Multisystemic Therapy versus management as usual in the treatment of adolescent antisocial behaviour (START): a randomised controlled pragmatic effectiveness superiority trial

Peter Fonagy, Stephen Butler, David Cottrell, Stephen Scott, Stephen Pilling, Ivan Eisler, Peter Fuggle, Abdullah Kraam, Sarah Byford, James Wason, Rachel Ellison, Elizabeth Simes, Poushali Ganguli, Elizabeth Allison, Ian M Goodyer

Research Department of Clinical, Educational and Health Psychology, University College London, London, UK (Prof P Fonagy PhD, S Butler PhD, Prof S Pilling PhD, R Ellison BSc, E Simes MA, E Allison DPhil); Leeds Institute of Health Sciences, University of Leeds, Leeds, UK (Prof D Cottrell FRCPsych); Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK (Prof S Scott FRCPsych, Prof I Eisler PhD, Prof S Byford PhD, P Ganguli MSc); Anna Freud National Centre for Children and Families, London, UK (P Fuggle PhD); University of Leeds and South West Yorkshire Partnership NHS Foundation Trust, Leeds, UK (A Kraam MD); MRC Biostatistics Unit, University of Cambridge, Cambridge, UK (J Wason PhD); Department of Psychiatry, University of Cambridge, Cambridge, UK (Prof I M Goodyer MD).

Correspondence to: Professor Peter Fonagy, Research Department of Clinical, Educational and Health Psychology, University College London, London WC1E 7HB, UK. E-mail: p.fonagy@ucl.ac.uk Telephone: +44 7679 1943

Summary

Background: Adolescent antisocial behaviour is a major health and social problem. Multisystemic Therapy (MST) has reduced symptoms and offending rate in US trials, but non-US findings are equivocal.

Methods: We conducted an 18-month multisite pragmatic randomised controlled superiority trial in England. Adolescents (aged 11–17) with moderate to severe antisocial behaviour received either management as usual (MAU; n=342) or 3–5 months of MST followed by MAU (n=342). Primary outcome was proportion of out-of-home placements. Secondary outcomes included offending data, service and criminal justice sector costs, participant wellbeing, and substance misuse, measured at baseline, 6, 12, and 18 months. We used logistic regression for the primary outcome and mixed-effects regression models for secondary outcomes.

Outcomes: At 18 months the treatment effect for out-of-home placement was not significant (OR 1·25, 95% CI 0·77–2·05; p=0·37). Time to first offence was also comparable but the number of offences was higher for the MST group at 18 months. There were consistent short-term symptom reductions from MST in some secondary outcomes, but no evidence of sustained superiority on most secondary outcomes. Conduct disorder diagnoses were reduced by >40% in both groups. Mean total service costs were not significantly different.

Interpretation: The findings do not support MST over MAU as the intervention of choice for adolescents with moderate to severe antisocial behaviour. MST achieves some early symptomatic gains on parent-rated outcomes, but not those based on independent records, which after 12 months favour MAU.

Funding: Department for Children, Schools and Families; Department of Health.

Research in context

Evidence before this study

We undertook a systematic review to identify randomised studies of Multisystemic Therapy (MST) for conduct disorder. We searched Embase, MEDLINE, and PsycINFO from inception to December 2016 using the terms “Multisystemic Therapy” or “MST” in combination with 49 terms covering conduct problems, to identify relevant RCTs and systematic reviews of MST published in the English language. The search terms were based on systematic searches originally conducted in 2012 by the National Collaborating Centre for Mental Health for National Institute for Health and Care Excellence (NICE) guidelines. We identified 495 papers with relevant abstracts, and full text screening of these yielded 22 primary randomised studies of MST for CD for inclusion. Previous reviews (eg, those for NICE) identified MST as a promising intervention for delinquent adolescents in reducing recidivism and improving individual and family pathology, mitigating this major public health problem; these findings justified the national rollout of MST in England and elsewhere in Europe. Our review, like others with similar scope, found the replicability of findings in some non-USA studies to be mixed, with MST failing in some reports to reduce antisocial behaviour more than usual services but even then often demonstrating significant economic advantages.

Added value of this study

To our knowledge, this is the only independently conducted, large-sample, community-based, superiority cost-effectiveness study assessing the medium-term effects and costs of MST. The study was conducted with the treatment developers’ full collaboration but with no involvement from them at any stage of data acquisition or data processing. Researchers were blinded to treatment condition and participants were representative of those likely to be referred to MST services in the UK. Treatment quality in all but one of the sites was well

above the carefully independently specified standard expected by the developers, and the majority (491, 75%) of the participants were retained; reliable data on out-of-home placement and offending were collected from official records even for untraced participants. No long-term benefit of MST was found, and no evidence of superior cost-effectiveness compared with management as usual (MAU). There was no indication of benefit in terms of reduction in custodial or other out-of-home arrangements, and there was a statistically significant beneficial effect associated with MAU versus MST in relation to offending behaviour at 18 months following recruitment. However, there was consistent evidence that MST brought about more rapid change in young people's behaviour as rated by their parents and, to a lesser extent, by themselves. Post-hoc analysis pointed to early-onset problems, and association with delinquent peers as contraindications for MST.

Implications of all the available evidence

Previous evidence from the USA and some European countries had suggested that MST is a very promising treatment, but the question of whether MST would be similarly effective in the UK had not been fully investigated before this study. Our results do not provide strong evidence for the continued national rollout of MST in child and adolescent health and social services. We found no evidence that major savings would ensue from further implementation of the model. The substantial improvements observed in both groups reflect the effectiveness of routinely offered interventions for this group of young people, at least when observed via trial methodology. Further post-hoc analysis of differences in MAU outcomes may provide suggestions for rational investment and/or disinvestment in this expensive domain of service provision. .

Introduction

Youth antisocial behaviour is a common and serious problem, with costly consequences for the young people, their families and wider society;[1] an elevated risk of health and social problems;[2] and a ten-fold increase in public sector costs by age 28.[3]

Multisystemic Therapy (MST) is an intensive family- and home-based intervention for young people with serious antisocial behaviour.[4] Recent high-quality, quantitative systematic reviews of 22 randomised controlled trials (RCTs)[1,5] identified MST as a promising intervention for improving the prognosis of adolescent antisocial and offending behaviour, mitigating public health impacts, and improving individual and family morbidity. However, outside the USA replicability of findings has been mixed, with MST failing to reduce antisocial behaviour more than usual services in some studies.[6-9]

A small UK-based RCT provided preliminary support for MST versus comprehensive targeted services delivered by Youth Offending Teams (YOTs) in reducing non-violent offending in the 18 months following randomisation.[10] The Systemic Therapy for At Risk Teens (START) study was a pragmatic multicentre superiority trial in which a large nationally representative sample of young people with moderate to severe antisocial behaviour were individually randomised to MST followed by management as usual (MAU) or MAU alone in order to determine the value added by MST in reducing the risk of out-of-home placements and criminal behaviour over the 18-month period following referral. The trial also assessed MST's impact on family relationships, wellbeing, educational performance, and cost-effectiveness, and the impact of previously identified moderating factors (callous–unemotional (CU) traits,[11] pre-adolescent onset,[12] delinquent peers[13]) and hypothesized mediators (parental attitudes and discipline practices[14]) in the context of a full economic evaluation.

Methods

Study design and participants

The study design and procedures are fully described in the published trial protocol.[15] (For the study protocol see http://www.ucl.ac.uk/start/START_research_protocol_v3_(Final)_05.11.2013.pdf) There were nine MST pilot sites in the UK with at least 12 months' experience of running the programme. Young people were recruited from social services, Youth Offending Teams (YOTs), schools, Child and Adolescent Mental Health Services (CAMHS), and voluntary services; all were referred to local multi-agency panels to standardise the referral process. These panels identified participants' suitability for MST (see below) and invited them for formal assessment.

All participants met one of five general antisocial behaviour inclusion criteria: (1) persistent (weekly) and enduring (≥6 months) violent and aggressive interpersonal behaviour; (2) at least one conviction plus three additional warnings, reprimands, or convictions; (3) a current DSM-IV diagnosis of CD that had not responded to treatment; (4) a permanent school exclusion for antisocial behaviour; (5) a significant risk of harm to others or self; and, additionally, at least three severity criteria indicating past difficulties across several settings (appendix). Exclusion criteria were kept to a minimum (appendix).

The MST supervisor and researcher visited proposed participants and their families to assess inclusion and exclusion criteria and discuss the trial, including the identification of an acceptable and credible MAU path. Written informed consent for randomisation was sought at the second visit, 3–7 days after the first, when a research assistant (RA) performed the baseline assessment. The study protocol was approved by the London South-East Research Ethics Committee (09/H1102/55).

Randomisation and masking

The RA initiated a secure randomisation by telephone from the trial centre (UCL), which in turn communicated to the referrer and family within 24 hours. Families were randomised to MST or MAU by an equal allocation ratio using stochastic minimisation, balancing for treatment centre, sex, current age (<15 or ≥15 because of differences in CAMHS service provision based on Gillick competence), and age at onset of antisocial behaviour (≤11 or >11, representing transition to secondary school with increased exposure to psychosocial risks and lower controls in the school environment). RAs remained blind to treatment allocation and were located separately to avoid leakage of trial information. Treatment fidelity assessments were carried out by a geographically separate research group without access to outcomes information. All coding, data entry, and data cleaning were done blind to allocation. Data were housed by a Mental Health Research Network data warehouse separate from the research teams. A sample (25%) of data was double-entered to reduce the chance of entry errors.

Interventions

Multisystemic Therapy (MST) is an intensive family- and home-based intervention for young people with serious antisocial behaviour.[4] The MST therapist works primarily with the young person's caregiver to improve parenting skills, enhance family relationships, increase support from social networks, develop skills and resources, address communication problems, encourage school attendance and achievement, and reduce the young person's association with delinquent peers. The intervention is tailored to each family's specific needs, using techniques from cognitive–behavioural, behavioural, and strategic and structural family therapies. Therapists meet the family three times a week for 3–5 months, and over this period are available 24 hours a day, 7 days a week.

Programme fidelity is maintained by (1) manualised weekly group supervision with an MST expert designated by MST Services;[16] (2) a well-developed quality assurance system[17] with twice-yearly implementation reviews; and (3) the Therapist Adherence Measure-Revised (TAM-R) based on independently administered interviews with parents.[18] All but one site averaged above criterion adherence (appendix).

Following MST, families received MAU from YOTs, CAMHS, and social and education services.

MAU was based on the best available local service(s) for the young person identified by the multi-agency referral panel and simply designed to be in line with current community practice informed by treatment guidelines offered on an as-needed basis.[1,19] MAU interventions were multicomponent, no less resource-intensive than MST, and consistent with the young people's complex mental health needs and behavioural difficulties.[20] Unlike MST, they were not coordinated in the context of a single overarching formulation, and were delivered without weekly expert supervision. No attempt was made to standardise MAU. See appendix for details of MAU interventions and services.

Outcomes

Outcome assessment measures were administered at baseline and 6, 12, and 18 months (primary endpoint chosen as at least 1 year after end of treatment to determine whether treatment gains were maintained). The primary outcome, chosen by the commissioners of the MST service because of high costs and poor long-term outcomes,[21] was the proportion of participants assigned to long-term (≥3 months) placement in specialist residential provision. We report a wide range of secondary outcomes, which reflect the diverse interests of Government policymakers who commissioned the investigation. To ensure comparability with other MST trials, antisocial behaviour was examined as time to first criminal offence

and the total number of offences, based on official records from the Police National Computer and Young Offender Information System. Further secondary outcomes were obtained from questionnaire measures concerning antisocial behaviour and attitudes, completed by parents and young people (Strengths and Difficulties Questionnaire [SDQ],[22] Inventory of Callous and Unemotional Traits[23]), by young people alone (Self-Report Delinquency Measure [SRDM], which includes a substance misuse scale,[24] Antisocial Beliefs and Attitudes Scale,[25] and Youth Materialism Scale)[26], and by teachers and parents (the ADHD scales from the Conners Comprehensive Behaviour Rating Scales [CBRS][27]). Intermediate outcome measures of parenting skills (Alabama Parenting Questionnaire [APQ][28]) and family functioning (Loeber Caregiver Questionnaire,[29] Family Adaptability and Cohesion Evaluation Scale [FACES-IV][30], Level of Expressed Emotion Questionnaire,[31] and Conflict Tactics Scale)[32] were completed by parents and/or young people, as appropriate. Only the Monitoring and Supervision subscale of the APQ is reported here, as it is central to adolescent antisocial behaviour.[33] Questionnaire measures concerning young people's and parental wellbeing and adjustment were completed by young people (Mood and Feelings Questionnaire [MFQ][34] and SDQ) and parents (SDQ, CBRS,[27] and General Health Questionnaire [GHQ])[35].

Data on educational participation (attendance and exclusions) were obtained from the National Pupil Database. Psychiatric disorders were identified at baseline and at 12 months by the Development and Well-Being Assessment (DAWBA).[36] Child IQ estimates were obtained using two subtests from the Wechsler Abbreviated Scale of Intelligence (WASI).[37] Two qualitative studies, to be reported separately, were also conducted with a subsample of families and professionals, exploring service characteristics and experiences of MST. We intended to use three additional questionnaires to characterise the nature and delivery of interventions in both the MST and MAU arms (the Expectancies Questionnaire,[38] the

California Psychotherapy Alliance Scale,[39] and the Reasons for Termination checklist)[40]. However, these measures were dropped following feedback from parents and young people about the burden of assessments and in consultation with the Trial Steering Committee. We intended to use the Child Attachment Interview to measure the quality of attachment relationships in a subsample of families.[41] However, the young people approached expressed concerns about completing the interview on camera (necessary for scoring) and no data were collected. All measures and schedules for data collection, together with observed reliability of the instruments, are described in the appendix.

Statistical analysis

On the basis of a previous UK trial[10] and official records, we anticipated that 30% of the MAU arm would have an out-of-home placement. We considered a reduction to 20% to be significant clinically and in terms of policy, and calculated that 700 participants would give 86% power to detect this difference (two-sided significance level of 5%). To take account of within-therapist correlation of outcomes in the MST arm, assuming based on a previous study[10] an intraclass correlation of 0·02 giving design effects of 1·22 in the MST arm and 1 in the MAU arm, power would be reduced to 83%. For the primary outcome, no loss to follow-up was expected, so this sample size was not increased.

Analysis was by intention to treat. The primary analyses entailed a logistic regression of out-of-home placement status at 18 months and a Cox regression for time-to-event outcomes for first criminal offence. Clustering by therapist was accounted for by including a random therapist effect. The logistic regression model included site, number of past convictions, sex, and age at onset of criminal behaviour as fixed effects, and was fitted using glmer() in the R package lme4 with a Wald test of the effect of intervention. Secondary outcomes were modelled using linear mixed-effects models (for continuous outcomes) adjusting for baseline values, and Poisson mixed models for count variables. For longitudinal outcomes, separate

treatment effects for 6-, 12-, and 18-month outcomes were used, together with two parameters representing the linear and quadratic time-trend in the outcome. Tests of interaction were planned to explore whether the intervention effects differed according to (1) sex, (2) age, (3) referral path, and (4) severity as indicated by the presence of a criminal record. Further non-prespecified moderator analyses were performed. These are exploratory and should be interpreted with caution.

As the primary outcome data were obtained independently of the subjects, negligible missing data were expected. For secondary outcomes, the analysis models used yield valid inferences under a missing-at-random assumption. As suggested by the Data Monitoring Committee, we performed a sensitivity analysis using post-baseline offending data (ie, total number of offences committed at each 6-month interval) as auxiliary variables in a multiple imputation analysis (appendix). As these made only minor differences to the results, the report is based on non-imputed outcomes; imputed outcomes are provided in the figures and the appendix. Statistical tests were deemed significant if their two-sided p value was <0·05. All analyses were performed in R version 3.3.0.

For the economic analysis, the costs and cost-effectiveness of treatment arms were compared at 18 months in terms of the proportion of participants requiring out-of-home placements. The economic evaluation took a broad societal perspective, including all health, social, education, and non-statutory sector services, as well as costs to the criminal justice sector resulting from crimes committed. Data on MST contacts to enable costing of the MST intervention were collected directly from pilot schemes to maintain the RAs' blindness to group allocation. RAs collected data on use of other services (number and duration of contacts) in interviews with families at baseline and at each follow-up using the Child and Adolescent Service Use Schedule (CA-SUS). The CA-SUS was based on previous economic studies in similar populations[42] and was adapted for use in the present study through a review

of the literature and pilot testing, to ensure comprehensive coverage and face validity. Data were collected in the following domains: delivery of MST intervention, accommodation services, education services, NHS secondary care services, community-based services, use of prescribed medication, out-of-pocket expenses, criminal justice system contacts, and criminal activity. The economic analysis uses all occurrences of criminal behaviour as reported in the CA-SUS rather than only convictions recorded in the Police National Computer or the Young Offender Information System database to capture all costs associated with criminal activity.

