**The Effectiveness and Acceptability of Empirically-Supported Treatments in**

**Gender Minority Youth Across Four Randomized Controlled Trials**

**Online Supplemental Materials**

|  |  |
| --- | --- |
|  **Table S1** |  **RCT-specific Demographics and Treatment Condition Allocation** |
|  **Part I** |  **Method, Validity, and Reliability of GM Classification** |
|  **Table S2** |  **GM-Classification Variable Correlations Over Time**  |
|  **Table S3** |  **Correlations in Acceptability and Symptom Improvement**  |
|  **Table S4** |  **Treatment Satisfaction by GM Classification**   |
|  **Part II** |  **Treatment Duration as an Indicator of Acceptability** |
|  **Table S5** |  **Treatment Duration by GM Classification** |
|  **Table S6** |  **Symptom Severity and Improvement in CG Youth by Birth-assigned Sex** |
|  **Table S7** |  **Symptom Severity and Improvement in GM Youth by Birth-assigned Sex** |
|  **Part III** |  **Examining Higher-Level Models of Symptom Improvement**    |

**Supplementary Table S1. RCT-specific Demographics and Treatment Condition Allocation**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Pretreatment characteristics | Chorpita et al.,2013(*N* = 174) | Weisz et al., 2019(*N* = 156) | Weisz, Uguetoet al., 2018(*N* = 168) | Weisz, Santucciet al., 2018(*N* = 200) | Pooled sample |
|
|
| Condition |  |  |  |  |  |
|  Usual carea | *n* = 53[0] | *n* = 79[0] | – | – | – |
|  Standard ESTb | *n* = 59[56] | – | – | – | 56 |
|  Modular ESTc  | *n* = 62[55] | *n* = 77[73] | *n* = 168[73] | *n* = 200[175] | 376 |
| Mean age | 10.6 | 10.5 | 9.5 | 10.7 | 10.6 |
| Birth-assigned sex |  |  |  |  |  |
|  Female | 30.5% | 51.9% | 40.5% | 46% | 44.7% |
|  Male | 69.5% | 48.1% | 59.5% | 54% | 55.3% |
| Race/ethnicity |  |  |  |  |  |
|  White | 45% | 79.5% | 85.1% | 32.5% | 55.1% |
|  Black | 9% | 4.5% | 2.4% | 27.5% | 14.6% |
|  Latinx | 6% | 1.9% | 1.8% | 24.0% | 11.6% |
|  Asian | 4% | – | 0.6% | 1.0% | 1.2% |
|  Multiracial | 32% | 12.8% | 10.1% | 13.5% | 15.9% |
|  Other | 4% | 1.3% | – | 1.5% | 1.6% |

*Note*. Brackets include [*n*] of youth in pooled sample, based on whether they had both self- and caregiver-reported data from at least two assessments. aUsual care = everyday practices of therapists in real-world settings. bStandard EST = interventions effective in targeting a specific disorder. cModular EST = treatments employing elements of ESTs to address multiple disorders.

**References**

Chorpita, B. F., Weisz, J. R., Daleiden, E. L., Schoenwald, S. K., Palinkas, L. A., Miranda, J.,

Higa-McMillan, C. K., Nakamura, B. J., Austin, A. A., Borntrager, C. F., Ward, A., Wells, K. C., Gibbons, R. D., & the Research Network on Youth Mental Health. (2013). Long term outcomes for the Child STEPs randomized effectiveness trial: A comparison of modular and standard treatment designs with usual care. *Journal of Consulting and Clinical Psychology,* *81*(6), 999–1009. <https://doi.org/10.1037/a0034200>

Weisz, J. R., Bearman, S. K., Ugueto, A. M., Herren, J. A., Evans, S. C., Cheron, D. M.,

Alleyne, A. R., Weissman, A. S., Tweed, J. L., Pollack, A. A., Langer, D. A., Southam-Gerow, M. A., Wells, K. C., & Jensen-Doss, A. (2019). Testing robustness of Child STEPs effects with children and adolescents: A randomized controlled effectiveness trial. *Journal of Clinical Child and Adolescent Psychology*. Advance online publication. <https://doi.org/10.1080/15374416.2019.1655757>

Weisz, J. R., Santucci, L., & Evans, S. (2018). *Transdiagnostic mental health care for*

*Connecticut’s children and adolescents: A randomized controlled trial of MATCH-ADTC*. Department of Psychology, Harvard University, Cambridge, MA.

