Supplemental Material A: Clinician Training and Adherence

Clinicians were clinical psychology doctoral students (n = 3), a clinical psychology postdoctoral fellow (n = 1), and a licensed clinical social worker (n = 1). All clinicians had extensive training in cognitive-behavioral therapy through our specialty clinic and, in addition, completed two 6-hour didactic trainings focused on using MATCH specifically. The majority of study cases (29 out of 40) were treated by clinicians who had already earned their terminal degree (i.e., PhD or LCSW). The principal investigator (PI) also supervised each clinician treating youth with MATCH skills prior to the clinician's first study case. To train in the SDM protocol, clinicians each studied the protocol, met individually with the PI for at least 6 hours of SDM-focused training, and used the SDM protocol with at least one family (or study staff role-playing family members).

The PI reviewed session recordings and provided weekly supervision to all study clinicians. Supervision only differed in relation to treatment planning. For the CG condition, the PI and clinician met ahead of the initial treatment session to plan the treatment based on baseline assessment data and the PWEBS database. For the SDM condition, the PI and clinician met prior to the initial treatment (SDM) session to discuss the baseline assessment data and review the PWEBS database, but the actual treatment plan was built in the SDM session with the family. The PI and SDM clinician then discussed the treatment plan after the session. To prevent contamination of the CG condition, CG clinicians were instructed to respond to family requests/suggestions for treatment plan changes with clear language that identified the clinician as the decision maker (e.g., "I will consider your request and make the decision I think is best."). In addition, CG clinicians did not use any of the SDM tools (e.g., the SDM activity), handouts, or other materials. The PI monitored and ensured the use of these responses and restrictions

through review of session recordings.

Clinicians were largely adherent to the SDM protocol. In addition to weekly supervision, observational coders rated all SDM sessions (n = 20) to test for the degree to which clinicians covered the prescribed content and did so in a manner consistent with SDM principles. SDM sessions included content from the majority of the protocol's 10 sections (M = 9.35, SD = 0.81; see Table 3 for a summary of the SDM protocol sections). Coders also used the OPTION-5 (Barr et al., 2015) to assess SDM practices, with clinicians scoring very high (M = 19.20, SD = 1.01; maximum score = 20).

To assess adherence to the use of the treatment procedures specified in the MATCH modules, we coded 75 sessions (11.1% of all sessions held), including a minimum of 10% from each participant in the study. Following the approach modelled by Weisz et al. (2012), we calculated the percentage of treatment elements in each session that align with MATCH. All (i.e., 100%) of intervention techniques delivered in the coded sessions aligned with MATCH. However, 22.7% of coded sessions (n = 17) did not include a MATCH treatment element; instead, these sessions focused on MATCH-consistent elements including review of treatment, maintenance, and/or termination (n = 11), homework review and planning (n = 2), or managing resistance to treatment engagement (n = 1). This aligned with the findings of Weisz et al. (2012) and Chorpita et al. (2017), showing the similarity of our study's delivery of MATCH to that of the trial implementations.

Supplemental Material B: Additional Information about Study Measures

Diagnosis, Symptom, and Functioning Measures. Youth diagnoses were determined using the Anxiety Disorders Interview Schedule for DSM-IV, Child/Parent (ADIS-IV; Silverman & Albano, 1997), an extensively tested semi-structured interview assessing the major anxiety, mood, and externalizing disorders. Youth symptoms were additionally assessed using two youth- and caregiver-report measures. First, the Multidimensional Anxiety Scale for Children (MASC; March et al., 1997) is a 39-item measure of youth anxiety symptoms. Items were rated on a 4-point scale from 0 (*never true*) to 3 (*often true*). Second, the Children's Depression Inventory (CDI; Kovacs, 1992) is a measure of depression symptoms in youth (child = 27 items; parent = 17 items) using a three-choice format (e.g., 0 = "*I am sad once in a while*," 1 = "*I am sad many times,*" 2 = "*I am sad all the time*"). Both measures demonstrated good to excellent reliability for youth and caregivers in the current study (MASC α = .91 and .89, respectively; CDI α = .90 and .79, respectively). Assessors completed the Clinical Global Impression Scale (CGIS; Guy, 1976), a 7-point clinician-rated scale measuring overall clinical severity (CGIS-S) and improvement (CGIS-I); lower scores indicate lower severity and greater improvement.