Unit costs for the financial year 2012–13 were applied to all resources used. The cost of MST was calculated using a standard micro-costing approach.[43] This involved estimation of indirect time spent on individual cases, including preparation, meetings, telephone calls and supervision, as well as detailed recording of face-to-face contacts. Unit costs were calculated using data on salaries, employer on-costs (National Insurance and superannuation), conditions of service, and appropriate administrative, managerial, and capital overheads, plus the cost of contributions from MST Services, which included MST training, MST supervision, and the MST licence. Nationally applicable unit costs were applied to all other services, including MAU. These are outlined in detail in the appendix, along with a costing schema for the MST intervention. Costs in the second year were discounted by 3·5%, as recommended by the National Institute for Health and Care Excellence.[44] Detailed information on the economic data and unit costs applied are provided in the appendix. For the cost-effectiveness analysis, we calculated incremental cost-effectiveness ratios (the difference in mean cost divided by the difference in mean effect) and explored uncertainty with cost-effectiveness acceptability curves, which show the probability that MST is the optimum choice, for a range of possible values of willingness to pay for improvements in outcome.[45] All economic analyses were adjusted for the prespecified covariates and for baseline cost and outcomes, as appropriate. Complete case analysis was used, with the effect of missing data

explored in sensitivity analyses. A prespecified secondary economic analysis using quality-adjusted life years measured by the three-level version of the EQ-5D[46,47] was planned but an administrative error at the start of the trial meant that the EQ-5D was not included in the outcome pack, resulting in extensive missing data, and this analysis had to be abandoned. In addition, out-of-pocket expenses had to be excluded from the cost-effectiveness analysis because of poor quality of reporting (less than one-quarter of the sample provided adequate data to enable these expenses to be costed).

This trial is registered with ISRCTN, number ISRCTN77132214.

Role of the funding source

Beyond the tender brief, funders had no role in the study design, data collection and analysis, or interpretation of the findings. Representatives of the funders and MST-UK were present at the Trial Steering Committee meetings and had the opportunity to comment on drafts of this paper. The corresponding author had full access to all the study data and had final responsibility for the decision to submit the findings for publication.

Results

Between February 4, 2010 and September 1, 2012, 1076 young people were referred to the nine multi-agency panels, the largest group from Children's Services and then YOTs (figure 1). Of these, 16% were inappropriate referrals for MST and a further 10% did not complete the referral process (4% refused to take part in the study and 6% turned down the clinical interventions on offer). The 684 who consented to baseline assessment and randomisation were clinically and demographically representative of appropriate referrals (for inclusion and exclusion criteria, see appendix). Of this sample, 85% was retained for 6-month assessment and 80% at 12 months. At the final time point more than three-quarters of those (491, 75%) who had not withdrawn from the study were available for assessment, with slightly fewer

from the MAU (234, 70%) than the MST (257, 77%) group; 91% of assessments were completed within 30 days of the assessment due date.

Three direct observational points were available for nearly 85% of the families. Official records were available for almost the complete sample (98%) for out-of-home placements, criminal convictions, and educational outcomes. Client and family baseline characteristics and moderators are displayed in table 1. The two groups were similar except there were slightly more young people with ADHD diagnoses in the MST arm. Over 80% of the sample met ICD-10 criteria for CD.

For the overall sample of 684 at baseline, 443 participants were identified to have persistent and enduring violent and aggressive interpersonal behaviour; 63 participants had at least one conviction plus three additional warnings, reprimands, or convictions; 531 currently met DSM-IV diagnosis of CD that had not responded to treatment; 179 participants had been permanently been excluded from school for antisocial behaviour; and 67 were at significant risk of harm to themselves (appendix). All 684 young people at baseline scored >65 on the WASI, with similar scores in the MST (mean 84·2, SD 13·2) and MAU (84·0, SD 13·2) groups.

Primary and key forensic outcomes

MST had no significant effect on the probability of out-of-home placement (12·6% *vs* 10·7%; OR 1·25, 95% CI 0·77 to 2·05; p=0·37) (table 2A), determined from a combination of parent-report and Local Authority computerised records. The key forensic analyses examined the time to first offence using a Cox proportional hazards model (table 2B). MST did not significantly delay the time to first offence (HR 1·06, 95% CI 0·84 to 1·33; p=0·64). The number of offences committed in 6-month periods after the end of the intervention, based on police records, are displayed in table 3. Overall, the numbers were low, with the mean

number of offences never exceeding 1. The Poisson mixed-effects model showed that a significantly higher mean number of offences were committed in the MST versus the MAU condition by 18 months (difference in mean number 0·65, 95% CI 0·28 to 1·02; p=0·00067). When violent and non-violent crimes were analysed separately, the difference was in the same direction, but not statistically significant. Reconviction rates cannot be reported because these were not reliably recorded on the databases available to the research team.

Secondary outcomes: Antisocial behaviour and attitudes

Further analyses of parent- and youth-reported secondary outcomes are reported in tables 4 to 6. Graphical illustrations are displayed in the appendix for summary results and individual variables alongside non-prespecified subscales and analyses based on multiple imputations.

Self-report and parent report of antisocial behaviour and attitudes (tables 4A and B) showed significant benefits from MST at 6 months, but mostly these were no longer significant by 12 months. Analysis of young people's self-ratings revealed smaller differences between the groups even at 6 months and no differences in self-reported behaviour on the SDQ at any time point. Self-reported attitudinal measures of antisociality yielded no group differences at any time, although CU traits were rated lower by young people in MST at 18 months. MST showed some benefit at 6 months on self-reported delinquency (SRDM) in terms of reduced volume and variety of substance misuse. Materialistic attitudes characteristic of conduct problems did not change significantly during the study period (table 4B).

Information obtained from the National Pupil Database indicated that MST had no significant effects on exclusion from school. The odds ratios (95% CI) for 6, 12, and 18 months were 1·00 (0·70 to 1·43), 0·93 (0·64 to 1·37), and 0·71 (0·45 to 1·13), respectively.

Secondary outcomes: Parenting skills and family functioning

Parents' reports of their own parenting behaviour (Loeber Caregiver Questionnaire and APQ Monitoring and Supervision subscale; table 5A) indicated increased parental support and involvement and reduced problems with monitoring and supervision in the MST group at 6 months. Young people's report on parenting behaviour on the APQ Monitoring and Supervision subscale or Level of Expressed Emotion (table 5A) indicated no significant effect of MST at any point. Parent-rated family functioning (FACES-IV) favoured the MST participants at 6 months, but differences were no longer significant at 18 months (table 5B). Parent reports of partner conflict on the CTS showed no significant group differences at any time point (table 5B).

Secondary outcomes: Young people's and parental wellbeing and adjustment

Young people's self-report of their emotional wellbeing on the SDQ and MFQ indicated statistically significant benefits from MST at 6 and 12 months but no differences at 18 months (table 6A). Parental reports of young people's wellbeing on the SDQ revealed some between-group differences but none were maintained at 18-month follow-up. On the parent-rated Conners ADHD scale, scores were significantly higher in the MAU condition at 6 months but not thereafter, but teachers were unable to detect this change (table 6B). Parental reports suggested larger effects at 6 and 12 months but these dissipated at 18 months (table 6B). Teachers' ratings using the other Conners behaviour rating scales (appendix) did not detect an impact of the MST intervention, although teachers reported less disruptive behaviour in the MST group at 12 months (estimate: –2·56, 95% CI –4·77 to –0·35; p=0·025). Parental wellbeing benefited from MST and differences on the GHQ continued to favour MST at 18 months post-baseline (table 6B). Clinician ratings on the DAWBA identified no significant between-group differences in psychiatric disorders at either baseline or 12 months (table 6C).

Moderator analyses

We considered several potential moderators (table 2). Onset of antisocial behaviour before 11 years powerfully moderated the effect of MST on out-of-home placements (interaction: OR 4·95, 95% CI 1·74 to 14·0; p=0·0026). There was a significant detrimental effect of MST (OR 3·11, 95% CI 1·40 to 6·93; p=0·0014) in the early-onset group when directly compared with the late-onset group, and a non-significant beneficial effect of MST in the late-onset group (OR 0·63, 95% CI 0·32 to 1·23; p=0·17).

CU traits at baseline also moderated the impact of MST on out-of-home placement (interaction: OR 0·95, 95% CI 0·90 to 1·00; p=0·048). MST was significantly detrimental relative to MAU in participants low on CU traits at baseline (those scoring below the median in CU traits) (OR 2·77, 95% CI 1·20 to 6·40; p=0·017). There was no significant moderating effect of high baseline CU traits on the MST group (OR 0·70, 95% CI 0·36 to 1·35; p=0·29).

In participants with few delinquent peers (≤the median peer delinquency score of 3), MST significantly decreased the time to first offence (HR 1·47, 95% CI 1·04 to 2·09; p=0·029), while in the group where delinquency was more socialised, MST significantly increased the time to first offence (HR 0·68, 95% CI 0·50 to 0·94; p=0·020).

Figure 2 shows Kaplan–Meier curves for each subgroup. Finally, there were no interaction effects with psychiatric comorbidities on these treatment outcomes.

The high level of provision (appendix) underscores (1) the participants' high service need and (2) the groups' comparability in terms of hours of face-to-face treatment, with almost no differences between the conditions, notwithstanding that the MST therapist contacts were not included in computing MAU.

Economic analyses

Total service costs and outcomes over the 18-month follow-up period are summarised in table 7, including a breakdown of costs by service-providing sector. The mean total costs over 18-month follow-up were £30,928 in the MAU group and £28,678 in the MST group; this difference was not statistically significant (adjusted difference –£1623, 95% CI –£7684 to £4438; p=0·60). The cost-effectiveness acceptability curve (appendix) indicates that the probability that MST is cost-effective compared with MAU is low and does not rise above 18% for a range of willingness-to-pay thresholds.

Discussion

We identified no long-term behavioural, mental health, social care, forensic, or educational benefit, nor any economic advantage, for this therapy compared with MAU by local services. MST may actually have worsened some of these outcomes for some young people. There was no evidence that MST reduces the likelihood of out-of-home placement; if anything, it was slightly increased, perhaps because of MST's greater attention to young people at risk triggering safeguarding arrangements for these young people. It should be noted that both arms achieved the reduction of 20% (from the actuarial estimate of 30% to the observed 10%) that we *a priori* identified as clinically significant.

In terms of the key secondary outcome of criminal behaviours, the reduction in convictions achieved by MST was no better than that achieved with MAU, and some advantage for MAU was noted by 18 months.

MST brought about change more rapidly than MAU, especially as noted by parents, although this change was no more likely to be sustained in the longer term. Parents valued MST even though its impact on participants dissipated by the end of the study. This may account for the improvements in parents' own overall mental health and reporting of improved family

functioning. Overall, and compared with the young people, parents may have somewhat benefitted from the MST programme, and sustained change in self-reported parenting in combination with improved mood may turn out to bring long-term behavioural benefit; this will be examined by an ongoing extended follow-up of this sample. In contrast, young people reported little change in parenting behaviour, including failing to confirm the lasting reduction in inconsistent parenting reported by parents in the MST group.

It is unclear why the young people themselves appeared less sensitive to the programme's benefits. Self-rated conduct problems and delinquent behaviour decreased across both groups with time. There were few between-group differences in antisocial attitudes, apart from an unpredicted difference in CU traits at 18 months favouring MST. Measures of emotional wellbeing (anxiety and depression) also indicated benefit from MST for the year following the interventions; the group differences were small in absolute terms and fell short of mean differences on the MFQ usually associated with clinical significance (5 points or more) but the pattern was statistically robust across two measures.

There was little indication of MST's educational benefit from either teachers or records of school attendance, although there were considerable missing data. Despite earlier pilot study evidence suggesting that MST led to cost savings,[48] in this larger economic evaluation there is no evidence that MST is more cost-effective than MAU. Although total costs were slightly lower, differences were not significant, and poorer outcomes in terms of out-of-home placement resulted in a low probability of MST being cost-effective compared with MAU.

Analysis of the severity moderators yielded findings worthy of further exploration. With early-onset antisocial behaviour, MST appeared to increase the likelihood of costly out-of-home placement, although it is possible that this was because close observation of family dynamics in MST revealed more instances where such placements were appropriate. MST

appeared to delay reoffending when delinquent peer influences were marked, while increasing risk of offending in young people without antisocial peers. MST appeared to be similarly detrimental relative to MAU for a low-risk group, namely low-CU individuals, whose time to first offence decreased following MST. The authors speculate that in relatively low-risk groups the focus of MST on criminal activity (eg, police involvement with acts of violence to family members as part of MST safety planning) may have the effect of enhancing adverse outcomes in individuals not previously sensitised to offending possibilities.

This trial is the most comprehensive study of MST reported so far and has a number of strengths. It was independently conducted, with the developers' collaboration but without their involvement at any stage of data acquisition or data processing. The participants were representative of individuals likely to be referred to MST services in the UK. We were able to independently assure treatment quality, all but one of the sites performed well above the standards expected by the developers, and no information on treatment assignment was available to anyone on the research team. The study retained the vast majority of participants, and reliable data on offending and out-of-home placement were collected for almost all participants. Multiple imputations using available data ensured representativeness of estimates where the young people, parents, or educators were unable to provide information. Outcomes covered the principal domains of interest, including offending; out-of-home placements; parent, educator, and self-rated behaviour; emotional wellbeing; family functioning; and societal and service costs. A putative mediator variable (parenting) was also incorporated.

However, significant limitations remain. The MAU group was not a homogenous comparison condition, with considerable between-site variation of what was offered. Future analysis will reveal whether differences between services significantly influenced outcomes. MAU may

have offered more flexibility in addressing the young people's specific needs, as opposed to MST, which focuses more on helping the family bring about behavioural change. While MST allows flexibility in the way specific problems are targeted, it also requires a high level of adherence to the interventions used, which may carry disadvantages. While the implementation of MST met formal fidelity criteria, the current average fidelity ratings for UK services significantly exceed levels achieved by these first-generation services. However, failures to replicate USA RCTs of interventions for youth antisocial behaviour are more likely due to the greater effectiveness of usual treatment rather than limitations of the UK implementation. A recent UK trial of Functional Family Therapy likewise found no improvement compared with controls, despite adequate implementation.[49] We tested a large number of secondary outcomes, so some significant results may be attributable to multiple testing and, along with our moderator analyses, are best considered exploratory and requiring replication. While the Cronbach's alpha (interclass reliability) coefficients were high or acceptable, some of the mean inter-item correlations (appendix) were outside the 0·15–0·20 range recommended as an indication of reasonable scale internal consistency.[50]

In conclusion, this rigorous and comprehensive evaluation found that MST did not significantly reduce dependence on MAU and brought no long-term advantages in terms of outcome. Although parents saw MST as bringing about more rapid and effective change, this was not reflected in objective indicators of delinquency. The medium-term gain from MST relative to MAU is limited in the behavioural domain, with some suggestion of adverse effect of MST in increased risk of criminal activity for individuals who are relatively low in risk in terms of the factors assessed in this study.

The findings also reflect the effectiveness of UK mental health, youth offending, and social care services, which were active in both arms of the trial, in reducing the risk of crime and

protecting young people and society, at least when under the scrutiny of a randomised controlled trial.

Contributors

PF, SBu, DC, SS, SP, IE, PFu, SBy, and IMG were responsible for the original proposal, for securing funding for the trial, and for drafting the original protocol with assistance from EA. PF as chief investigator had overall responsibility for the management of the study, with support from SBu as clinical research lead. IMG had responsibility for the East Anglia site; DC and AK for the Northern site; and IE, SS, and PFu for the South-East sites. SP, PFu and SBu were responsible for the development of the measure of MAU interventions. RE was project manager throughout the trial and developed and coordinated the randomisation and minimisation protocol. ES was the Senior Research Assistant and supported RE with the trial coordination. RE (with PF and SBu) set up and coordinated the database, with all data held in a single repository managed by the MHRN at the East Anglia site. RE and ES coordinated and supervised the treatment fidelity project managed by MST Inc. JW and SBy wrote the statistical analysis plan. JW, PG, and SBy did the statistical analyses. RE and ES were responsible for data cleaning. PF wrote the initial draft of the manuscript with support from EA. All authors contributed to and approved the final manuscript.

Declaration of interests

We declare no competing interests.

Acknowledgments

This study received funding from the Department for Children, Schools and Families in conjunction with the Department of Health. Peter Fonagy is in receipt of a National Institute for Health Research (NIHR) Senior Investigator Award (NF-SI-0514-10157), and was in part supported by the NIHR Collaboration for Leadership in Applied Health Research and Care

(CLAHRC) North Thames at Barts Health NHS Trust. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. Nicole Hickey (Imperial College London) entered, coded, analysed and prepared tables for the youth offending data from official records from the Police National Computer and Youth Offender Information System for all nine MST sites. The authors acknowledge the excellent work of the team of RAs involved in acquiring the data across the nine MST sites; the wise counsel of Professor Eric Taylor and members of the Trial Steering Committee he chaired, and of Professor Philip Graham and members of the Data Management and Ethics Committee he chaired; and the consistent and thoughtful support and oversight of Cathy James, National Multisystemic Therapy Programme Lead, National Implementation Service supported by Kathryn Harney, Associate Director of Research, Greater Manchester West Mental Health NHS Foundation Trust, without whom this trial could not have been completed.

References

1. National Institute for Health and Clinical Excellence. Conduct disorders and antisocial behaviour in children and young people: Recognition, intervention and management (CG158). London, UK: British Psychological Society and Royal College of Psychiatrists, 2013.
2. Colman I, Murray J, Abbott RA, et al. Outcomes of conduct problems in adolescence: 40 year follow-up of national cohort. *BMJ* 2009; **338**: a2981.
3. Khan L, Parsonage M, Stubbs J. Investing in children's mental health: A review of evidence on the costs and benefits of increased service provision. London, UK: Centre for Mental Health, 2015.
4. Henggeler SW, Schoenwald SK, Borduin CM, Rowland MD, Cunningham PB. Multisystemic therapy for antisocial behavior in children and adolescents. 2nd ed. New York: Guilford Press; 2009.
5. van der Stouwe T, Asscher JJ, Stams GJ, Dekovic M, van der Laan PH. The effectiveness of Multisystemic Therapy (MST): a meta-analysis. *Clin Psychol Rev* 2014; **34**(6): 468-81.
6. Leschied AW, Cunningham A. Seeking effective interventions for serious young offenders: Interim results of a four-year randomized study of multisystemic therapy in Ontario, Canada. London, ON: Centre for Children and Families in the Justice System, 2002.
7. Sundell K, Hansson K, Lofholm CA, Olsson T, Gustle LH, Kadesjo C. The transportability of multisystemic therapy to Sweden: Short-term results from a randomized trial of conduct-disordered youths. *J Fam Psychol* 2008; **22**(4): 550-60.
8. Ogden T, Christensen B, Sheidow AJ, Holth P. Bridging the gap between science and practice: The effective nationwide transport of MST programs in Norway. *J Child Adolesc Subst Abuse* 2008; **17**(3): 93-109.