Weisz, J. R., Ugueto, A. M., Herren, J., Marchette, L. K., Bearman, S. K., Lee, E. H.,

Thomassin, K., Alleyne, A., Cheron, D. M., Tweed, J. L., Hersh, J., Raftery-Helmer, J. N., Weissman, A. S., & Jensen-Doss, A. (2018). When the torch is passed, does the flame still burn? Testing a “train the supervisor” model for the Child STEPs treatment program. *Journal of Consulting and Clinical Psychology,* *86*(9), 726–737. <http://dx.doi.org/10.1037/ccp0000331>

**Supplemental Part I. Method, Validity, and Reliability of GM Classification**

To establish the most valid method of GM classification, three measurement methods were considered: youth who endorsed a “wish to be the opposite sex” (Achenbach & Rescorla, 2001) (a) at least once across timepoints (*n* = 64), (b) at pretreatment (*n* = 42), and (c) on two or more occasions (*n* = 24). As GM youth who more strongly identify with the opposite sex in childhood have been shown to be more likely to socially transition (i.e., change their gender expression to align with their gender identity; Rae et al., 2019), the third grouping might better represent *persistent* GM identities. However, research indicates that youth who wish to be the opposite sex for a limited period, question their gender identity, and/or express gender nonconformity (e.g., birth-assigned boys who wear dresses but identify as CG) may also be considered GMs and face similar risks (e.g., suicidality) to youth with more persistent GM identities (Jackman et al., 2019; Toomey et al., 2018). To determine the most appropriate criteria for GM classification, outcomes were compared across these three GM groups, and primary analyses were conducted separately with each subsample. Results did not differ by classification method. Specifically, no significant differences in pretreatment and mean (i.e., aggregated across all timepoints) outcome scores were found across GM groups. Similarly, results of primary analyses (i.e., models comparing symptom improvement in CG and GM youth) did not differ by subsample. Thus, all 64 youth who positively endorsed YSR Item #110 at least once were included in the GM subsample to be inclusive of all youth who may be considered GMs and to maximize statistical power.

To further asses the validity of this method, we examined responses on an idiographic measure of youth’s top problems (i.e., Top Problems Assessment; Weisz et al., 2011), which asked youth to identify the three most important behavioral or emotional problems they hoped to address in therapy. Examination of these data failed to produce any overt instances of gender-related concerns for GM youth (e.g., no mention of “gender” or “sex”). The absence of such concerns, however, may be partially attributable to the RCTs’ inclusion criteria (i.e., referral for a primary problem of anxiety, depression, conduct, or trauma) and is consistent with extant literature on GMs’ pursuit of therapy, which is often not explicitly related to gender (Shipherd et al., 2010). Additionally, the measure’s focus on emotional and behavioral problems may have discouraged youth from describing concerns related to gender (e.g., gender dysphoria, GM stressors) and instead captured their symptomatology. Finally, we compared the prevalence of GM youth in current study with past research. Research using this measurement method with clinically-referred samples has collected data at only one timepoint and found similar prevalence rates (4%–6.5%; Achenbach & Rescorla, 2001; Steensma et al., 2013; van der Miesen et al., 2018) compared to the current study (3.2%–9.7% at a given timepoint).

Test-retest reliability of YSR #110 was examined, though interpretation of these results is limited due to inconsistent assessment completion (i.e., not all youth completed the YSR at each timepoint) and the complexity of GM identities and classification. For instance, GM youth referred to treatment in specialty gender clinics have exhibited variability in their response to this item in past research (e.g., birth-assigned girls reported a stronger wish to be the opposite sex compared to birth-assigned boys; de Graaf et al., 2018). Variability in gender expression and identification across GM youth, as well as across time, is increasingly supported in research (Becker et al., 2017; Steensma et al., 2011). Thus, in addition to assessing test-retest reliability from pretreatment to 6 months and again from 6 months to one year, we examined variation in responses to YSR #110 by conducting bivariate correlations across assessment periods. Responses were significantly correlated at all timepoints (*r*s = .11–.62, *p*s < .05; see Table S2 below), and evidence of test-retest reliability was fair to moderate (κ = .3–.5).