Decision-Making Measures. Youth and caregivers completed a number of measures relevant to treatment-related decision-making. The 16-item Decisional Conflict Scale (DCS; O'Connor, 1995) was used to measure personal perceptions of a) uncertainty in choosing options, b) factors contributing to uncertainty (e.g., feeling uninformed), and c) factors contributing to effective decision making. Items were rated on a 5-point scale from 0 (*strongly agree*) to 4 (*strongly disagree*). The DCS demonstrated excellent reliability for youth and caregivers in the current study (α = .96 and .97, respectively). The 6-item Satisfaction with Decision Scale (SWD; Holmes-Rovner et al., 1996) was used to assess satisfaction with

treatment-related decisions. Items were rated on a 5-point scale from 1 (*strongly disagree*) to 5 (*strongly agree*). The SWD demonstrated excellent reliability for youth and caregiver in the current study (α = .96 and .96, respectively). The 11-item Decision Self-Efficacy Scale (DSES; O'Connor, 1995) was used to measure belief in one's ability to make treatment decisions (e.g., understand options, ask questions, express opinions). Items were rated on a 5-point scale from 0 (*not at all confident*) to 4 (*very confident*). The DSES demonstrated good to excellent reliability for youth and caregiver in the current study (α = .86 and .92, respectively). The 5-item Decisional Regret Scale (DRS; Brehaut et al., 2003) was used to measure caregiver and youth remorse after specific treatment-related decisions (i.e., treatment targets, participants, and components). Items were rated on a 5-point scale from 1 (*strongly agree*) to 5 (*strongly disagree*). The DRS demonstrated excellent reliability for youth and caregivers in the current study (α = .99 and .99, respectively). The 9-item Shared Decision-Making Questionnaire-9 (SDMQ-9; Kriston et al., 2010) was used to assess the extent of patient involvement in decision making (e.g., the degree to which the clinician discussed treatment information, elicited and valued patient opinions). Items were rated on a 6-point scale from 0 (*no effort was made*) to 5 (*every effort was made*). The SDMQ-9 demonstrated excellent reliability for youth and caregiver in the current study (α = .90 and .93, respectively). Each of the above-described decision-making measures was originally developed for adult patients to report on their own experiences; wording was minimally modified for youth to report on their own psychotherapeutic treatment and caregivers to report on their experiences related to their child's psychotherapeutic treatment.

Treatment Process and Related Measures. Clinicians completed a clinical note for each treatment session that included content covered (e.g., treatment skill, crisis of the week) as well as treatment planning discussions (i.e., was treatment plan discussed, questioned, and/or

modified). To measure treatment engagement, clinicians also tracked total number of sessions attended, number of sessions missed, and who participated in each session (i.e., youth alone, caregiver alone, caregiver and youth together). In addition, youth and caregivers completed a number of self-report measures related to treatment process. The 8-item Treatment Outcomes Expectation Scale (TOES; Bickman et al., 2010) was completed by youth and caregiver to assess expectations about the anticipated outcomes of treatment. Items were rated on a 3-point scale from 1 (*I do not expect this*) to 3 (*I do expect this*). The TOES demonstrated acceptable to good reliability for youth and caregiver in the current study (α = .79 and .67, respectively). Youth completed the 8-item Motivation for Youth Treatment Scale (MYTS; Bickman et al., 2010) to assess readiness to participate in treatment. Items were rated on a 5-point scale from 1 (*strongly disagree*) to 5 (*strongly agree*). The MYTS demonstrated good reliability in the current study (α = .85).

The 7-item Therapeutic Alliance Scale for Children (TASC; Shirk & Saiz, 1992) measured perceived relationship strength with the clinician. Youth rated items on a 4-point scale from 1 (*not true*) to 4 (*very much true*). The TASC demonstrated good reliability in the current study (α = .86).

Family and Youth Background. Caregivers reported youth and family demographic information including youth age, gender, ethnicity, and race, as well as household income and caregiver education level.

Table B1: Assessment Timing

Domain	Measure	Inf	B	S1	S8	PT
Diagnosis, Symptom, & Functioning Measures	Anxiety Disorders Interview Scheduled for DSM-IV; Clinical Global Impression Scale	Cl	●			●
	Multidimensional Anxiety Scale for Children; Children's Depression Inventory	C,Y	●		●	●
Decision-Making Measures	Decision Self-Efficacy Scale	C,Y	●	●		
	Decisional Conflict Scale; Satisfaction with Decision Scale	C,Y		●		
	SDM Questionnaire-9	C,Y		●		
	Decisional Regret Scale	C,Y			●	
Treatment Process and Related Measures	Session Notes, Sessions Attended, Sessions Missed	Cl				
	Treatment Outcomes Expectation Scale	C,Y		●	●	
	Motivation for Youth Treatment Scale	Y		●	●	
	Therapeutic Alliance Scale for Children	C,Y		●	●	●
Family and Youth Background	Demographics	C	●			

Note: Inf=Informant; C=Caregiver; Y=Youth; Cl=Clinician; B=Baseline assessment; S1=After initial SDM or CG session; S8=after the 8th treatment session; PT=Posttreatment assessment