9. Gustle LH, Hansson K, Sundell K, Andree-Lofholm C. Implementation of evidence-based models in social work practice: practitioners' perspectives on an MST trial in Sweden. *J Child Adolesc Subst Abuse* 2008; **17**(3): 111-25.
10. Butler S, Baruch G, Hickey N, Fonagy P. A randomized controlled trial of multisystemic therapy and a statutory therapeutic intervention for young offenders. *J Am Acad Child Adolesc Psychiatry* 2011; **50**(12): 1220-35.
11. Wilkinson S, Waller R, Viding E. Practitioner Review: Involving young people with callous unemotional traits in treatment—does it work? A systematic review. *J Child Psychol Psychiatry* 2016; **57**(5): 552-65.
12. Bakker MJ, Greven CU, Buitelaar JK, Glennon JC. Practitioner Review: Psychological treatments for children and adolescents with conduct disorder problems – a systematic review and meta-analysis. *J Child Psychol Psychiatry* 2017; **58**(1): 4-18.
13. Dopp AR, Borduin CM, White MH, Kuppens S. Family-based treatments for serious juvenile offenders: A multilevel meta-analysis. *J Consult Clin Psychol* 2017; **85**(4): 335-54.
14. Patel CC, Fairchild AJ, Prinz RJ. Potential mediators in parenting and family intervention: Quality of mediation analyses. *Clin Child Fam Psychol Rev* 2017; **20**(2): 127-45.
15. Fonagy P, Butler S, Goodyer I, et al. Evaluation of multisystemic therapy pilot services in the Systemic Therapy for At Risk Teens (START) trial: study protocol for a randomised controlled trial. *Trials* 2013; **14**: 265.
16. Strother KB, Swenson ME, Schoenwald SK. Multisystemic therapy organizational manual. Charleston, SC: MST Institute; 1998.
17. Schoenwald SK, Garland AF. A review of treatment adherence measurement methods. *Psychol Assess* 2013; **25**(1): 146-56.
18. Huey SJ, Jr., Henggeler SW, Brondino MJ, Pickrel SG. Mechanisms of change in multisystemic therapy: reducing delinquent behavior through therapist adherence and improved family and peer functioning. *J Consult Clin Psychol* 2000; **68**(3): 451-67.
19. National Institute for Health and Clinical Excellence. Antisocial personality disorder: Treatment, management and prevention. London, UK: The British Psychological Society and the Royal College of Psychiatrists, 2010.
20. Chitsabesan P, Kroll L, Bailey S, et al. Mental health needs of young offenders in custody and in the community. *Br J Psychiatry* 2006; **188**(6): 534-40.
21. Bazalgette L, Rahilly T, Trevelyan G. Achieving emotional wellbeing for looked after children. London, UK: NSPCC, 2015.
22. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 1997; **38**(5): 581-6.
23. Essau CA, Sasagawa S, Frick PJ. Callous-unemotional traits in a community sample of adolescents. *Assessment* 2006; **13**(4): 454-69.
24. Smith DJ, McVie S. Theory and method in the Edinburgh study of youth transitions and crime. *Br J Criminol* 2003; **43**(1): 169-95.
25. Butler SM, Leschied AW, Fearon P. Antisocial beliefs and attitudes in pre-adolescent and adolescent youth: the development of the antisocial beliefs and attitudes scales (ABAS). *J Youth Adolesc* 2007; **36**(8): 1058-71.
26. Goldberg ME, Gorn GJ, Peracchio LA, Bamossy G. Understanding materialism among youth. *J Consum Psychol* 2003; **13**(3): 278-88.
27. Conners CK. Conners' Rating Scales-Revised: technical manual. Toronto: Multi-Health Systems; 1995.
28. Elgar FJ, Waschbusch DA, Dadds MR, Sigvaldason N. Development and validation of a short form of the Alabama Parenting Questionnaire. *J Child Fam Stud* 2006; **16**(2): 243-59.

29. Loeber R, Stouthamer-Loeber M, Van Kammen W, Farrington DP. Initiation, escalation and desistance in juvenile offending and their correlates. *J Crim Law Criminol* 1991; **82**(1): 36-82.
30. Olson D. FACES IV and the Circumplex Model: validation study. *J Marital Fam Ther* 2011; **37**(1): 64-80.
31. Gerlsma C, Hale WW, 3rd. Predictive power and construct validity of the Level of Expressed Emotion (LEE) scale. Depressed out-patients and couples from the general community. *Br J Psychiatry* 1997; **170**(6): 520-5.
32. Straus MA. Measuring intrafamily conflict and violence: the Conflict Tactics (CT) Scales. In: Straus MA, Gelles RJ, eds. Physical violence in American families: Risk factors and adaptations to violence in 8,145 families. New Brunswick, NJ: Transaction Publishing; 1990: 403-24.
33. Dishion TJ, McMahon RJ. Parental monitoring and the prevention of child and adolescent problem behavior: a conceptual and empirical formulation. *Clin Child Fam Psychol Rev* 1998; **1**(1): 61-75.
34. Messer SC, Angold A, Costello EJ, Loeber R, van Kammen W, Stouthamer-Loeber M. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents: factor composition and structure across development. *Int J Methods Psychiatr Res* 1995; **5**(4): 251-62.
35. Goldberg DP, Gater R, Sartorius N, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Med* 1997; **27**(1): 191-7.
36. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: Description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry* 2000; **41**(5): 645-55.
37. Wechsler D. Wechsler Abbreviated Scale of Intelligence (WASI). Oxford: Harcourt Assessment; 1999.
38. Shaw BF, Elkin I, Yamaguchi J, et al. Therapist competence ratings in relation to clinical outcome in cognitive therapy of depression. *J Consult Clin Psychol* 1999; **67**(6): 837-46.
39. Gaston L, Thompson L, Gallagher D, Cournoyer L-G, Gagnon R. Alliance, technique, and their interactions in predicting outcome of behavioral, cognitive, and brief dynamic therapy. *Psychother Res* 1998; **8**(2): 190-209.
40. Moras KK. Early termination and the outcome of psychotherapy: patients' perspectives. *Diss Abstr Int* 1986; **46**(B): 2817-8.
41. Shmueli-Goetz Y, Target M, Fonagy P, Datta A. The child attachment interview: A psychometric study of reliability and discriminant validity. *Dev Psychol* 2008; **44**(4): 939-56.
42. Barrett B, Byford S, Chitsabesan P, Kenning C. Mental health provision for young offenders: service use and cost. *Br J Psychiatry* 2006; **188**(6): 541-6.
43. Drummond MF, Sculpher J, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2015.
44. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. London, UK: National Institute for Health and Care Excellence, 2013.
45. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ* 2001; **10**(8): 779-87.
46. Brooks R. EuroQol: the current state of play. *Health Policy* 1996; **37**(1): 53-72.
47. Byford S. The validity and responsiveness of the EQ-5D measure of health-related quality of life in an adolescent population with persistent major depression. *J Ment Health* 2013; **22**(2): 101-10.

48. Cary M, Butler S, Baruch G, Hickey N, Byford S. Economic evaluation of multisystemic therapy for young people at risk for continuing criminal activity in the UK. *PLoS One* 2013; **8**(4): e61070.
49. Humayun S, Herlitz L, Chesnokov M, Doolan M, Landau S, Scott S. Randomized controlled trial of Functional Family Therapy for offending and antisocial behavior in UK youth. *J Child Psychol Psychiatry* 2017; **58**(9): 1023-32.
50. Clark LA, Watson D. Constructing validity: Basic issues in objective scale development. *Psychol Assess* 1995; **7**(3): 309-19.

Figure captions

Figure 1: Trial profile

Figure 2: Time to first offence of young people with high or low levels of peer delinquency

Del=peer delinquency. MAU=management as usual. MST=Multisystemic Therapy.

Table 1. Baseline characteristics

	MST		MAU	
	n or mean	**SD or %**	**n**	**SD or %**
Demographics				
Number	342		342	
Mean age (years)	13·7	1·4	13·9	1·4
Female	126	36·8	124	36·4
Mean SES (range 1–6)	3·0	1·4	2·9	1·3
Family income				
% on state benefits or <£20k pa	258	75·4	267	78·3
Ethnicity				
White British/European	261	76·5	274	80·4
Black African/Afro-Caribbean	38	11·1	33	9·7
Asian	6	1·8	10	2·9
Mixed/Other	34	10	17	5
Parents' marital status				
Single or widowed	142	41·5	131	38·4
Separated or divorced	77	22·5	59	17·3
Married or cohabiting	123	36	147	43·1
Number of siblings	2·5	1·3	2·5	1·4
Siblings offending	118	36·9	126	39·4
Offences in year prior to referral				
Non-offender on referral	124	36·5	111	32·7
Total number of offences	1·1	2·2	1·2	2·5
Violent offences	0·4	1	0·4	0·9
Non-violent offences	0·5	1·2	0·6	1·3
Number with custodial sentences	4	1·2	6	1·8
Comorbid diagnosis				
Conduct disorder	262	77·7	270	79·4
Oppositional defiant disorder	14	4·2	14	4·1
Any conduct disorder	274	81·3	280	82·4
Social phobia	12	3·6	9	2·6
Obsessive-compulsive disorder	1	0·3	2	0·6
Posttraumatic stress disorder	25	7·4	26	7·6
Separation anxiety disorder	7	2·1	15	4·4
Specific phobia	6	1·8	13	3·8
Generalised anxiety disorder	6	1·8	9	2·6
Panic disorder	5	1·5	3	0·9
ADHD Combined	113	33·5	91	26·8

ADHD Hyperactive–Impulsive	8	2·4	3	0·9
ADHD Inattentive	13	3·9	12	3·5
PDD/autism	3	0·9	4	1·2
Eating disorders	2	0·6	2	0·6
Tic disorder	7	2·1	4	1·2
Major depression	30	8·9	42	12·4
Any emotional disorder	73	21·7	90	26·5
Mixed anxiety/conduct disorder	46	13·6	56	16·5
Number without diagnosis	50	14·8	50	14·7
Average number of Axis I diagnoses	1·5	1	1·5	1·1
Onset of conduct disorder	148	43·3	149	43·7
ICUT score	33·5	9·7	32·7	9·6
Peer delinquency score (SRDM)	5·0	4·7	4·9	4·7

ADHD=attention deficit hyperactivity disorder. ICUT=Inventory of Callous and Unemotional Traits. MAU=management as usual. MST=Multisystemic Therapy. PDD=pervasive developmental disorder. SES=socioeconomic status. SRDM=Self-Report Delinquency Measure. We have not tested for baseline differences in line with custom and practice in large trials recommendations.[43]

Table 2: A. Results of logistic regression of out-of-home placement and Cox proportional hazards model of time to first offence. B. Results of moderator analyses, where additional variables were included as interaction parameters in the primary analysis models

A. Analyses

Outcome	Effect of MST	95% CI	p value
Out-of-home placement (OR)	1·25	(0·77 to 2·05)	0·37
Time to first offence (HR)	1·06	(0·84 to 1·33)	0·64

B. Moderator analysis

Variable	Out-of-home placement			Time to first offence		
	Interaction OR	95% CI for OR	p value	Interaction HR	95% CI for HR	p value
Sex	1·01	(0·38 to 2·74)	0·98	0·94	(0·87 to 1·03)	0·19
Age	0·91	(0·63 to 1·33)	0·64	1·25	(0·97 to 1·62)	0·084
Early-onset CD	4·95	(1·74 to 14·0)	0·0026	1·19	(0·75 to 1·89)	0·47
Baseline ICUT score†	0·95	(0·90 to 1·00)	0·048	1·01	(0·99 to 1·03)	0·49
Baseline peer delinquency score†	0·91	(0·81 to 1·01)	0·085	0·92	(0·88 to 0·97)	0·00071
Baseline ABAS score†	1·00	(0·98 to 1·03)	0·69	1·00	(0·99 to 1·01)	0·93
No prior offence at baseline	0·53	(0·25 to 1·11)	0·39	NA*	NA*	NA*
CD + ADHD at baseline†	0·53	(0·18 to 1·58)	0·25	1·31	(0·79 to 2·17)	0·29
CD + depression at baseline†	1·29	(0·25 to 6·55)	0·76	0·94	(0·43 to 2·03)	0·87
Referral path‡	0·22	(0·02 to 2·48)	0·22	0·73	(0·39 to 1·35)	0·31

ABAS=Antisocial Beliefs and Attitudes Scale. ADHD=attention deficit hyperactivity disorder. CD=conduct disorder. HR=hazard ratio. ICUT=Inventory of Callous and Unemotional Traits. MST=Multisystemic Therapy. OR=odds ratio. *Non-offender at baseline parameter was not identifiable in the analysis of time to first offence as no individuals who were non-offenders at baseline went on to offend during the trial. All analyses are also adjusted for fixed centre effects. Each interaction is between a baseline measurement and treatment arm, and in each case regressions included all variables used in the main analysis but with an additional interaction term (and the main effect if the variable was not originally adjusted for). †Non-prespecified but recommended by the trial's Data Monitoring Committee. ‡Most significant result out of six referral path tests.

Table 3: Secondary key forensic and behavioural outcomes: Police records of offending behaviour and recorded crimes

	Group (n) and between-group significance	Proportion free of offending behaviour (%)	All crimes Mean (SD) [% offenders]	Violent crimes Mean (SD) [% offenders]	Non-violent crimes Mean (SD) [% offenders]
6 months–baseline	MST (n=338)	63·4	0·7 (1·5) [32]	0·24 (0·7) [17]	0·3 (0·7) [21]
	MAU (n=338)	67·6	0·7 (1·4) [37]	0·24 (0·6) [16]	0·4 (0·8) [23]
6-month follow-up	MST (n=338)	75·3	0·5 (1·2) [25]	0·2 (0·7) [13]	0·2 (0·6) [14]
	MAU (n=338)	71·1	0·7 (1·6) [29]	0·2 (0·6) [14]	0·3 (0·8) [18]
	Effect (95% CI)	Not estimated	–0·21 (–0·55 to 0·13)	–0·12 (–0·59 to 0·35)	–0·28 (–0·69 to 0·13)
	p value	Not estimated	0·23	0·61	0·18
12-month follow-up	MST (n=338)	77·4	0·5 (1·4) [23]	0·2 (0·7) [11]	0·2 (0·8) [13]
	MAU (n=338)	75·8	0·6 (1·4) [24]	0·2 (0·6) [11]	0·2 (0·6) [14]
	Difference (95% CI)	Not estimated	0·02 (–0·33 to 0·37)	–0·02 (–0·5 to 0·46)	0·2 (–0·23 to 0·63)
	p value	Not estimated	0·91	0·94	0·37
18-month follow-up	MST (n=338)	80·3	0·5 (1·7) [20]	0·2 (0·7) [8]	0·2 (0·8) [10]
	MAU (n=338)	84·4	0·3 (0·8) [16]	0·1 (0·3) [6]	0·1 (0·4) [8]
	Difference (95% CI)	Not estimated	0·65 (0·28 to 1·02)	0·51 (–0·05 to 1·07)	0·48 (–0·01 to 0·97)
	p value	Not estimated	0·00067	0·076	0·052

Data were obtained from the Police National Computer database. The models are mixed-effects random intercept logistic models for binary data and Poisson regression models for count data. MAU=management as usual. MST=Multisystemic Therapy. *Proportion free of offending behaviour was not tested for difference between arms as it was not a prespecified endpoint.

Table 4: Secondary outcomes: A. Parent report and young people's self-report of antisocial behaviour and callous–unemotional traits.

B. Young people's self-report of delinquent behaviour, antisocial beliefs and attitudes, and materialism.

A.

	Group and between-group significance (*t*-test)	SDQ conduct problems (YP) Mean (SD) [n]	SDQ conduct problems (P) Mean (SD) [n]	ICUT (YP) Mean (SD) [n]	ICUT (P) Mean (SD) [n]
6 months–baseline	MST	5·0 (2·1) [n=340]	6·59 (2·41) [n=340]	33·5 (9·7) [n=341]	42·91 (11·58) [n=341]
	MAU	4·9 (2·3) [n=340]	6·62 (2·45) [n=340]	32·7 (9·6) [n=339]	41·96 (11·74) [n=339]
6-month follow-up	MST	4·2 (2·1) [n=290]	4·8 (2·5) [n=290]	30·4 (9·9) [n=292]	36·1 (11·1) [n=292]
	MAU	4·4 (2·1) [n=264]	5·5 (2·5) [n=268]	30·8 (9·4) [n=268]	39·5 (12·3) [n=268]
	Effect	–0·26 (–0·57 to 0·05)	–0·62 (–0·99 to –0·25)	–0·92 (–2·31 to 0·47)	–4·61 (–6·37 to –2·85)
	p value	0·11	<0·0001	0·20	<0·0001
12-month follow-up	MST	3·8 (2·2) [n=252]	4·6 (2·6) [n=246]	29·0 (9·5) [n=248]	36·3 (12·6) [n=248]
	MAU	4·0 (2·2) [n=237]	4·8 (2·7) [n=237]	29·3 (9·9) [n=238]	36·1 (12·0) [n=238]
	Difference (95% CI)	–0·22 (–0·55 to 0·11)	–0·25 (–0·66 to 0·16)	–0·59 (–2·10 to 0·92)	–0·55 (–2·43 to 1·33)
	p value	0·20	0·22	0·44	0·57
18-month follow-up	MST	3·5 (2·0) [n=221]	4·4 (2·5) [n=232]	29·1 (9·8) [n=234]	35·0 (12·5) [n=234]
	MAU	3·4 (1·9) [n=193]	4·6 (2·5) [n=209]	30·9 (9·4) [n=217]	34·9 (11·9) [n=217]
	Difference (95% CI)	–0·07 (–0·42 to 0·28)	–0·16 (–0·57 to 0·25)	–1·92 (–3·39 to –0·45)	–0·69 (–2·61 to 1·23)
	p value	0·69	0·46	0·011	0·48

Data were obtained using the Strengths and Difficulties Questionnaire (SDQ) and Inventory of Callous and Unemotional Traits (ICUT) MAU=management as usual. MST=Multisystemic Therapy. P=completed by parent. YP=completed by young person.