**References**

Achenbach, T., & Rescorla, L. (2001). *The manual for the ASEBA school-age forms & profiles*.

Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Becker, I., Ravens-Sieberer, U., Ottová-Jordan, V., & Schulte-Markwort, M. (2017). Prevalence

of adolescent gender experiences and gender expression in Germany. *Journal of Adolescent Health, 61*(1), 83–90. <https://doi.org/10.1016/j.jadohealth.2017.02.001>

de Graaf, N. M., Cohen-Kettenis, P. T., Carmichael, P., de Vries, A. L. C., Dhondt, K., Laridaen,

J., Pauli, D., Ball, J., & Steensma, T. D. (2018). Psychological functioning in adolescents referred to specialist gender identity clinics across Europe: A clinical comparison study between four clinics. *European Child and Adolescent Psychiatry,* *27*(7), 909–919. <https://doi.org/10.1007/s00787-017-1098-4>

Jackman, K. B., Caceres, B. A., Kreuze, E. J., & Bockting, W. O. (2019). Suicidality among

gender minority youth: Analysis of 2017 Youth Risk Behavior Survey Data. *Archives of Suicide Research,* 1–16. Advance online publication. <https://doi.org/10.1080/13811118.2019.1678539>

Rae, J. R., Gülgöz, S., Durwood, L., DeMeules, M., Lowe, R., Lindquist, G., & Olson, K. R.

(2019). Predicting early-childhood gender transitions. *Psychological Science, 30*(5), 669–681. <https://doi.org/10.1177/0956797619830649>

Shipherd, J. C., Green, K. E., & Abramovitz, S. (2010). Transgender clients: Identifying and

minimizing barriers to mental health treatment. *Journal of Gay and Lesbian Mental Health, 14*(2), 94–108. <https://doi.org/10.1080/19359701003622875>

Steensma, T. D., Biemond, R., de Boer, F., & Cohen-Kettenis, P. T. (2011). Desisting and persisting gender dysphoria after childhood: A qualitative follow-up study. *Clinical Child Psychology and Psychiatry, 16*(4), 499–516. <https://doi.org/10.1177/1359104510378303>

Steensma, T. D., van der Ende, J., Verhulst, F. C., & Cohen-Kettenis, P. T. (2013). Gender

variance in childhood and sexual orientation in adulthood: A prospective study. *Journal of Sexual Medicine, 10*(11), 2723–2733.
<https://doi.org/10.1111/j.1743-6109.2012.02701.x>

Toomey, R. B., Syvertsen, A. K., & Shramko, M. (2018). Transgender adolescent suicide behavior. *Pediatrics, 142*(4): e20174218. [https://dx.doi.org/10.1542%2Fpeds.2017-4218](https://dx.doi.org/10.1542/peds.2017-4218)

van der Miesen, A. I. R., Hurley, H., Bal, A. M., & de Vries, A. L. C. (2018). Prevalence of the

wish to be the opposite gender in adolescents and adults with autism spectrum disorder. *Archives of Sexual Behavior, 47*(8), 2307–2317.
<https://doi.org/10.1007/s10508-018-1218-3>

Weisz, J. R., Chorpita, B. F., Frye, A., Ng, M. Y., Lau, N., Bearman, S. K., Ugueto, A. M., Langer, D. A., Hoagwood, K. E., & the Research Network on Youth Mental Health. (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*(3), 369–380. <https://doi.org/10.1037/a0023307>

**Supplementary Table S2. GM-Classification Variable Correlations Over Time**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Timepoint | 0 months | 3 months | 6 months | 9 months | 12 months | 18 months |
| 0 months | – |  |  |  |  |  |
| 3 months | .257\*\* | – |  |  |  |  |
| 6 months | .293\*\* | .603\*\* | – |  |  |  |
| 9 months | .110\* | .345\*\* | .508\*\* | – |  |  |
| 12 months | .159\*\* | .325\*\* | .478\*\* | .478\*\* | – |  |
| 18 months | .146\* | .355\*\* | .492\*\* | .331\*\* | .618\*\* | – |

*Note*. Youth’s responses to YSR Item #110 (“I wish I were of the opposite sex”; Achenbach & Rescorla, 2001), which were used for GM classification, were positively correlated at all timepoints.

\**p* < .05. \*\**p* < .01.

**References**

Achenbach, T., & Rescorla, L. (2001). *The manual for the ASEBA school-age forms & profiles*.

Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

**Supplementary Table S3. Correlations in Acceptability and Symptom Improvement**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Pooled sampleAlliance | CG youthAlliance | GM youthAlliance |
| Variable | Satisfaction | Alliance | Satisfaction | Alliance | Satisfaction | Alliance |
| Alliance | -.133 | – | -.173\* | – | -.085 | – |
| YSRint | -.015 | -.020 | -.047 | -.089 | .156 | .323\* |
| YSRext | .065 | -.069 | .075 | -.097 | .012 | .213 |
| CBCLint | .010 | -.066 | -.019 | -.056 | .345\* | -.075 |
| CBCLext | .021 | -.034 | -.034 | -.020 | .468\*\* | -.058 |

*Note*. Satisfaction = Youth Satisfaction Questionnaire (YSQ-8; Hawley & Weisz, 2005). Alliance = Therapeutic Alliance Scale for Children (TASC; Shirk & Saiz, 1992). YSR/CBCL internalizing/externalizing problems (Achenbach & Rescorla, 2001) represent symptom improvement in youth- and caregiver-reported problems over treatment (i.e., MLM-predicted pre-post treatment change, with higher scores indicating more improvement).

\**p* < .05. \*\**p* < .01.

**References**

Achenbach, T., & Rescorla, L. (2001). *The manual for the ASEBA school-age forms & profiles*.

Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Hawley, K. M., & Weisz, J. R. (2005). Youth versus parent working alliance in usual clinical

care: Distinctive associations with retention, satisfaction, and treatment outcome. *Journal of Clinical Child and Adolescent Psychology, 34*(1), 117–128. <https://doi.org/10.1207/s15374424jccp3401_11>

Shirk, S. R., & Saiz, C. C. (1992). Clinical, empirical, and developmental perspectives on the

therapeutic relationship in child psychotherapy. *Development and Psychopathology, 4*(4), 713–728. <https://doi.org/10.1017/S0954579400004946>

**Supplementary Table S4. Treatment Satisfaction by GM Classification**

|  |
| --- |
| ANCOVAs of Youth Satisfaction Controlling for Posttreatment Symptom Severity  |
| Variable | Type III Sum of Squares | *F*(1,204) | η2 |
| 1. Covariate (YSRint) | .06 | <.01 | <.001 .000 |
|  Group (CG v. GM youth) | 131.2 | **4.84\*** | **.023**  |
| 2. Covariate (YSRext) | 20.1 | .72 | <.01  |
|  Group (CG v. GM youth) | 163.8 | **5.89\*** | **.028** |
| 3. Covariate (CBCLint) | 20.7 | .74 | <.01 |
|  Group (CG v. GM youth) | 140.7 | **5.06\*** | **.024** |
| 4. Covariate (CBCLext) | 34.5 | 1.24 | <.01 |
|  Group (CG v. GM youth) | 154.9 | **5.58\*** | **.027**  |
| Youth Satisfaction Questionnaire (YSQ-8; Hawley & Weisz, 2005) Item-Level Analysis  |
|  | CG youth | GM youth |  |
| Item | *M(SD)* | *M(SD)* | *t*(205) |
| 1. Overall, how happy are you with the help you got?
 | 3.36(.87) | 3.12(.95) | 1.47 |
| 1. If you had a friend that needed help, would you tell your friend about the help you got?
 | 3.24(.84) | 2.97(.94) | 1.66 |
| 1. **Did you get the kind of help you wanted?**
 | 3.32(.81) | 2.97(.94) | **2.27\*** |
| 1. **If you had your choice, would you choose to do the same kinds of things you did here?**
 | 3.19(.87) | 2.74(.93) | **2.72\*\*** |
| 1. **Do you think the help you got here will make things better for you later on?**
 | 3.41(.76) | 3.03(.97) | **2.57\*** |
| 1. How would you rate the help you got?
 | 3.43(.76) | 3.27(.84)  | 1.06 |
| 1. How happy are you with how much help you got?
 | 3.39(.81) | 3.21(.95) | 1.16 |
| 1. Would you come back to counseling if you need help again?
 | 3.39(.81) | 3.15(.82) | 1.59 |
| **Total score** | 26.7(5.16) | 24.5(5.80) | **2.29\*** |

*Note*. ANCOVAs showed that GM youth reported lower satisfaction when controlling for MLM-predicted youth- and caregiver-reported posttreatment symptom severity. These lower satisfaction scores were most prominent on the three bolded items. Independent samples *t*-tests were used to examine item-level differences, as results did not differ from ANCOVAs.

