mulley, A. G., & Wennberg, J. E. (2011). Reducing unwarranted variation in clinical practice by supporting clinicians and patients in decision-making. In G. Gigerenzer & J. A. Muir Gray (Eds.), *Better doctors, better patients, better decisions* (pp. 45-52). Cambridge, MA: MIT Press.

Powsner, S. M., & Tufte, E. R. (1994). Graphical summary of patient status. *The Lancet, 344*, 368-389. doi:10.1016/S0140-6736(94)91406-0

Rettew, D. C., Lynch, A. D., Achenbach, T. M., Dumenci, L., & Ivanova, M. Y. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research, 18*, 169-184. doi:10.1002/mpr.289

Straus, S. E., Glasziou, P., Richardson, W. S., & Haynes, R. B. (2011). *Evidence-based medicine: How to practice and teach EBM* (4th ed.). New York, NY: Churchill Livingstone.

Weisz, J. R., Chorpita, B. F., Frye, A., Ng, M. Y., Lau, N., Bearman, S. K., . . . Hoagwood, K. E. (2011). Youth Top Problems: using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. *Journal of Consulting and Clinical Psychology, 79*, 369-380. doi:10.1037/a0023307

Youngstrom, E. A. (2013). Future directions in psychological assessment: Combining Evidence-Based Medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical Child & Adolescent Psychology, 42*, 139-159. doi:10.1080/15374416.2012.736358

Youngstrom, E. A., & Frazier, T. W. (2013). Evidence-based strategies for the assessment of children and adolescents: Measuring prediction, prescription, and process. In D. J. Miklowitz, W. E. Craighead, & L. Craighead (Eds.), *Developmental psychopathology* (2nd ed., pp. 36-79). New York: Wiley.