B.

	Group (n) and between-group significance (*t*-test)	SRDM Variety of delinquent acts Mean (SD) [n]	SRDM Volume of delinquent acts Mean (SD) [n]	SRDM Variety of substance misuse Mean (SD) [n]	SRDM Volume of substance misuse Mean (SD) [n]	SRDM Peer delinquency Mean (SD) [n]	ABAS Mean (SD) [n]	Youth Materialism Scale Mean (SD) [n]
6 months–baseline	MST	4·8 (3·6) [n=337]	19·7 (18·3) [n=337]	0·8 (1·7) [n=337]	1·6 (3·7) [n=337]	5·0 (4·7) [n=337]	60·8 (23·1) [n=341]	38·8 (8·4) [n=342]
	MAU	3·1 (3·7) [n=335]	20·9 (19·0) [n=335]	0·7 (1·3) [n=335]	1·5 (3·0) [n=335]	4·9 (4·7) [n=335]	61·7 (24·4) [n=339]	38·7 (8·8) [n=341]
6-month follow-up	MST	3·9 (3·5) [n=288]	15·6 (17·0) [n=288]	0·7 (1·5) [n=288]	1·4 (3·0) [n=288]	4·9 (4·5) [n=288]	55·5 (24·7) [n=292]	37·2 (8·8) [n=293]
	MAU	4·5 (3·9) [n=262]	18·0 (18·1) [n=262]	0·8 (1·5) [n=262]	1·8 (3·2) [n=262]	4·7 (4·7) [n=262]	57·1 (23·6) [n=268]	37·9 (9·0) [n=263]
	Effect (95% CI)	–0·15 (–0·35 to 0·05)	–0·18 (–0·38 to 0·02)	–0·41 (–0·68 to –0·14)	–0·13 (–0·25 to –0·01)	0·12 (–0·70 to 0·94)	–1·58 (–4·81 to 1·65)	–0·75 (–2·06 to 0·56)
	p value	0·12	0·068	0·0033	0·016	0·77	0·34	0·26
12-month follow-up	MST	3·3 (3·5) [n=243]	12·3 (16·3) [n=243]	0·8 (1·9) [n=243]	1·9 (4·0) [n=243]	5·0 (5·2) [n=243]	54·4 (24·2) [n=248]	36·0 (9·5) [n=252]
	MAU	3·4 (3·4) [n=230]	12·7 (14·4) [n=230]	0·7 (1·2) [n=230]	1·5 (2·6) [n=230]	5·2 (4·9) [n=230]	55·5 (22·2) [n=238]	36·6 (8·9) [n=238]
	Difference (95% CI)	–0·02 (–0·22 to 0·18)	–0·15 (–0·35 to 0·05)	–0·06 (–0·33 to 0·21)	–0·03 (–0·15 to 0·09)	–0·16 (–1·02 to 0·70)	–0·75 (–4·14 to 2·64)	–0·53 (–1·90 to 0·84)
	p value	0·86	0·15	0·68	0·66	0·72	0·67	0·45
18-month follow-up	MST	2·8 (3·3) [n=231]	10·0 (13·7) [n=231]	0·7 (1·4) [n=231]	1·5 (2·6) [n=231]	4·6 (5·0) [n=231]	53·1 (24·6) [n=234]	36·3 (9·5) [n=241]
	MAU	2·4 (2·6) [n=215]	9·2 (11·2) [n=215]	0·7 (1·2) [n=215]	1·4 (2·3) [n=215]	4·7 (5·1) [n=215]	51·4 (22·7) [n=217]	37·1 (9·0) [n=211]
	Difference (95% CI)	0·17 (–0·05 to 0·39)	0·04 (–0·16 to 0·24)	–0·12 (–0·41 to 0·17)	–0·02 (–0·14 to 0·10)	–0·03 (–0·91 to 0·85)	2·63 (–0·86 to 6·12)	–0·45 (–1·88 to 0·98)
	p value	0·11	0·73	0·44	0·72	0·95	0·14	0·54

Data were obtained using the Self-Report Delinquency Measure (SRDM), Antisocial Beliefs and Attitudes Scale (ABAS), and Youth Materialism Scale. MAU=management as usual. MST=Multisystemic Therapy.

Table 5: Secondary outcomes: A. Young people's and parent's report on parenting skills and family functioning. B. Parents' report on family functioning.

A.

	Group and between-group significance (*t*-test)	APQ Problems of monitoring and supervision (YP) Mean (SD) [n]	APQ Problems of monitoring and supervision (P) Mean (SD) [n]	Loeber parental support score (P) Mean (SD) [n]	Level of Expressed Emotion (YP) Mean (SD) [n]
6 months–baseline	MST	8·4 (3·0) [n=341]	9·29 (3·33) [n=341]	44·44 (6·40) [n=337]	88·8 (20·0) [n=341]
	MAU	8·8 (2·8) [n=339]	9·37 (3·34) [n=339]	44·57 (6·04) [n=335]	89·1 (19·1) [n=339]
6-month follow-up	MST	7·7 (2·9) [n=292]	7·7 (3·2) [n=292]	47·7 (5·7) [n=288]	83·1 (18·9) [n=292]
	MAU	7·9 (2·9) [n=261]	8·5 (3·4) [n=268]	45·4 (6·6) [n=262]	85·3 (18·3) [n=268]
	Effect (95% CI)	–0·12 (–0·59 to 0·35)	–0·71 (–1·20 to –0·22)	2·05 (1·09 to 3·01)	–2·58 (–5·32 to 0·16)
	p value	0·62	0·0039	<0·0001	0·065
12-month follow-up	MST	7·7 (3·1) [n=246]	7·8 (3·3) [n=248]	46·7 (6·4) [n=243]	81·4 (19·7) [n=248]
	MAU	7·9 (3·1) [n=233]	8·1 (3·3) [n=238]	45·5 (6·4) [n=230]	82·3 (17·3) [n=238]
	Difference (95% CI)	0·02 (–0·49 to 0·53)	–0·10 (–0·61 to 0·41)	0·88 (–0·14 to 1·90)	–0·82 (–3·72 to 2·08)
	p value	0·94	0·72	0·093	0·58
18-month follow-up	MST	7·6 (3·1) [n=235]	7·7 (3·2) [n=234]	46·0 (7·1) [n=231]	78·7 (19·3) [n=234]
	MAU	7·7 (3·0) [n=206]	7·7 (3·4) [n=217]	44·9 (6·8) [n=215]	79·9 (18·6) [n=217]
	Difference (95% CI)	0·14 (–0·39 to 0·67)	0·11 (–0·42 to 0·64)	0·76 (–0·30 to 1·82)	–0·98 (–3·96 to 2·00)
	p value	0·61	0·70	0·16	0·52

Data were obtained using the Alabama Parenting Questionnaire (APQ), Loeber Caregiver Questionnaire (Loeber), and Level of Expressed Emotion. MAU=management as usual. MST=Multisystemic Therapy. YP=completed by young person. P=completed by parent. Other APQ scales are reported in the appendix.

B.

	Group (n) and between-group significance (*t*-test)	FACES-IV family satisfaction Mean (SD)	FACES-IV cohesion Mean (SD)	FACES-IV family communication Mean (SD)	CTS Mean (SD)
6 months–baseline	MST (n=337)	27·94 (8·73)	53·36 (22·54)	34·24 (8·11)	8·90 (9·87)
	MAU (n=335)	28·24 (9·09)	53·52 (24·21)	34·22 (8·55)	8·77 (9·72)
6-month follow-up	MST (n=288)	33·7 (8·0)	61·3 (18·7)	37·4 (7·2)	7·1 (9·9)
	MAU (n=262)	30·0 (9·1)	55·7 (22·1)	35·0 (8·4)	8·0 (8·1)
	Effect (95% CI)	3·85 (2·60 to 5·10)	5·80 (2·49 to 9·11)	2·60 (1·48 to 3·72)	–1·13 (–2·85 to 0·59)
	p value	<0·0001	0·00059	<0·0001	0·20
12-month follow-up	MST (n=243)	32·6 (8·6)	59·5 (20·8)	37·3 (7·0)	6·1 (9·5)
	MAU (n=230)	30·5 (8·9)	56·5 (22·3)	36·2 (8·1)	6·6 (7·1)
	Difference (95% CI)	2·43 (1·10 to 3·76)	3·27 (–0·20 to 6·74)	0·99 (–0·21 to 2·19)	0·10 (–1·74 to 1·94)
	p value	0·00037	0·065	0·11	0·92
18-month follow-up	MST (n=231)	32·5 (8·2)	59·7 (20·2)	37·7 (6·9)	5·0 (8·7)
	MAU (n=215)	32·0 (9·5)	59·5 (21·5)	37·0 (8·1)	4·8 (5·5)
	Difference (95% CI)	0·13 (–1·24 to 1·50)	0·80 (–2·79 to 4·39)	0·59 (–0·64 to 1·82)	–0·06 (–1·94 to 1·82)
	p value	0·85	0·66	0·35	0·95

Data were obtained using the Family Adaptability and Cohesion Evaluation Scale (FACES-IV) and the Conflict Tactics Scale (CTS). MAU=management as usual. MST=Multisystemic Therapy.

Table 6: Secondary outcomes: A. Young people's self-report of their wellbeing and behaviour. B. Parents' and teachers' report of young people's wellbeing and behaviour and parents' own wellbeing. C. Clinician-rated mental health outcomes using multiple imputation with baseline educational outcomes and demographic covariates.

A.

	Group and between-group significance	Total SDQ score Mean (SD) [n]	SDQ impact score Mean (SD) [n]	SDQ emotional problems score Mean (SD) [n]	SDQ hyperactivity/ inattention Mean (SD) [n]	SDQ prosocial behaviour Mean (SD) [n]	MFQ Mean (SD) [n]
6 months–baseline	MST	17·4 (5·7) [n=340]	2·5 (2·8) [n=340]	3·4 (2·6) [n=340]	6·5 (2·5) [n=340]	6·8 (2·3) [n=340]	8·7 (6·4) [n=341]
	MAU	17·2 (6·3) [n=340]	2·6 (2·9) [n=340]	3·5 (2·6) [n=340]	6·4 (2·6) [n=340]	6·7 (2·1) [n=340]	8·7 (6·4) [n=339]
6-month follow-up	MST	16·1 (5·7) [n=290]	1·9 (2·6) [n=290]	3·2 (2·4) [n=290]	6·0 (2·3) [n=290]	6·6 (2·1) [n=290]	6·8 (5·7) [n=292]
	MAU	16·4 (6·1) [n=264]	2·0 (2·4) [n=264]	3·5 (2·5) [n=264]	6·0 (2·4) [n=264]	6·6 (2·1) [n=264]	7·9 (6·6) [n=268]
	Difference (95% CI)	–0·33 (–1·17 to 0·51)	–0·03 (–0·42 to 0·36)	–0·28 (–0·63 to 0·07)	–0·09 (–0·46 to 0·28)	0·05 (–0·28 to 0·38)	–1·05 (–1·93 to –0·17)
	p value	0·45	0·87	0·11	0·62	0·78	0·018
12-month follow-up	MST	14·9 (5·7) [n=252]	1·4 (2·2) [n=252]	3·0 (2·3) [n=252]	5·7 (2·6) [n=252]	6·8 (2·3) [n=252]	5·9 (5·4) [n=248]
	MAU	16·0 (6·2) [n=237]	1·8 (2·3) [n=237]	3·5 (2·5) [n=237]	5·8 (2·4) [n=237]	6·7 (2·0) [n=237]	7·0 (5·7) [n=238]
	Difference (95% CI)	–1·09 (–1·97 to –0·21)	–0·32 (–0·73 to 0·09)	–0·43 (–0·80 to –0·06)	–0·11 (–0·50 to 0·28)	0·05 (–0·30 to 0·40)	–1·07 (–1·99 to –0·15)
	p value	0·016	0·13	0·020	0·58	0·77	0·022
18-month follow-up	MST	14·9 (6·0) [n=221]	1·4 (2·2) [n=221]	3·3 (2·5) [n=221]	5·4 (2·6) [n=221]	6·8 (2·2) [n=221]	6·4 (6·2) [n=234]
	MAU	15·3 (6·2) [n=193]	1·6 (2·5) [n=193]	3·5 (2·7) [n=193]	5·5 (2·6) [n=193]	6·9 (2·1) [n=193]	6·8 (6·2) [n=217]
	Difference (95% CI)	–0·55 (–1·51 to 0·41)	–0·13 (–0·58 to 0·32)	–0·25 (–0·64 to 0·14)	–0·15 (–0·56 to 0·26)	–0·06 (–0·43 to 0·31)	–0·29 (–1·25 to 0·67)
	p value	0·26	0·58	0·20	0·46	0·77	0·55

Data were obtained using the Strengths and Difficulties Questionnaire (SDQ) and Moods and Feelings Questionnaire (MFQ). MAU=management as usual. MST= Multisystemic Therapy.

B.

	Group and between-group significance (*t*-test)	Total SDQ score (P) Mean (SD) [n]	SDQ impact score (P) Mean (SD) [n]	SDQ emotional problems score (P) Mean (SD) [n]	SDQ hyperactivity/ inattention (P) Mean (SD) [n]	SDQ prosocial behaviour (P) Mean (SD) [n]	Conners ADHD (P) Mean (SD) [n]	Conners ADHD (T) Mean (SD) [n]	GHQ (P) Mean (SD) [n]
6 months–baseline	MST	21·6 (6·2) [n=340]	5·30 (2·73) [n=340]	4·21 (2·75) [n=340]	7·60 (2·38) [n=340]	5·25 (2·51) [n=340]	80·2 (12·3) [n=341]	74.2 (12.9) [n=213]	64·1 (16·5) [n=341]
	MAU	21·6 (6·5) [n=340]	5·29 (2·95) [n=340]	4·22 (2·64) [n=340]	7·56 (2·53) [n=340]	5·38 (2·50) [n=340]	79·0 (13·2) [n=339]	73.7 (12.8) [n=217]	62·3 (18·3) [n=339]
6-month follow-up	MST	17·5 (6·7) [n=290]	3·5 (3·0) [n=290]	3·3 (2·6) [n=290]	6·3 (2·5) [n=290]	5·9 (2·4) [n=290]	72·8 (14·5) [n=292]	69·3 (16·2) [n=150]	52·5 (15·5) [n=292]
	MAU	19·2 (7·1) [n=268]	4·0 (3·0) [n=268]	3·8 (2·7) [n=268]	6·7 (2·6) [n=268]	5·5 (2·5) [n=268]	76·5 (14·9) [n=268]	69·1 (16·6) [n=155]	59·8 (18·7) [n=268]
	Difference (95% CI)	–2·00 (–3·02 to –0·98)	–0·63 (–1·10 to –0·16)	–0·62 (–0·99 to –0·25)	–0·43 (–0·82 to –0·04)	0·56 (0·21 to 0·91)	–5·16 (–7·45 to –2·87)	0·27 (–1·63 to 2·17)	–6·89 (–9·38 to –4·40)
	p value	0·00011	0·0089	0·0013	0·029	0·0025	<0·0001	0·78	<0·0001
12-month follow-up	MST	16·6 (7·3) [n=246]	3·4 (3·1) [n=246]	3·1 (2·6) [n=246]	6·0 (2·9) [n=246]	5·8 (2·5) [n=246]	71·7 (15·6) [n=248]	67·5 (17·2) [n=134]	53·9 (16·6) [n=248]
	MAU	18·0 (7·3) [n=237]	3·6 (3·0) [n=237]	3·8 (2·7) [n=237]	6·4 (2·8) [n=237]	6·3 (2·4) [n=237]	72·8 (15·5) [n=238]	68·4 (16·5) [n=123]	57·5 (18·1) [n=238]
	Difference (95% CI)	–1·24 (–2·32 to –0·16)	–0·31 (–0·82 to 0·20)	–0·55 (–0·94 to –0·16)	–0·42 (–0·83 to –0·01)	–0·33 (–0·70 to 0·04)	–1·53 (–3·94 to 0·88)	–0·64 (–2·74 to 1·46)	–3·58 (–6·23 to –0·93)
	p value	0·023	0·23	0·0066	0·045	0·086	0·21	0·55	0·0079
18-month follow-up	MST	16·5 (6·9) [n=232]	3·3 (3·0) [n=232]	3·1 (2·5) [n=232]	6·0 (2·6) [n=232]	5·9 (2·5) [n=232]	69·5 (16·8) [n=234]	68·6 (17·0) [n=87]	52·7 (15·7) [n=234]
	MAU	16·6 (7·4) [n=209]	3·4 (3·2) [n=209]	3·4 (2·7) [n=209]	5·9 (2·8) [n=209]	6·3 (2·5) [n=209]	70·0 (16·6) [n=217]	68·7 (16·7) [n=90]	56·2 (18·5) [n=217]
	Difference (95% CI)	–0·43 (–1·55 to 0·69)	–0·19 (–0·72 to 0·34)	–0·40 (–0·81 to 0·01)	0·00 (–0·41 to 0·41)	–0·33 (–0·72 to 0·06)	–1·01 (–3·50 to 1·48)	–0·05 (–1·95 to 1·85)	–2·84 (–5·54 to –0·14)
	p value	0·44	0·48	0·058	0·99	0·11	0·43	0·96	0·040

Data were obtained using the Strengths and Difficulties Questionnaire (SDQ), Conners ADHD Rating Scale – Parent and Teacher form (Conners Comprehensive Behaviour Rating Scale), and General Health Questionnaire (GHQ). MAU=management as usual. MST=Multisystemic Therapy. P=completed by parent. T=completed by teacher. Because of the high proportion of incomplete and missing data for the Conners measure, only the results of multiple imputation are shown (see appendix for description of the imputation procedure followed).