\**p* < .05. \*\**p* < .01.

**References**

Achenbach, T., & Rescorla, L. (2001). *The manual for the ASEBA school-age forms & profiles*.

Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Hawley, K. M., & Weisz, J. R. (2005). Youth versus parent working alliance in usual clinical

care: Distinctive associations with retention, satisfaction, and treatment outcome. *Journal of Clinical Child and Adolescent Psychology, 34*(1), 117–128. <https://doi.org/10.1207/s15374424jccp3401_11>

**Supplemental Part II. Treatment Duration as an Indicator of Acceptability**

Treatment duration, in days and sessions attended, was examined as a potential indicator of acceptability. Though youth psychotherapy duration is perhaps most robustly related to pretreatment characteristics (e.g., initial symptom severity; Brookman-Frazee et al., 2008), it has also been associated with youth- and caregiver-reported satisfaction (Garland et al., 2000; Garland et al., 2007). Modular EST has been shown to be more efficient than standard ESTs in one study (i.e., requiring fewer days/sessions; Chorpita et al., 2017). However, duration did not vary across EST conditions in the present study, allowing for comparisons between GM and CG youth regardless of EST condition.

Group differences in treatment duration were evaluated via ANCOVA, controlling for pretreatment symptom severity, as greater initial symptom severity might necessitate more sessions and span more days. Across the full sample, duration ranged from 8 to 589 days (*M* = 201.7, *SD* = 111.5) and did not significantly differ between CG (*M* = 202.7, *SD* = 112.0) and GM (*M* = 196.2, *SD* = 109.8) youth (see Table S5). Similarly, number of sessions attended, ranging from 1 to 87 sessions (*M* = 14.5, *SD* = 10.9), did not significantly vary across CG (*M* = 14.6, *SD* = 10.3) and GM (*M* = 14.2, *SD* = 14.1) groups (see Table S5).

Though treatment duration has been utilized as an indicator of youth satisfaction with mental health treatment previously (e.g., Garland et al., 2007), the course and length of treatment is highly variable within modular ESTs (the condition to which most youth were allocated in the present study), as therapists have flexibility in selecting and implementing the most appropriate modules (Weisz et al., 2019). Thus, similar duration across CG and GM youth is difficult to interpret. While this finding may suggest comparable levels of satisfaction, it may be confounded by other factors associated with treatment duration, such as case complexity (e.g., comorbidity). Instead, treatment retention (i.e., the proportion of youth retained though completion), which was not collected consistently across RCTs in the present study, might better reflect satisfaction and should be examined in future studies.

**References**

Brookman-Frazee, L., Haine, R. A., Gabayan, E. N., & Garland, A. F. (2008). Predicting

frequency of treatment visits in community-based youth psychotherapy. *Psychological Services, 5*(2), 126–138. [https://dx.doi.org/10.1037%2F1541-1559.5.2.126](https://dx.doi.org/10.1037/1541-1559.5.2.126)

Chorpita, B. F., Daleiden, E. L., Park, A. L., Ward, A. M., Levy, M. C., Cromley, T., Chiu, A.

W., Letamendi, A. M., Tsai, K. H., & Krull, J. L. (2017). Child STEPs in California: A cluster randomized effectiveness trial comparing modular treatment with community implemented treatment for youth with anxiety, depression, conduct problems, or traumatic stress. *Journal of Consulting and Clinical Psychology,* *85*(1), 13–25. <https://doi.org/10.1037/ccp0000133>

Garland, A. F., Aarons, G. A., Saltzman, M. D., & Kruse, M. I. (2000). Correlates of

adolescents’ satisfaction with mental health services. *Mental Health Services Research, 2*(3), 127–139. [https://doi.org/10.1023/a:1010137725958](https://doi.org/10.1023/a%3A1010137725958)

Garland, A. F., Haine, R. A., Lewczyk Boxmeyer, C., & Child and Adolescent Services

Research Center. (2007). Determinants of youth and parent satisfaction in usual care psychotherapy. *Evaluation and Program Planning, 30*(1), 45–54. [https://dx.doi.org/10.1016%2Fj.evalprogplan.2006.10.003](https://dx.doi.org/10.1016/j.evalprogplan.2006.10.003)