C.

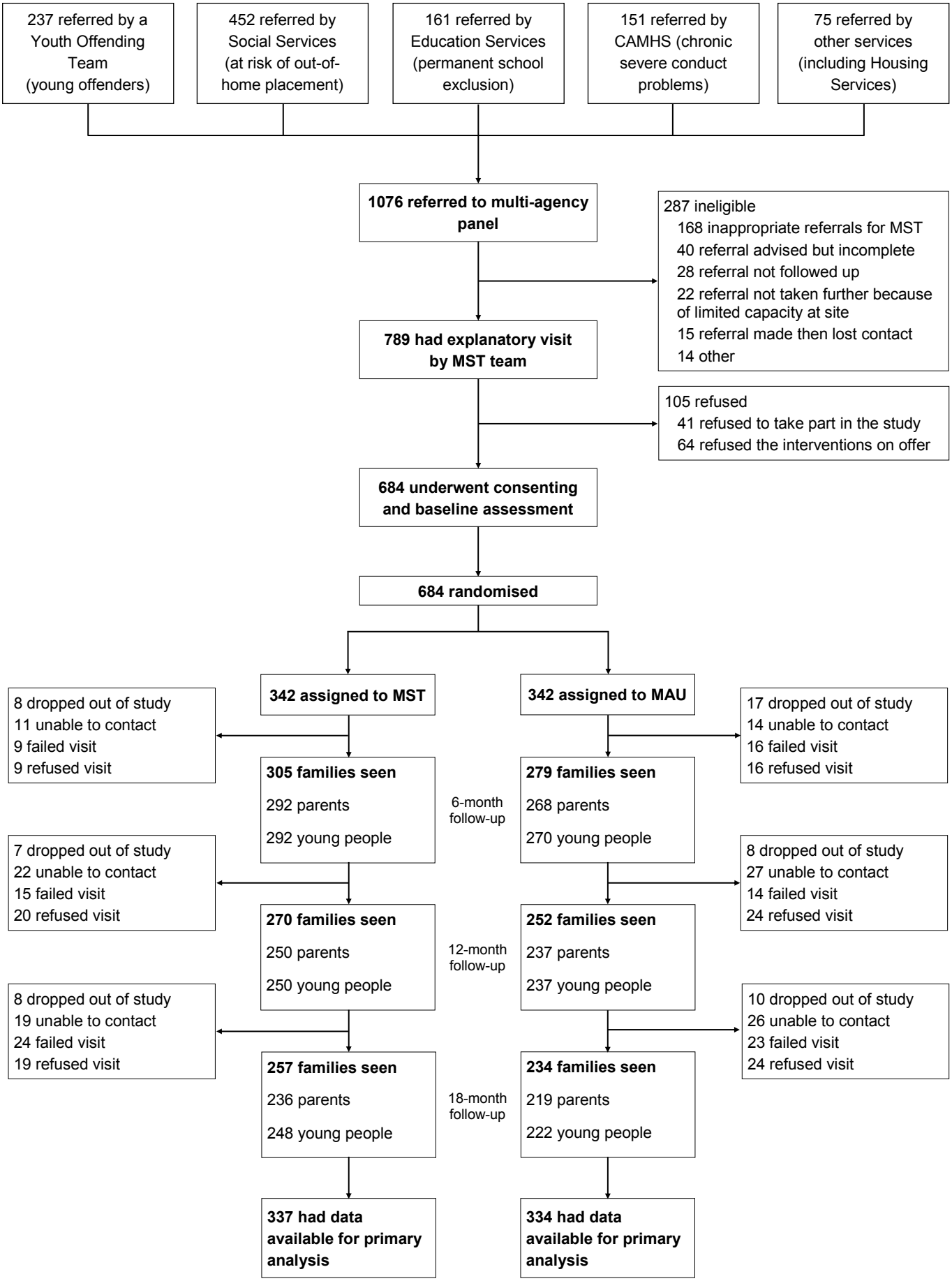
	Group (n) and between-group significance	Any disorder (proportion)	CD (proportion)	ADHD (proportion)	Major depression (proportion)	Anxiety disorder (proportion)	CD with anxiety (proportion)
Baseline	MST (n=342)	0·843	0·778	0·389	0·089	0·146	0·798
	MAU (n=342)	0·853	0·794	0·302	0·124	0·182	0·824
	p value	0·797	0·677	0·021	0·174	0·243	0·441
12-month follow-up	MST (n=249)	0·596	0·456	0·304	0·064	0·132	0·497
	MAU (n=238)	0·616	0·484	0·296	0·070	0·161	0·554
	Difference (95% CI)	0·75 (0·53 to 1·06)	0·90 (0·62 to 1·30)	0·71 (0·46 to 1·10)	1·21 (0·57 to 2·55)	0·83 (0·49 to 1·40)	0·81 (0·56 to 1·18)
	p value	0·12	0·54	0·12	0·62	0·49	0·25

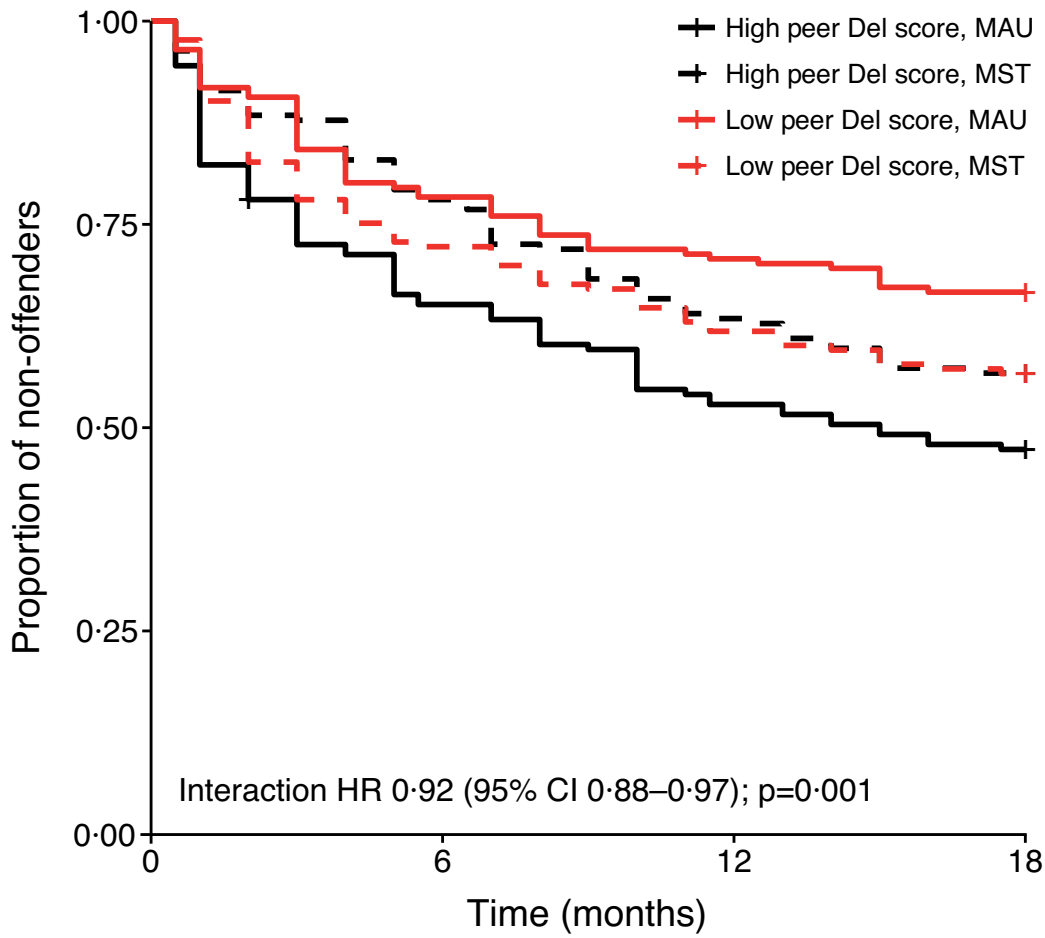
Data were obtained using the Development and Well-Being Assessment. ADHD=attention deficit hyperactivity disorder. CD=conduct disorder. MAU=management as usual. MST=Multisystemic Therapy. Because of the high proportion of incomplete and missing data for the DAWBA, only the results of multiple imputation are shown (see appendix for description of the imputation procedure followed).

Table 7: Differences in costs (£ at 2012/2013 prices) and outcomes per participant over the 18-month follow-up period

	MST (n=226)	MAU (n=209)			
	Mean (SD)	Mean (SD)	Mean difference*	95% CI*	p value*
Costs					
MST intervention	2116·17 (1793·29)	0 (0)	2115·32	1876·78 to 2353·85	<0·0001
Accommodation	819·55 (4446·35)	1433·41 (7977·41)	–599·67	–1809·46 to 610·12	0·33
Education	7869·97 (11378·52)	6602·44 (9913·74)	1424·77	–377·98 to 3227·51	0·12
Secondary health care	500·28 (1773·41)	798·26 (3920·04)	–210·68	–735· 90 to 314·55	0·43
Community services	4127·71 (13338·89)	4674·40 (9991·88)	–617·11	–2780·61 to 1546·38	0·58
Medication	8·34 (111·13)	1·39 (4·02)	–0·47	–1·11 to 0·16	0·15
Criminal justice	13245·30 (23072·32)	17417·79 (29244·66)	–3341·22	–8140·65 to 1458·22	0·17
Total	28678·32 (34175·21)	30927·68 (36106·37)	–1622·94	–7684·45 to 4438·57	0·60
Outcomes					
Out-of-home placement	9·73	8·17	1·56		

MAU=management as usual. MST=Multisystemic Therapy. *Adjusted for stratification variables.


237 referred by a Youth Offending Team (young offenders)
452 referred by Social Services (at risk of out-of-home placement)
161 referred by Education Services (permanent school exclusion)
151 referred by CAMHS (chronic severe conduct problems)
75 referred by other services (including Housing Services)
1076 referred to multi-agency panel
287 ineligible
168 inappropriate referrals for MST
40 referral advised but incomplete
28 referral not followed up
22 referral not taken further because of limited capacity at site
15 referral made then lost contact
14 other
789 had explanatory visit by MST team
105 refused
41 refused to take part in the study
64 refused the interventions on offer
684 underwent consenting and baseline assessment
684 randomised
342 assigned to MST
342 assigned to MAU
8 dropped out of study
11 unable to contact
9 failed visit
9 refused visit
17 dropped out of study
14 unable to contact
16 failed visit
16 refused visit
305 families seen
292 parents
292 young people
6-month follow-up
279 families seen
268 parents
270 young people
7 dropped out of study
22 unable to contact
15 failed visit
20 refused visit
8 dropped out of study
27 unable to contact
14 failed visit
24 refused visit
270 families seen
250 parents
250 young people
12-month follow-up
252 families seen
237 parents
237 young people
8 dropped out of study
19 unable to contact
24 failed visit
19 refused visit
10 dropped out of study
26 unable to contact
23 failed visit
24 refused visit
257 families seen
236 parents
248 young people
18-month follow-up
234 families seen
219 parents
222 young people
337 had data available for primary analysis
334 had data available for primary analysis



Number at risk

High peer Del score, MAU	164	106	86	77
High peer Del score, MST	164	128	104	93
Low peer Del score, MAU	171	134	121	114
Low peer Del score, MST	173	125	107	98

Appendices: Supplementary materials

Multisystemic Therapy versus management as usual in the treatment of adolescent antisocial behaviour (START): a randomised controlled trial

Peter Fonagy, Stephen Butler, David Cottrell, Stephen Scott, Stephen Pilling, Ivan Eisler, Peter Fuggle, Abdullah Kraam, Sarah Byford, James Wason, Rachel Ellison, Elizabeth Simes, Poushali Ganguli, Elizabeth Allison, Ian M Goodyer

Appendix i: Study Design and Methods

A comprehensive listing of inclusion criteria by referral source is provided in Table A1.

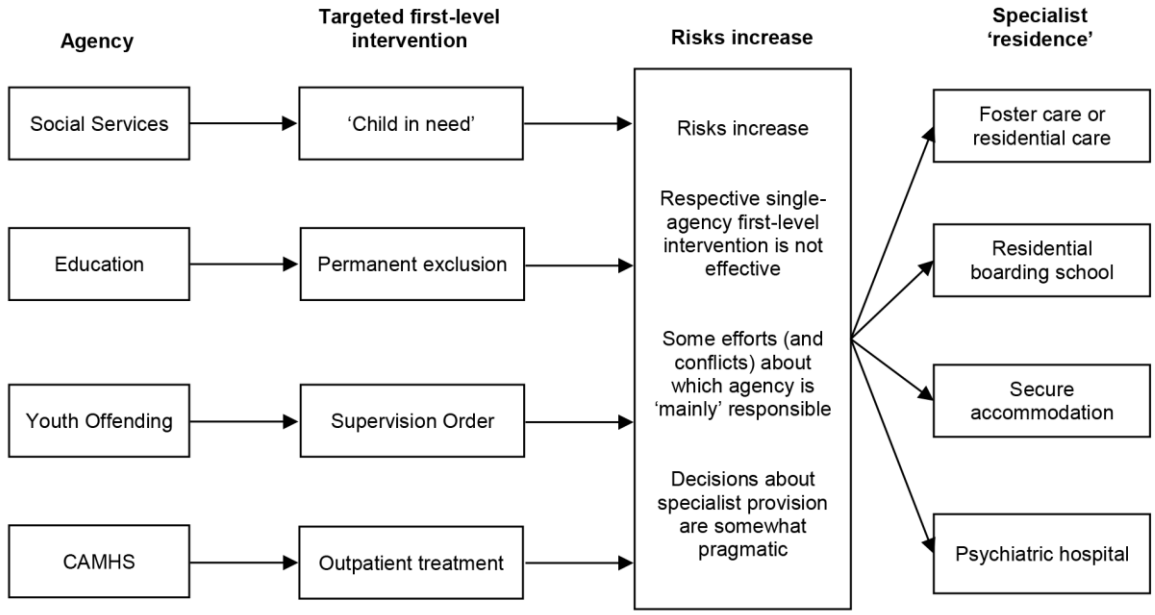
Table A1: Inclusion criteria by referral source, additional severity criteria, and exclusion criteria in the START trial of multisystemic therapy

Referral source	Operationalised inclusion criteria specific to the referral source*
Social services	• Designated as 'child in need' where this is associated with antisocial behaviour on the part of the adolescent • Exhibiting extremely challenging behaviour by either persistent (weekly) and enduring (6 months or longer) violent and aggressive interpersonal behaviour and/or a significant risk of harm to self or to others (for example, self-harming, substance misuse, sexual exploitation, absconding)
Youth Offending Teams	• At least one conviction within the past 12 months, or referral via a supervision order with multisystemic therapy as a specified activity • A warning, reprimand, and/or conviction on at least three occasions in the past 18 months
Child and Adolescent Mental Health Services	• Current diagnosis of conduct disorder, substance misuse, major depression, or anxiety • History of at least one unsuccessful outpatient intervention • Either history of school exclusion or assessment as 'child in need'
Education services	• Currently permanently excluded from school • History of having been excluded from at least one other school for aggressive conduct
Additional severity criteria	**At least three of the following indicators of risk status:** • Excluded from school or at significant risk of exclusion; • High levels of non-attendance at school • A history of offending, or at significant risk of offending; • Previous episodes on the Child Protection Register • Previous episodes of being 'looked after', that is, placed outside of the home (whether via incarceration, psychiatric hospitalisation, residential schooling or assignment to residential local authority care) • Previous referral to a Family Group Conference (usually a meeting between the family members and sometimes also friends or neighbours, the young person and his/her supporter or advocate if requested, and professionals from the health, education, or social services to discuss, plan and make decisions regarding a child at risk to prevent the young person from becoming looked after) • History of siblings being looked after and taken into local authority care
Exclusion criteria	• History or current diagnosis of psychosis • Generalised learning problems (clinical diagnosis) as indicated by intelligence quotient (IQ) below 65 • Identified serious risk of injury or harm to a therapist or researcher • Presenting issues for which MST has not been empirically validated, in particular, substance abuse in the absence of criminal conduct or sex offending as the sole presenting issue • High suicidality • Committed offences likely to bring a custodial sentence • Insufficient family involvement for MST to be applied

MST=Multisystemic Therapy. *All participants must also meet the general inclusion criteria described in the main text.

Figure A1 gives an idealised schematic of the prototypical care pathway an individual young person with moderate or severe antisocial behaviour might follow, depending on the agency of first contact.

Figure A1: Schematic care pathway for an antisocial young person

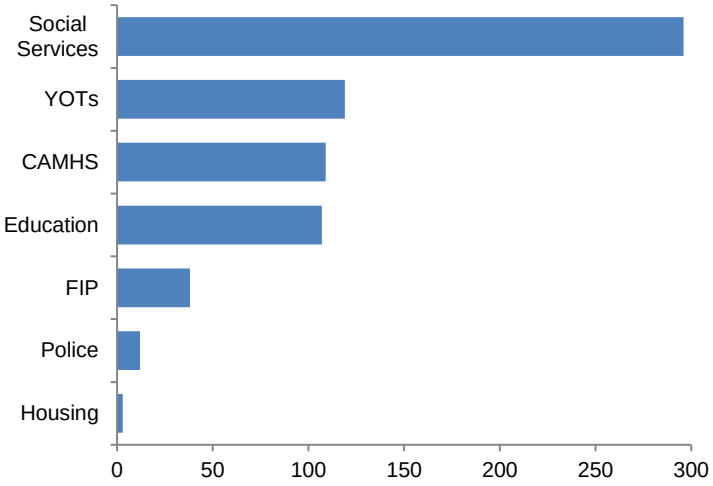


CAMHS=Child and Adolescent Mental Health Services.

Figure A2 and the accompanying table display the agencies that provided participants for the START trial. Social Care was the primary source of referrals, with Youth Offending Teams (YOTs), Child and Adolescent Mental Health Services (CAMHS) and Education each providing about the same number.

Statistical tests revealed no major demographic or clinical differences between the subgroups referred by the five major categories of providers (Police and Housing provided too few cases for the differences to be examined statistically).

Figure A2: Referral sources for the START trial, including only randomised cases



Source	n	%
Social Services	296	43·3%
YOTs	119	17·4%
CAMHS	109	15·9%
Education	107	15·6%
Police	12	1·8%
FIP	38	5·6%
Housing	3	0·4%
Total	**684**	**100%**

CAMHS=Child and Adolescent Mental Health Services. FIP=Family Intervention Project. YOTs=Youth Offending Teams.

Table A2 shows the number of participants who met the severity of antisocial behaviour criteria in MST and MAU, respectively. The severity criteria incorporate objective data (ie, offending, school exclusions), young people's reports of their antisocial behaviour, and diagnoses of conduct disorder from a semi-structured psychiatric interview (the Development and Well-Being Assessment; DAWBA). Table A3 documents the prevalence of current and historical self-harm based on data gathered from the DAWBA.

Table A2: Number of participants meeting each of the severity criteria

Severity criteria	n	%	MST n	MST %	MAU n	MAU %
Number of participants with persistent (weekly) and enduring (>6 months) violent and aggressive interpersonal behaviour (endorsing two or more items violent and/or aggressive behaviour)	443	64·8	220	64·3	223	65·2
Number of participants with at least one conviction plus three additional warnings, reprimands, or convictions (44 with 4 convictions, 10 with 3 convictions, 6 with 2 convictions)	63	9·2	27	7·9	36	10·5
Number of participants with a current DSM-IV diagnosis of conduct disorder at baseline	531	77·6	262	76·6	269	78·7
Number of participants with a permanent school exclusion for antisocial behaviour at baseline	179	26·2	93	27·2	86	25·1

Data were obtained from the Self-report Delinquency Measure, the Police National Computer database, the Development and Well-Being Assessment and from school exclusion records. MAU=management as usual. MST=Multisystemic Therapy.

Table A3: Prevalence of reported self-harm

Prevalence of reported self-harm	n	%	MST n	MST %	MAU n	MAU %
Rates of self-harm at baseline (n=683)						
Recent discussion of self-harm	88	12·9	44	12·9	44	12·9
Report of recent deliberate self-harm	67	9·8	31	9·1	36	10·5
Report ever self-harmed	197	28·8	85	24·9	112	32·7
Rates of self-harm at follow up (n=510)						
Recent discussion of self-harm	33	6·5	14	4·1	19	5·6
Report of recent deliberate self-harm	23	4·5	7	2·0	16	4·7
Report ever self-harmed	146	28·6	67	19·6	79	23·1

Data were obtained from the Development and Well-Being Assessment. MAU=management as usual. MST=Multisystemic Therapy

Table A4 shows the nine sites where recruitment to the trial took place. At all sites the first patients were recruited within 30 days of the site becoming active. Because one site was closed after the recruitment numbers were specified, five sites recruited more than the contracted 70 participants in order to achieve close to the target of 700 participants on which power calculations were based. As sites were included in the randomisation algorithm there were no significant deviations from the 50% split between management as usual (MAU) and MST allocations. The table also lists the mean adherence rating (Therapist Adherence Measure-Revised score) of each site (see below for a description of how ratings were obtained).

Table A4: Recruitment and therapist adherence scores at the nine trial sites

Site	Became active	Date first family recruited	Recruitment (n)	MST (%):MAU (%)	TAM-R score	
					Mean	SE
Barnsley	June 2010	June 28, 2010	80	38 (49):41 (51)	0·698	0·035
Greenwich	February 2010	February 4, 2010	80	38 (48):42 (52)	0·790	0·035
Hackney	February 2010	March 16, 2010	70	35 (50):35 (50)	0·640	0·035
Leeds	February 2010	March 8, 2010	83	44 (53):39 (47)	0·733	0·033
Merton & Kingston	July 2010	July 29, 2010	80	41 (51):39 (49)	0·610	0·033
Peterborough	February 2010	March 4, 2010	81	41 (51):40 (49)	0·615	0·034
Reading	September 2010	October 11, 2010	70	36 (51):34 (49)	0·704	0·036
Sheffield	December 2010	January 20, 2011	70	35 (50):35 (50)	0·705	0·039
Trafford	December 2010	January 13, 2011	70	33 (47):37 (53)	0·806	0·038
Total			**684**	**342 (50):342 (50)**	**0·698**	**0·012**

MAU=management as usual. MST=Multisystemic Therapy. TAM-R=Therapist Adherence Measure-Revised.

Details of the planned interventions

Multisystemic Therapy
Multisystemic Therapy (MST) is an integrative, manualised, licensed programme with a substantial evidence base for engaging young people exhibiting antisocial behaviour and their families. Although the intervention is manualised, it is also individualised, highly flexible, and adaptable to various constellations of needs. Young people with severe conduct problems (violence, substance misuse, school expulsion) were treated over a period of 3 to 6 months with a community-based multicomponent treatment programme focused on the family but also engaging schools, neighbourhoods, and community resources. The programme was administered by specifically trained professionals (MST workers) with relatively low caseloads of 4 to 6 cases. The average treatment duration was 139 days. Young people and families requiring this approach are assumed to respond poorly to engagement by existing services (see inclusion criteria in Table A1). The cases referred tend to require intensive outreach services, probably associated with complex family problems, including substance misuse and mental health problems, which are likely to affect parenting. The frequency of contact with the MST workers is monitored but not controlled. MST addresses specific individual risk factors in line with the Risk–Need–Responsivity model[1] specifically designed for hard-to-reach troubled families. This includes a duty cover system available 24 hours a day, 7 days a week. In recognition of this commitment, an individual therapist normally works with no more than four to six families at a time. The treatment uses multiple interventions, in combinations indicated by the clinical picture. The constituent treatments include techniques from systemic and structural family therapy, parent training, marital therapy, supportive therapy related to interpersonal problems, social skills components, social perspective training, behavioural methods (eg, contingency contracting) and cognitive therapy techniques (eg, self-instructional training), as well as case management with the therapist acting as an advocate to outside agencies.

A family focus is central to the intervention. The overriding goals of MST are to give parents the skills and resources needed to address the inevitable difficulties of raising adolescents, and to empower the young people to cope with familial and extra-familial problems. Assessment and treatment explore the young person's role in various systems and consider the inter-relationship between these systems. Specific attention is given to strengthening the various systems, and an attempt is made to promote appropriate and responsible behaviour among all family members. The therapist aims to develop the family's skills and resources and to address communication problems and other challenges with social, educational, and youth justice services. MST is more than a mere amalgamation of techniques and approaches, and the focus on the interrelationship between systems is retained. Interventions are individualised and highly flexible but are documented in treatment manuals.[2]

Each MST site was led by an accredited supervisor with experience of delivering MST, including experience and resources to offer group and one-to-one supervision of therapists. MST was delivered by a team of at least three specially trained clinicians under the supervision of an MST supervisor, with weekly 1-hour conference calls for consultation with an MST Services staff member. In addition, MST therapists had the support of local consultation from mental health professionals with postgraduate qualifications in disciplines such as social work, psychology, or counselling. In view of the breadth and complexity of this input, it was essential to monitor consultation as well as contact time of the MST team in order to arrive at accurate assessments of health and social care costs. We endeavoured to ensure that the MST therapists and MST supervisors would not be allowed to see participants in the management as usual (MAU) arm of the trial.

The nine trial sites were all licensed by MST Services and the quality of treatment they provided was closely and continuously monitored. There was a weekly telephone consultation between the therapists and an MST expert designated by MST Services, and booster training sessions were provided four times a year. There were twice-yearly implementation reviews. Adherence was monitored in relation to each treatment using the Therapist Adherence Measure-Revised (TAM-R), a 28-item instrument based on parent interviews in which they are asked about the intervention they actually received. The TAM-R was administered independently from the MST team by a research assistant (RA) not associated with that site.[3] A minimum score of 0·61 on the TAM-R is specified for the treatment to be classed as adherent. The average rating was 0·698 (SE 0·012), with all but three of the sites averaging statistically significantly above criterion adherence (see Table A4).

Management as usual
MAU was the standard care offered to young people and their families who met eligibility criteria for the trial. This treatment was diverse and often involved no therapeutic intervention or individual or family-orientated work. It was likely to be delivered by a wide range of appropriately qualified practitioners with quite different theoretical orientations and professional groups, including social workers, probation officers, and specialist therapists. Recommended interventions included individual support to re-engage the young person with

education, treatment of substance misuse, anger management, social problem-solving skills training, family-based interventions, and awareness programmes (including victim awareness and reparation interventions). The average duration of these interventions varied considerably. It was expected that practitioners were working in line with best practice as specified in relevant Social Care Institute for Excellence (SCIE) and National Institute for Health and Care Excellence (NICE) guidance. It is unlikely that practitioners in MAU received the extent or quality of supervision available for the MST therapists.

It was not intended even during the trial period for the MAU interventions to be less intensive or less costly than MST. However, they were likely to be delivered in a less focused and far less well specified manner and thus to be less effective. MAU interventions were carefully monitored using the Child and Adolescent Service Use Schedule (CA-SUS; described below) designed specifically for the trial, which recorded contact with all services (health, social, YOT, education, voluntary sector, etc.), including number of contacts and, where possible, average duration of contacts. This gave a realistic sense of the level of intensity of MAU that was available in conjunction with and also independent of the MST arm, to give an indication of shifts in intensity of service provision—that is, whether the addition of MST reduced the need for other (particular kinds of) support. As this was a pragmatic trial involving a number of collaborating services even within each of nine sites, it was never possible to specify in advance what MAU would consist of.

Routine service use in both arms of the trial

Routine interventions offered in both arms of the trial were monitored using a service use schedule, supplemented by a rigorous and exhaustive independent simultaneous search of service records for health, social care, YOT, and school teams associated with any of our trial cases.

The coding was independently carried out by two RAs, with inter-coder agreements in all cases being >80%. We were surprised by the consistency of provision across participants attained for the MAU arm as well as the MST arm; perhaps the systematic delivery of MAU was a side effect of the rigour of the multi-agency panels that reviewed cases, generating greater rigour and integrated delivery of MAU.

Table A5a–d displays the routine care reported by participants and obtained from social care, health care, and YOT records across the follow-up period. The mean number of contacts, their average duration, and the number and percentage of young people making use of the type of care were obtained. The data show that the overall routine care effort spent increased with time over the study period but did not differ for the two treatment arms. At 6 months the young people in the MST group had fewer social worker contacts but no overall difference in either social care or total routine care use. At 12 months the MST group had slightly briefer contacts across the services in all three categories ($t(484)=2.03$; $p=0.0429$). By 18 months there were no differences in routine service use. The introduction of MST appeared to lead to neither an increase nor a decrease or a change in the pattern of service provision. As it is part of the task of the MST worker to ensure that barriers to access to routine services are removed, an increase in the initial intensity of contacts might be anticipated, but this was not achieved in the MST arm. Nor was there evidence of an immediate decrease in use of routine care, as might be anticipated in a study design where MST was followed by routine care provision after 3 to 6 months of treatment.

Table A5a: Routine care received by the two intervention groups at baseline

Baseline	MAU (n=284)			MST (n=291)		
	Mean (SD) number of contacts	Mean (SD) duration (hours)	Number (%) used	Mean (SD) number of contacts	Mean (SD) duration (hours)	Number (%) used
Care Coordinator	0·15 (1·8)	6·89 (77·9)	4 (1·4%)	1·1 (8·1)	46·63 (372·4)	9 (3%)
Psychiatrist	0·21 (0·9)	10·49 (50·6)	18 (6·3%)	0·13 (0·8)	7·15 (44·1)	12 (4·1%)
Clinical Psychologist	0·34 (2·1)	19·26 (127·5)	16 (5·6%)	0·48 (4)	30·73 (248·6)	13 (4·4%)
CAMHS worker	0·67 (2·5)	36·53 (145·8)	45 (15·8%)	0·68 (3·4)	39·84 (208·1)	43 (14·7%)
Community Psychiatric Nurse	0·04 (0·3)	2·43 (21·6)	4 (1·4%)	0·07 (1·2)	4·43 (73·8)	2 (0·6%)
Total routine CAMHS	**1·42 (3·9)**	**75·61 (222·9)**	**72 (25·3%)**	**2·47 (9·7)**	**128·79 (495·1)**	**72 (24·7%)**
Social worker	3·07 (7·7)	159·62 (633·1)	100 (35·2%)	3·37 (7·5)	221·74 (687·7)	100 (34·3%)
Family support worker	1·18 (5·8)	58·25 (293·6)	23 (8%)	1·91 (9·2)	113·24 (598·3)	29 (9·9%)
Social services youth worker	0·49 (3·8)	34·46 (382·4)	12 (4·2%)	0·17 (1·1)	9·83 (58·9)	11 (3·7%)
Total routine social care	**4·74 (10·7)**	**252·33 (815·5)**	**122 (42·9%)**	**5·45 (12·4)**	**344·77 (949·7)**	**123 (42·2%)**
Total routine YOT	**6·12 (14·2)**	**290·6 (715·9)**	**87 (30·6%)**	**5·17 (11·8)**	**321·96 (1644·6)**	**82 (28·1%)**
Total	**12·28 (18)**	**618·55 (1136)**	**199 (70%)**	**13·09 (19·9)**	**795·53 (2032·2)**	**190 (65·2%)**

CAMHS=Child and Adolescent Mental Health Services. MAU=management as usual. MST=Multisystemic Therapy. YOT=Youth Offending Team.
*Indicates significant differences between the trial conditions on *t*-test or χ^2 test.

Table A5b: Routine care received by the two intervention groups at 6-month follow-up

6-month follow-up	MAU (n=266)			MST (n=251)		
	Mean (SD) number of contacts	Mean (SD) duration (hours)	Number (%) using	Mean (SD) number of contacts	Mean (SD) duration (hours)	Number (%) using
Care Coordinator	0·19 (1·7)	21·64 (238·9)	6 (2·2%)	0·94 (7·3)	203·84 (2435·8)	10 (3·9%)
Psychiatrist	0·18 (0·9)	8·12 (42·2)	15 (5·6%)	0·14 (0·7)	9·08 (47·7)	14 (5·5%)
Clinical Psychologist	0·55 (5)	32·01 (300·5)	14 (5·2%)	0·3 (2·7)	16·25 (161·2)	13 (5·1%)
CAMHS worker	0·58 (2·4)	32·63 (146)	32 (12%)	0·69 (2·5)	34·82 (133·2)	30 (11·9%)
Community Psychiatric Nurse	0 (0)	0 (0)	0 (0%)	0·05 (0·4)	3·94 (34·4)	4 (1·5%)
Total routine CAMHS	**1·5 (6)**	**94·41 (417·3)**	**53 (19·9%)**	**2·13 (8·4)**	**267·95 (2443·8)**	**56 (22·3%)**
Social worker	2·59 (6·6)	116·16 (308·2)	86 (32·3%)	2·72 (7·4)	134·39 (413·5)	78 (31%)
Family support worker	1·99 (10·3)	96·72 (437·2)	28 (10·5%)	1·38 (6·3)	93·79 (545·7)	20 (7·9%)
Social services youth worker	1·24 (6·8)	73·41 (386)	16 (6%)	0·31 (2·5)*	21·87 (208·7)	6 (2·3%)*
Total routine social care	**5·82 (16·2)**	**286·28 (711·2)**	**102 (38·3%)**	**4·42 (11·5)**	**250·05 (821·4)**	**91 (36·2%)**
Total routine YOT	**4·47 (10·8)**	**222·07 (613·2)**	**67 (25·1%)**	**4·93 (11·3)**	**240·7 (600·3)**	**70 (27·8%)**
Total	**11·78 (21·8)**	**602·77 (1112)**	**158 (59·3%)**	**11·49 (21·8)**	**758·7 (2719·9)**	**150 (59·7%)**

CAMHS=Child and Adolescent Mental Health Services. MAU=management as usual. MST=Multisystemic Therapy. YOT=Youth Offending Team.
*Indicates significant differences between the trial conditions on *t*-test or χ^2 test.