Weisz, J. R., Bearman, S. K., Ugueto, A. M., Herren, J. A., Evans, S. C., Cheron, D. M.,

Alleyne, A. R., Weissman, A. S., Tweed, J. L., Pollack, A. A., Langer, D. A., Southam-Gerow, M. A., Wells, K. C., & Jensen-Doss, A. (2019). Testing robustness of Child STEPs effects with children and adolescents: A randomized controlled effectiveness trial. *Journal of Clinical Child and Adolescent Psychology*. Advance online publication. <https://doi.org/10.1080/15374416.2019.1655757>

**Supplementary Table S5. Treatment Duration by GM Classification**

|  |
| --- |
| ANCOVAs of Treatment Duration Controlling for Pretreatment Symptom Severity |
| Treatment Duration in Days |
| Variable | Type III Sum of Squares | *F*(1,402) | η2 |
| 1. Covariate (YSRint) | 95.5 | <.01 | <.001 .000 |
|  Group (CG v. GM youth) | 2281.9 | .18 | <.001  |
| 2. Covariate (YSRext) | 41547.4 | 3.35 | <.01 |
|  Group (CG v. GM youth) | 4.14 | <.001 | <.001 |
| 3. Covariate (CBCLint) | 92279.1 | **7.52\*\*** | **.02** |
|  Group (CG v. GM youth) | 3.38 | <.001 | <.001 |
| 4. Covariate (CBCLext) | 123.3 | .01 | <.001 |
|  Group (CG v. GM youth) | 2260.3 | .18 | <.001 |
| Treatment Duration in Number of Sessions |
| Variable | Type III Sum of Squares | *F*(1,418) | η2 |
| 1. Covariate (YSRint) | 178.7 | 1.50 | <.01 |
|  Group (CG v. GM youth) | 39.5 | .33 | <.01 |
| 2. Covariate (YSRext) | 333.6 | 2.81 | <.01 |
|  Group (CG v. GM youth) | 1.88 | .02 | <.001 |
| 3. Covariate (CBCLint) | 2619.0 | **23.1\*\*** | **.05** |
|  Group (CG v. GM youth) | 23.0 | .20 | <.001 |
| 4. Covariate (CBCLext) | 23.4 | .20 | <.001 |
|  Group (CG v. GM youth) | 4.20 | .04 | <.001 |

*Note*. Treatment duration, in days and number of sessions, did not significantly differ across CG and GM youth when controlling for pretreatment symptom severity.

\*\**p* < .01.

**References**

Achenbach, T., & Rescorla, L. (2001). *The manual for the ASEBA school-age forms & profiles*.

Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

**Supplementary Table S6. Symptom Severity and Improvement in CG Youth by Birth-assigned Sex**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Birth-assigned boys(*n* = 206) | Birth-assigned girls(*n* = 162) | *t*(df) | ES | 95% CI  |
| Pretreatment symptom severity  |
|  YSRint | *M(SD)* = 56.9(10.8) | *M(SD)* = 55.1(10.8) | *t*(366) = 1.57 | .17 | [-.45, 4.01] |
|  YSRext | *M(SD)* = 52.9(11.0) | *M(SD)* = 51.1(10.4) | *t*(366) = 1.64 | .17 | [-.36, 4.09] |
|  CBCLint | *M(SD)* = 66.0(8.5) | *M(SD)* = 67.0(7.8) | *t*(366) = 1.15 | .12 | [-2.69, .71] |
|  CBCLext | *M(SD)* = 66.1(8.5) | *M(SD)* = 63.1(10.2) | *t*(311.4) = 2.98\*\* | .31 | [1.05, 4.87] |
| MLM-predicted pre-post treatment symptom improvement |
|  YSRint | *M(SD)* = 13.8(6.2) | *M(SD)* = 13.8(6.0) | *t*(366) = .01 | .01 | [-1.25, 1.26] |
|  YSRext | *M(SD)* = 7.52(3.1) | *M(SD)* = 7.92(2.8) | *t*(366) = 1.30 | .13 | [-.21, 1.01] |
|  CBCLint | *M(SD)* = 12.1(6.6) | *M(SD)* = 12.4(6.9) | *t*(366) = .44 | .04 | [-1.08, 1.70] |
|  CBCLext | *M(SD)* = 9.05(6.1) | *M(SD)* = 8.75(6.3) | *t*(366) = .46 | .05 | [-1.57, .98] |