Table A5c: Routine care received by the two intervention groups at 12-month follow-up

12-month follow-up	MAU (n=245)			MST (n=239)		
	Mean (SD) number of contacts	Mean (SD) duration (hours)	Number (%) using	Mean (SD) number of contacts	Mean (SD) duration (hours)	Number (%) using
Care Coordinator	2·93 (18·4)	487·9 (4047·6)	8 (3·2%)	0·77 (6·8)	33·4 (240·6)	10 (4·1%)
Psychiatrist	0·29 (1·7)	15·51 (93·9)	17 (6·9%)	0·15 (0·7)	7·97 (41·5)	14 (5·8%)
Clinical Psychologist	0·27 (1·5)	15·69 (94·1)	14 (5·7%)	0·11 (0·6)	7·99 (47·9)	10 (4·1%)
CAMHS worker	0·42 (1·8)	21·2 (89·4)	27 (11%)	0·63 (2·5)	27·68 (95·1)	34 (14·2%)
Community Psychiatric Nurse	0·11 (1·7)	6·73 (103·5)	2 (0·8%)	0 (0)	0 (0)	0 (0%)
Total routine CAMHS	**4·02 (19·1)**	**547·04 (4050·1)**	**50 (20·4%)**	**1·66 (7·3)**	**77·03 (270·6)**	**57 (23·8%)**
Social worker	2·95 (7)	134·14 (373·5)	80 (32·6%)	3·18 (7·5)	163·75 (466·9)	77 (32·2%)
Family support worker	1·9 (11·3)	142·4 (1003·7)	20 (8·1%)	1 (4·8)	73·76 (425·8)	23 (9·6%)
Social services youth worker	0·59 (6·6)	42·33 (432·7)	8 (3·2%)	0·33 (3)	18·72 (179·4)	6 (2·5%)
Total routine social care	**5·44 (15·4)**	**318·85 (1179)**	**92 (37·5%)**	**4·52 (9·8)**	**256·24 (679·3)**	**91 (38%)**
Total routine YOT	**5·07 (13·7)**	**228 (587)**	**57 (23·2%)**	**4·59 (14·7)**	**194·18 (554·9)**	**55 (23%)**
Total	**14·53 (28·2)**	**1093·9 (4238·6)**	**138 (56·3%)**	**10·78 (20·6)**	**527·45 (1002·1)***	**136 (56·9%)**

CAMHS=Child and Adolescent Mental Health Services. MAU=management as usual. MST=Multisystemic Therapy. YOT=Youth Offending Team.
*Indicates significant differences between the trial conditions on *t*-test or χ^2 test.

Table A5d: Routine care received by the two intervention groups at 18-month follow-up

18-month follow-up	MAU (n=222)			MST (n=209)		
	Mean (SD) number of contacts	Mean (SD) duration (hours)	Number (%) using	Mean (SD) number of contacts	Mean (SD) duration (hours)	Number (%) using
Care Coordinator	2·89 (17·4)	511·33 (4188·4)	12 (5·4%)	2·76 (18)	292·61 (3038·8)	20 (9·5%)
Psychiatrist	0·59 (2·4)	32·21 (140·8)	31 (13·9%)	0·44 (1·6)	25·53 (89·3)	23 (11%)
Clinical Psychologist	1·28 (7·1)	74 (424·3)	28 (12·6%)	0·79 (5·2)	51·13 (316·7)	19 (9%)
CAMHS worker	1·89 (5)	101·09 (287·3)	63 (28·3%)	2·09 (5·8)	106·17 (317·4)	60 (28·7%)
Community Psychiatric Nurse	0·18 (1·8)	10·54 (111·2)	6 (2·7%)	0·17 (1·9)	10·91 (106·6)	4 (1·9%)
Total routine CAMHS	**6·84 (21·3)**	**729·19 (4250)**	**89 (40%)**	**6·27 (19·9)**	**486·35 (3066·3)**	**89 (42·5%)**
Social worker	7·44 (13·2)	330·03 (642·6)	116 (52·2%)	8·43 (16·7)	463·42 (1097·5)	104 (49·7%)
Family support worker	4·68 (17·5)	288·54 (1225·3)	43 (19·3%)	3·38 (10·6)	226·35 (902·4)	41 (19·6%)
Social services youth worker	1·82 (9·5)	98·09 (536·3)	23 (10·3%)	0·63 (3·1)	32·93 (171·7)	16 (7·6%)
Total routine social care	**13·93 (27)**	**716·67 (1553·1)**	**138 (62·1%)**	**12·43 (22·4)**	**722·71 (1576·8)**	**122 (58·3%)**
Total routine YOT	**14·21 (29·9)**	**640·53 (1415·9)**	**87 (39·1%)**	**12·92 (24)**	**584·38 (1175·1)**	**92 (44%)**
Total	**34·99 (44·6)**	**2086·4 (4750·1)**	**189 (85·1%)**	**31·62 (40·5)**	**1793·44 (3827·4)**	**169 (80·8%)**

CAMHS=Child and Adolescent Mental Health Services. MAU=management as usual. MST=Multisystemic Therapy. YOT=Youth Offending Team.
*Indicates significant differences between the trial conditions on *t*-test or χ^2 test.

Assessments and outcome measures

To maximise the clinical validity of the outcome evaluations, assessments were made across multiple domains using multiple methods and sources. Several UK government departments had a stakeholding interest in the study, and the variety and number of measures reflect the desire to incorporate measures relevant to particular policy concerns (eg, mental health outcomes for the Department of Health, classroom behaviour outcomes for the Department for Children, Schools and Families, criminality outcomes for the Home Office, etc.).

Primary outcome

The primary outcome, specified by the funders (the UK Department for Children, Schools and Families, the Department of Health, and the Home Office) was the proportion of cases assigned to long-term (3 months or longer) *out-of-home placement* in specialist residential provision, including placement into local authority care, incarceration, long-term hospitalisation and residential schooling, at 18 months following randomisation. The investigators expected this trial to give information on how many young people assigned to MAU and MST require specialist residential provision either immediately or during the follow-up period. There were concerns that out-of-home-placement may be a reactive measure of outcome. A situation could be envisaged whereby the presence of the MST team would influence the likelihood of the courts or other systems deciding to place the young person away from the family. It was also possible that the presence of the MST team, affording a more accurate view of family functioning, may precipitate the placement of the young person outside the home. These types of influences suggest that the primary outcome measure may be 'reactive' with the planned intervention, and would compromise randomisation and compromise the trial. In order to minimise these problems we placed the primary endpoint of the study at 18 months in order to see whether the impact of MST would be apparent over the course of the year following the intervention. It was considered unlikely that over this period the primary outcome measure (long-term out-of-home placement) would be reactive with the intervention.

The research team strongly felt that while the rate of out-of-home placement was an important primary outcome, it was not in every instance an indication of the failure of the system to provide adequate support to the young person and his/her family. Findings have to be interpreted in the context of other outcomes, including general wellbeing, which may in some cases improve following out-of-home placement. Placement into specialist residential provision in the researchers' view reflects four types of outcome based on two separate factors—the first about family functioning and the second concerning decisions about where the young person lives. If, following intervention, the family functions in a way that more adequately meets the young person's needs and the young person continues to reside in the family, this constitutes an unequivocally preferred outcome. If, despite intervention, family functioning remains unchanged and is unable to meet young person's needs and the young person is placed out of the family, this constitutes a failure of the intervention (preservation of the family did not succeed), but it is likely to be the best outcome for the young person in the circumstances. The third possible outcome is that, despite intervention, family functioning is still unable to meet the young person's needs but the young person remains in the family. This is the critical instance where an apparently good outcome (family preservation) in fact reflects a non-preferred (poor) outcome for the young person. The fourth outcome, which is perhaps less likely, is that the intervention results in better family functioning after intervention but the young person is still placed out of home. It was hoped that this outcome would be rare, but it could represent an 'effective' intervention with respect to psychological outcomes but not with respect to family preservation. Thus, while out-of-home placement was a critical indicator that was considered relevant by all stakeholders, it could not be considered the sole arbiter of effectiveness. This qualification was made clear to the funders in the tender document submitted, on the basis of which the competitive contract was awarded. The schedule for collecting secondary outcome data is shown in Table A6.

Table A6: Schedule of measures together with internal consistency coefficients of the scales used

Assessment	Timeline (months) Baseline (T1)	6 (T2)	12 (T3)	18 (T4)	Cronbach's α (reliability coefficient)	Mean inter-item correlation*
Eligibility and consent						
Eligibility assessed by MST panel	X					
Consent taken	X					
Randomisation information provided	X					
Parent Questionnaires						
Child and Adolescent Service Use Schedule (CA-SUS)	X	X	X	X		
Family Information Part 1	X					
General Health Questionnaire (GHQ)	X	X	X	X	0·95	0·41
Conners Comprehensive Behaviour Rating Scale – Parent form (CBRS)	X	X	X	X	0·89	0·26
Inventory of Callous and Unemotional Traits (ICUT)	X	X	X	X	0·85	0·20
Strengths and Difficulties Questionnaire (SDQ)	X	X	X	X	0·72	0·06
Development and Well-Being Assessment (DAWBA)	X		X			
Conflict Tactics Scale (CTS2S)	X	X	X	X	0·83	0·20
Alabama Parenting Questionnaire (APQ)	X	X	X	X	0·62	0·07
Family Adaptability and Cohesion Evaluation Scale (FACES-IV)	X	X	X	X	0·73	0·08
Family Information Form Part II	X					
Loeber Caregiver Questionnaire	X	X	X	X	0·76	0·15
Young Person Questionnaires						
Child and Adolescent Service Use Schedule – last two questions (CA-SUS)	X	X	X	X		
Short Mood and Feelings Questionnaire (MFQ)	X	X	X	X	0·89	0·58
Inventory of Callous and Unemotional Traits (ICUT)	X	X	X	X	0·78	0·13
Self-Report Delinquency Measure (SRDM)	X	X	X	X	0·92	0·19
Levels of Expressed Emotion (LEE)	X	X	X	X	0·98	0·08
Wechsler Abbreviated Scale of Intelligence (WASI)	X					
Antisocial Beliefs and Attitude Scale (ABAS)	X	X	X	X	0·93	0·17
Strength and Difficulties Questionnaire (SDQ)	X	X	X	X	0·70	0·08
Development and Well-Being Assessment (DAWBA)	X		X			
Alabama Parenting Questionnaire (APQ)	X	X	X	X	0·61	0·10
Youth Materialism Scale	X	X	X	X	0·84	0·27
EQ-5D	X	X	X	X		
Education Data						
Conners Comprehensive Behaviour Rating Scale – Teacher form (CBRS)	X	X	X	X	0·89	0·26
Attendance/Exclusion rates	X	X	X	X		
Youth Offending Data						
Offending history	X	X	X	X		

*Clark and Watson[4] have recommended a mean inter-item correlation between 0.15 and 0.20 for broad constructs and between 0.40 and 0.50 for more narrow constructs.

In the light of this argument, which was accepted by the Department of Health/Department for Children, Schools and Families, while the study retained out-of-home placement as a primary outcome in deference to the funders, a further key outcome was proposed and agreed by the Trial Steering Committee and Data Monitoring Committee in order to achieve a comprehensive and definitive evaluation of the intervention. *Forensic outcomes* related to the antisocial behaviours that remain a key part of the definition of the target population, most meaningfully assessed in terms of the time to offences being committed, and reconvictions, which were adopted as a key forensic secondary outcomes to complement the primary outcome of out-of-home placement. Criminal behaviour (the number of violent and non-violent crimes leading to convictions) as registered on the Police Database (categorised as per annual statistical reports)[5] that resulted in a pre-court disposal (Reprimand or Final Warning) or a court disposal was used as an indicator of the severity of antisocial behaviour. Objective outcomes were collected from reports of offending behaviour based on police computer records, including details of custodial sentences. These measures were taken at 6-monthly intervals for the 6 months before randomisation, the 6 months covering the intervention period, and then 6-monthly until the 18-month follow-up point. In addition to the number of records of offending behaviour (count data), we also obtained 6-month periods free of any offending behaviour (binary data). Crime records were obtained from the Police National Computer as well as from the Young Offender Information System database at each study site. These records detail information on offences, court appearances, criminal orders, police custody records and arrest rates.

Additional forensic outcome measures that have been used in previous randomised controlled trials of MST include arrests (based on archival data) or survival rates to first arrest (time to arrest), number of arrests, or dichotomously coded arrests (ie, arrested *vs* not arrested). In some studies, seriousness of crime (tariff) for which the individual was arrested was also included. An obvious alternative forensic outcome would be number of arrests where the mean reduction associated with MST in previous studies was significant (SMD= –0.39, 95% CI –0.81 to 0.02, based on seven studies; n=677). Arrest as an outcome measure is known to be confounded by the efficiency of police forces and to some extent policing policy, both of which can vary considerably across between sites in a national sample. Given that the study covered a range of policing regions, arrests and other measures confounded by local practices were considered unsuitable as outcome measures. Eighteen months was selected as the time for primary outcome measurement to enable identification of any changes subsequent to cessation of therapy. This length of follow-up also facilitated the collection of more meaningful forensic data.

Secondary outcomes

While the number of secondary outcomes may appear large, it was actually reduced relative to initial plans in order to reduce the measurement burden of the study. We found extensive measurement to be a disincentive to continued participation in a similar study.[6] Data on MST contacts were collected directly from the MST sites to avoid participants revealing their group allocation to the researchers. Data on the use of all other services were collected at each time point via an interview using the Child and Adolescent Service Use Schedule (CA-SUS), which was developed and successfully employed by the research team in previous evaluations with young people with complex mental health and social care needs[7-10] but was considerably modified for the present investigation. Monitoring participants' receipt of a range of usual services and documenting outcomes in relation to this in both arms of the trial also enabled us to obtain data on the transition from child to adult services for this population. Data concerning the nature of service provision 'normally' extended to this group are currently unavailable, particularly in relation to the transition years. The RA administered pre-testing questionnaires during the initial contact with the young person and family after they had given consent to participate in the trial, prior to group assignment. Post-testing by the RA was scheduled for 6 months after entry into the study; at the time of planning the study it was envisaged that this would be a minimum of 2 weeks after the family completed the intervention. Follow-up assessments were made at 12 and 18 months post-randomisation.

Self-report of antisocial behaviour. The prevalence and incidence of delinquent behaviour such as vandalism, theft and burglary was monitored using the Self-Report Delinquency measure.[11] Noncompliance and increasingly serious forms of antisocial behaviour, together with young people's perceptions of law-abiding behaviour and institutions, were measured using the Antisocial Beliefs and Attitudes Scale.[12] Peer delinquency was assessed using the Self-Report Delinquency measure.[11] It was predicted that MST would achieve decreases in associations with antisocial peers, increases in positive peer relations, and greater commitment to prosocial activities (eg, education). This prediction was consistent with the model and hypothesised mediating mechanisms[3] and relevant to current social policy initiatives and concerns.

Parenting skills and family functioning. The study was also designed to collect data on variables relating to key mechanisms of change identified in previous studies of MST (parent–adolescent relationships) and to evaluate parenting skills in detail, given that MST aims to improve young people's lives by targeting their

parents/caregivers as being primarily responsible for facilitating change. Adolescent symptoms have been shown to decrease in association with increased supportiveness and decreased conflict between parents[13,14] and with increased follow-through by caregivers on discipline practices.[15] Furthermore, adherence to the MST manual by therapists appears to improve family functioning, which in turn decreases deviant peer affiliations and consequently delinquent behaviour.[3] The quality of the parent–adolescent relationship, family functioning, and parenting practices were evaluated using the Family Adaptability and Cohesion Evaluation Scales (FACES-IV)[16] and the monitoring and supervision subscale from the Alabama Parenting Questionnaire-Short Form (APQ).[17] Outcomes from other APQ subscales are reported in this appendix. Parental disruption was assessed using the short form of the Conflict Tactics Scale (CTS),[18] and the level of expressed emotion in the home (as conceptualised in the Camberwell Family Interview) was assessed using the Level of Expressed Emotion questionnaire.[19,20]

Wellbeing and adjustment. A general assessment of wellbeing used the Strengths and Difficulties Questionnaire (SDQ)[21], a self-report measure completed by both the young people and their parents/caregivers. Depression was specifically monitored using the Short Mood and Feelings Questionnaire (MFQ)[22] completed by the young people. A brief assessment of parental mental health was obtained using the General Health Questionnaire-28 (GHQ),[23] a commonly used instrument for the identification of mental health problems.

Psychiatric screening. Psychiatric disorders were identified and a psychosis screen provided by the Development and Well-Being Assessment (DAWBA).[24] This computerised structured interview measure was administered to both the young person and parents at baseline and 12 months.

Child psychometrics. IQ estimates were obtained for youths using the Wechsler Abbreviated Scale of Intelligence.[25]

Demographics interview. A bespoke interview (Demographic Interview for Parents) covering general family information, including parental forensic history, schooling, and economic information, was developed specifically for this study by one of the authors (SBu) and was administered to all parents.

Statistical analysis plan

Objectives

Primary

1. To investigate whether the provision of MST could reduce the incidence of out-of-home placements for young people at risk of being removed from their homes because of antisocial behaviour, severe mental health problems, educational problems, or unmet need.
2. To investigate whether the provision of MST could delay the time to first offence and reduce the frequency of offending, as directed by the TSC.

Secondary

1. To investigate whether MST is associated with:
   - increases in wellbeing
   - improved educational outcomes
   - improved family functioning.
2. To establish the cost of MST relative to MAU and the cost-effectiveness of providing MST.

Endpoints

The primary endpoint was the proportion of cases assigned to long-term (≥3 months) out-of-home placement in specialist residential provision between randomisation and the 18-month time point. The outcome was coded as treatment failure when there was no out-of-home placement but home observation data and self-report measures suggested that the young person's situation was markedly suboptimal. A sensitivity analysis was conducted to exclude out-of-home placements that were judged to be beneficial.

Antisocial behaviour was measured as the time to an offence resulting in a pre-court disposal or a court disposal as well as self-report and parent report measures of anti-social activity. In addition, as a previous smaller UK study found callous–unemotional traits (as assessed using the Inventory of Callous and Unemotional Traits [ICUT])[26] was sensitive to treatment effects, ICUT was included as a measure of asociality as well as a moderator of treatment effects. Antisocial behaviour outcomes relevant to the educational context included school attendance (measured as the percentage of days attended), reports from teachers (as measured by the Conners Comprehensive Behaviour Rating Scale-Teacher report form)[27] and the Self-Report Delinquency measure.[11]

Parenting was measured using youth and parent versions of the monitoring and supervision subscale of the APQ, as well as the total score from the parent-rated Loeber Caregiver Questionnaire. Family functioning was measured by the change on the FACES-IV questionnaire and the CTS.