*Note*. For CG youth, birth-assigned boys had significantly higher caregiver-reported pretreatment externalizing problems than birth-assigned girls, consistent with previous research (Nock et al., 2006). YSR/CBCL internalizing/externalizing problems (Achenbach & Rescorla, 2001) represent (a) pretreatment symptom severity and (b) improvement in youth- and caregiver-reported problems over treatment (i.e., MLM-predicted pre-post pretreatment change, with higher scores indicating more improvement), respectively. Effect sizes are reported as Cohen’s *d*.

\*\**p* < .01.

**References**

Achenbach, T., & Rescorla, L. (2001). *The manual for the ASEBA school-age forms & profiles*.

Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2006). Prevalence, subtypes, and

correlates of DSM-IV conduct disorder in the National Comorbidity Survey Replication. *Psychological Medicine, 36*(5), 669–710. [https://dx.doi.org/10.1017%2FS0033291706007082](https://dx.doi.org/10.1017/S0033291706007082)

**Supplementary Table S7. Symptom Severity and Improvement in GM Youth by Birth-assigned Sex**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Birth-assigned boys(*n* = 33) | Birth-assigned girls(*n* = 31) | *t*(df) | ES | 95% CI  |
| Pretreatment symptom severity |
|  YSRint | *M(SD)* = 65.3(10.4) | *M(SD)* = 59.4(6.9) | *t*(55.8) = 2.71\*\* | .67 | [1.55, 10.3] |
|  YSRext | *M(SD)* = 58.0(10.6) | *M(SD)* = 55.8(9.6) | *t*(62) = .89 | .22 | [-2.80, 7.31] |
|  CBCLint | *M(SD)* = 62.3(10.8) | *M(SD)* = 65.0(10.1) | *t*(62) = 1.03 | .26 | [-7.92, 2.53] |
|  CBCLext | *M(SD)* = 63.0(10.5) | *M(SD)* = 60.9(9.8) | *t*(62) = .84 | .21 | [-2.96, 7.21] |
| MLM-predicted pre-post treatment symptom improvement |
|  YSRint | *M(SD)* = 14.7(7.2) | *M(SD)* = 13.0(7.0) | *t*(62) = .92 | .24 | [-5.20, 1.93] |
|  YSRext | *M(SD)* = 8.32(3.8) | *M(SD)* = 7.83(3.1) | *t*(62) = .56 | .14 | [-2.23, 1.25] |
|  CBCLint | *M(SD)* = 10.5(5.9) | *M(SD)* = 8.60(6.6) | *t*(62) = 1.19 | .30 | [-5.01, 1.27] |
|  CBCLext | *M(SD)* = 6.71(5.9) | *M(SD)* = 5.13(6.2) | *t*(62) = 1.04 | .26 | [-4.60, 1.44] |

*Note*. In the GM subsample, birth-assigned boys had significantly higher youth-reported internalizing problems at pretreatment compared to birth-assigned girls. YSR/CBCL internalizing/externalizing problems (Achenbach & Rescorla, 2001) represent (a) pretreatment symptom severity and (b) improvement in youth- and caregiver-reported problems over treatment, respectively. Effect sizes are reported as Cohen’s *d*.

\*\**p* < .01.

**References**

Achenbach, T., & Rescorla, L. (2001). *The manual for the ASEBA school-age forms & profiles*.

Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

**Supplemental Part III. Examining Higher-Level Models of Symptom Improvement**

We examined whether higher-level models (e.g., 3-level models) of symptom improvement were necessary and found that the variance accounted for in outcomes by RCT, EST-condition, and study site was near zero (average intraclass correlation [ICC]RCT = .039, ICCEST-condition = .043, ICCsite = .035). Using the recommended cut-off (ICC > .1; Vajargah & Nikbakht, 2015), we determined that the more parsimonious 2-level model was well-supported.

**References**

Vajargah, K. F., & Nikbakht, M. (2015). Application REML model and determining cut off of

ICC by multi-level model based on Markov chains simulation in health. *Indian Journal of Fundamental and Applied Life Sciences, 5*(17), 1432–1448.