Wellbeing was assessed by the change on the SDQ and MFQ for youths and the GHQ for parents.

All outcomes were measured for all participants at 6, 12, and 18 months after randomisation.

Analysis population

All randomised participants were included in the analysis.

Statistical analysis

Primary outcome

The primary analysis was a logistic regression. Clustering by therapist was accounted for by including a random therapist effect. The analysis included centre and participants' number of past convictions, sex, age at onset of criminal behaviour, and other risk indicators (eg, criminal associations) as fixed effects. The logistic regression was fitted using generalised estimating equations. A Wald test of the effect of intervention was used as the primary analysis. As a secondary analysis, tests of interaction were used to explore whether the interventions differed according to participants' (1) sex, (2) age, (3) presence of a criminal record, and (4) referral path. Clustering by therapist was accounted for by computing robust standard errors.

Key forensic outcomes

The antisocial behaviour outcome (time to offence) was analysed using a Cox regression, as for the primary outcome.

Other secondary outcomes

All other secondary outcomes were modelled using linear mixed-effects models, with separate treatment effects for the 6-, 12-, and 18-month outcomes and an unstructured covariance matrix. The intervention effect on the

18-month outcome was tested using a Wald test. We also opted for an explicit modelling of the temporal effects, with treatment effects on the linear (and if necessary quadratic) slope, followed up by tests of marginal effects of treatment at each time point applying a linear mixed-effects model with a linear effect of time, random participant effect with robust standard errors (SEs) as above) and a treatment × time interaction, tested using a Wald test. In fact, we carried out these analyses as well and found almost identical patterns of significant findings; these are not reported but the observed and modelled mean scores are incorporated in appendix ii.

Tests of interaction were performed for all secondary outcomes for which a nominally significant treatment effect was found.

Missing data

It was anticipated that the primary outcome would have very little missing data, as the data were obtained independently of the study participants. For the secondary outcomes, linear mixed models and Cox regression yield valid inferences when data are missing at random (ie, the probability of a particular data point being missing depends only on observed data). It is possible that missing data may be missing not at random, so we conducted sensitivity analyses to explore the impact of missing data before the imputations were undertaken. These are reported in Table A7.

The questionnaire data and offending data, together with the other clinical baseline covariates and treatment arm, were included in the multiple imputations, and 30 replicates were generated. Each replicate was analysed with the same linear mixed-effects model used for the secondary outcomes. Results were combined using Rubin's rules to account for between-replicate variability and estimates obtained were used in computing group differences. Sensitivity analyses revealed that the multiple imputation made only minor differences to the results except in reducing the confidence intervals around estimates, so the report was based on the non-imputed outcomes but results based on imputed outcomes are shown in the tables in Appendix ii, with divergent findings also noted in the text of the main paper. There was a high proportion of missing data for the educational and teacher-rated outcomes at both baseline and follow-up, and for DAWBA variables at 12 months, so we used multiple imputation (without post-baseline offending data) with 30 replicates for the primary analysis of these outcomes.

Appendix ii: Results

In order to test the impact of missingness we contrasted the baseline scores of participants who completed the study with those participants who failed to provide data at the 12-month time point. There were no significant differences found between these groups on the *t*-test or χ^2 test.

Table A7: Baseline variables grouped according to whether the young person dropped out by 12 months or not for the entire sample (N=683)

	Dropped out by 12 months		Not dropped out by 12 months	
	n or mean	SD or %	n	SD or %
Demographics				
Number	194		489	
Mean age (years)	13·8	1·4	13·8	1·4
Female	61	31·4	189	38·7
Mean SES (range 1–6)	3·0	1·5	2·9	1·3
Family income				
% on state benefits or <£20k pa	147	75·8	378	77·3
Ethnicity				
White British/European*	141	73·1	394	80·6
Black African/Afro-Caribbean	26	13·5	45	9·2
Asian	3	1·6	13	2·7
Mixed/Other	19	9·8	32	6·5
Parents' marital status				
Single or widowed	75	38·7	198	40·5
Separated or divorced	37	19·1	99	20·2
Married or cohabiting	80	41·2	190	38·9
Number of siblings*	2·6	1·4	2·4	1·3
Siblings offending	68	37·4	176	38·4
Offences in year prior to referral				
Non-offender on referral	59	30·6	176	36·2
Total number of offences	1·3	2·3	1·1	2·3
Violent offences*	0·5	1·1	0·3	0·9
Non-violent offences	0·6	1·2	0·5	1·3
Number with custodial sentences	3	1·6	7	1·4
Comorbid diagnosis				
Conduct disorder	149	77·6	383	79
Oppositional defiant disorder	6	3·1	22	4·5
Any conduct disorder	152	79·2	402	82·9
Social phobia	5	2·6	16	3·3
Obsessive-compulsive disorder	1	0·5	2	0·4
Posttraumatic stress disorder	11	5·7	40	8·2
Separation anxiety disorder	10	5·2	12	2·5

Specific phobia	2	1	17	3·5
Generalised anxiety disorder*	0	0	15	3·1
Panic disorder	3	1·6	5	1
ADHD	59	30·7	145	29·9
ADHD Hyperactivity	1	0·5	10	2·1
ADHD Inattention	5	2·6	20	4·1
PDD/autism	2	1	5	1
Eating disorders	2	1	2	0·4
Tic disorder	4	2·1	7	1·4
Major depression	27	14·1	45	9·3
Any emotional disorder	48	25	115	23·7
Mixed anxiety/conduct disorder	21	10·9	81	16·7
Number without diagnosis	32	16·7	68	14
Average number of Axis I diagnoses	1·5	1·1	1·6	1
Onset of conduct disorder	81	41·8	216	44·2
ICUT score	33·4	9·6	33	9·7
Peer delinquency score (SRDM)	4·9	4·6	5	4·7

ADHD=attention deficit hyperactivity disorder. ICUT=Inventory of Callous Emotional Traits. PDD=pervasive developmental disorder. SES=socioeconomic status. SRDM-Self-Report Delinquency Measure. *Significant at $p<0.05$ (none significant at $p<0.01$)—Wilcoxon rank sum test used for continuous outcomes and Wilson proportion test for binary outcomes.

Secondary outcomes

As described in the main paper, the trial made an effort to collect a comprehensive set of outcome measures in order to inform a variety of stakeholders with keen interest in this evaluation. In this Appendix we report a slightly expanded set of variables pertaining to the secondary outcomes reported in the main paper. The outcome domains of the trial were objective offending and self-reported youth offending and antisociality outcomes, adolescent wellbeing outcomes, and family functioning outcomes. Young people's antisocial behaviour has been shown to decrease in association with increased parental supportiveness and decreased conflict between parents.[13,14] We collected data on variables that target key mechanisms of change identified in previous studies of MST, that is, parenting skills, family functioning and young people's associations with deviant peers. As the expected mechanism of change was through improvement in parenting capacity,[15] we aimed to evaluate parenting skills in some detail. Thus, parent-report and youth-report measures of parenting skills and family functioning were collected. As antisocial behaviour is highly likely to co-occur with internalizing mental health problems, we measured both self-report and parent-report of wellbeing in the young people, as well as an indication of parental mental health and adjustment. Diagnostic data collected at baseline and 12 months are also reported. Economic analyses are reported separately in detail at the end of this expanded results section (Appendix iii).

Overview

In order to offer the reader a simple overview of the findings from the secondary outcomes collected we provide a set of forest plots summarising the difference between the groups at 6, 12, and 18 months (Figure A3). The plots are organised according to the source of information (young people or parents) and for completeness include the key scales of the questionnaires used in the study. The plots are helpful in showing graphically that young people's behaviour and experience observed by their parents indicated greater benefit from MST than those noted by the young people themselves. Further, they illustrate how effect size estimates are larger at 6 months, immediately after treatment ended, and generally disappear at later times of testing.

Figure A3: Standardised differences between Multisystemic Therapy and management as usual groups on secondary outcome variables
a–c: Parent-rated variables (A, 6 months; B, 12 months; C, 18 months). d–f: Young people's self-rated variables (D, 6 months; E, 12 months; F, 18 months). ADHD=Conners Attention Deficit and Hyperactivity T-score. FACES=Family Adaptability and Cohesion Evaluation Scale. L&L=Conners Language & Learning T-score. SDQ=Strengths and Difficulties Questionnaire.

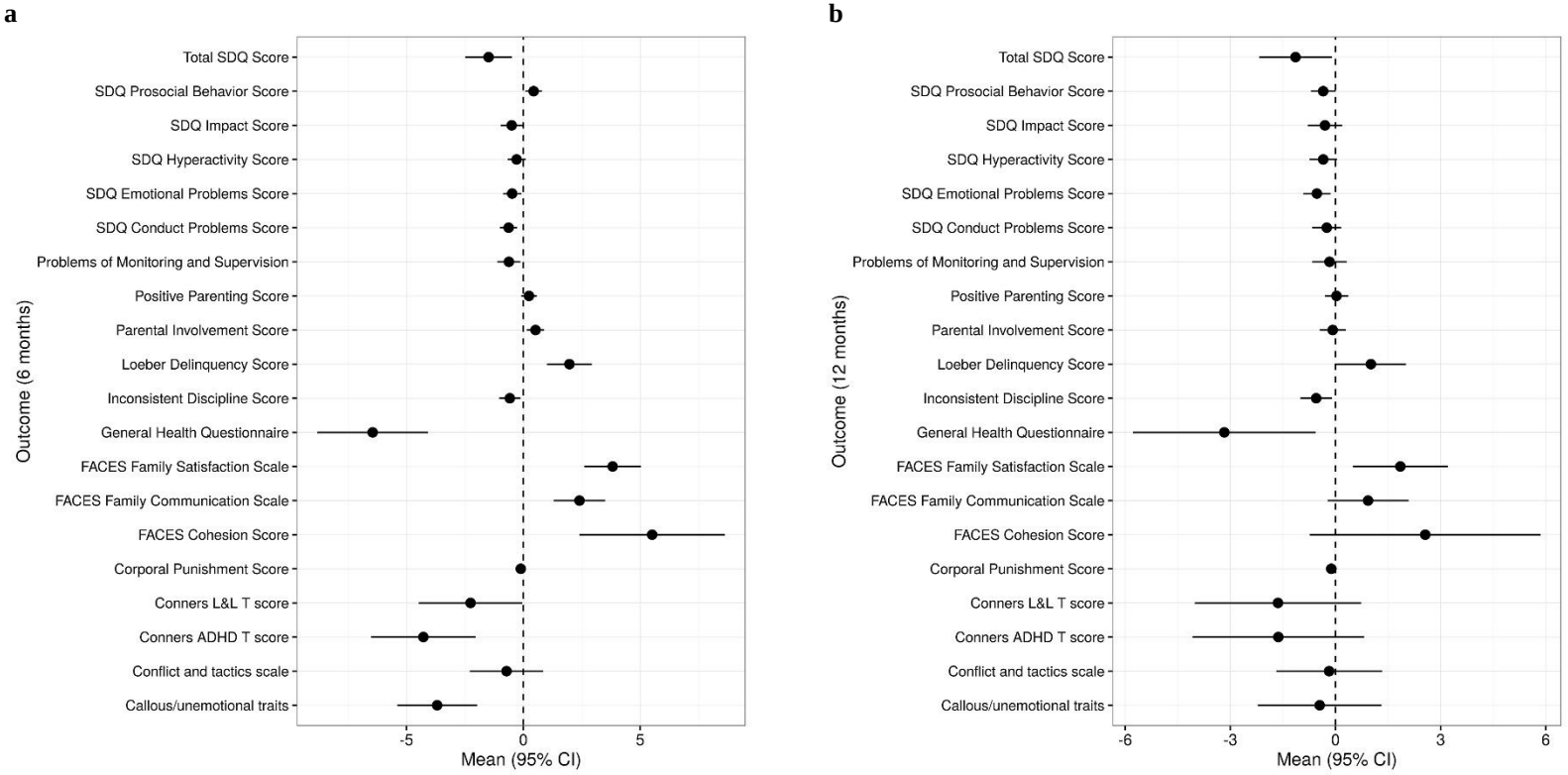
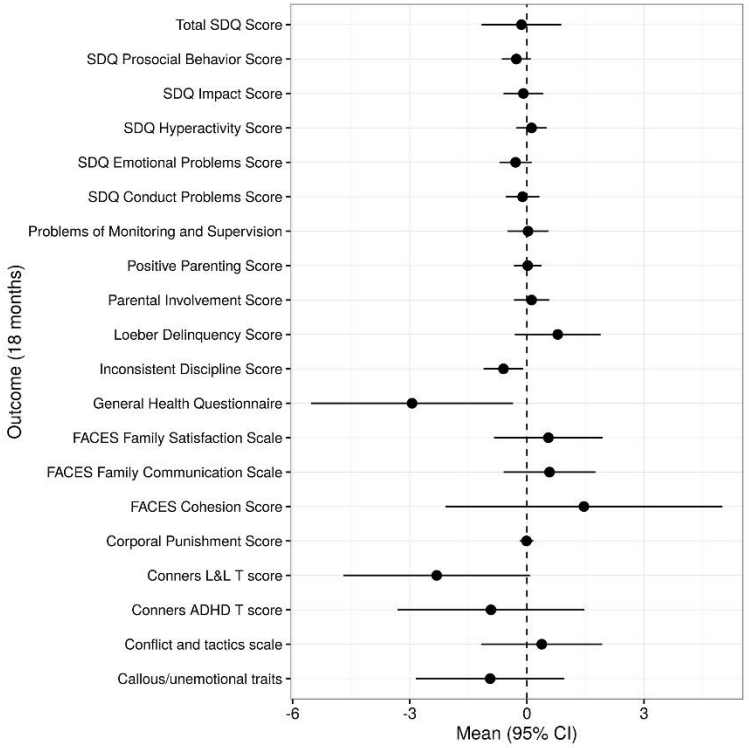


Figure A3, continued

c



d

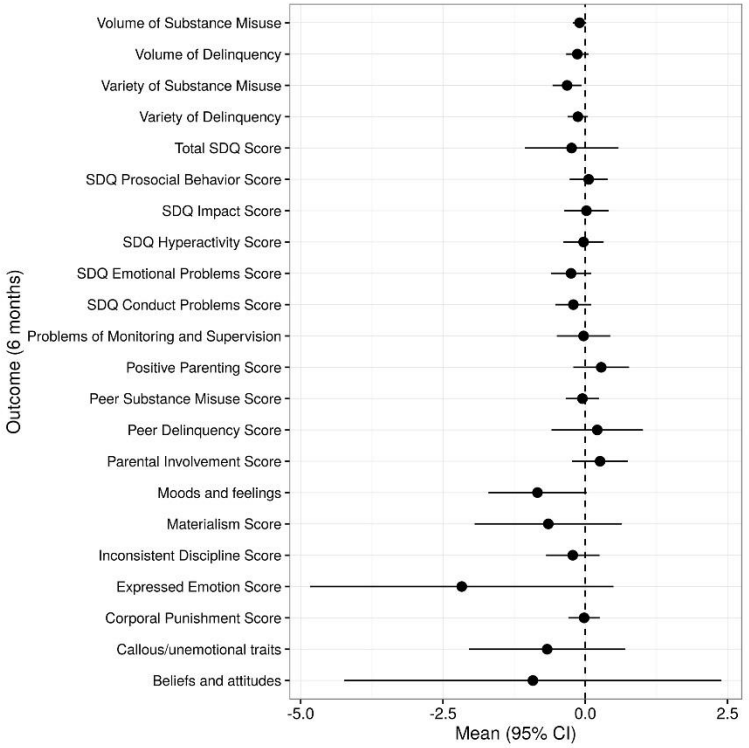
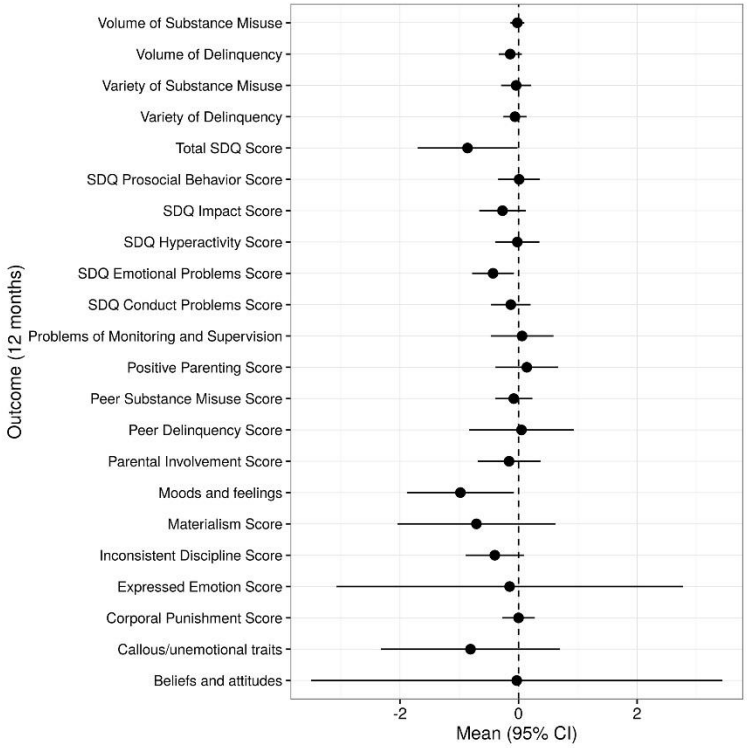
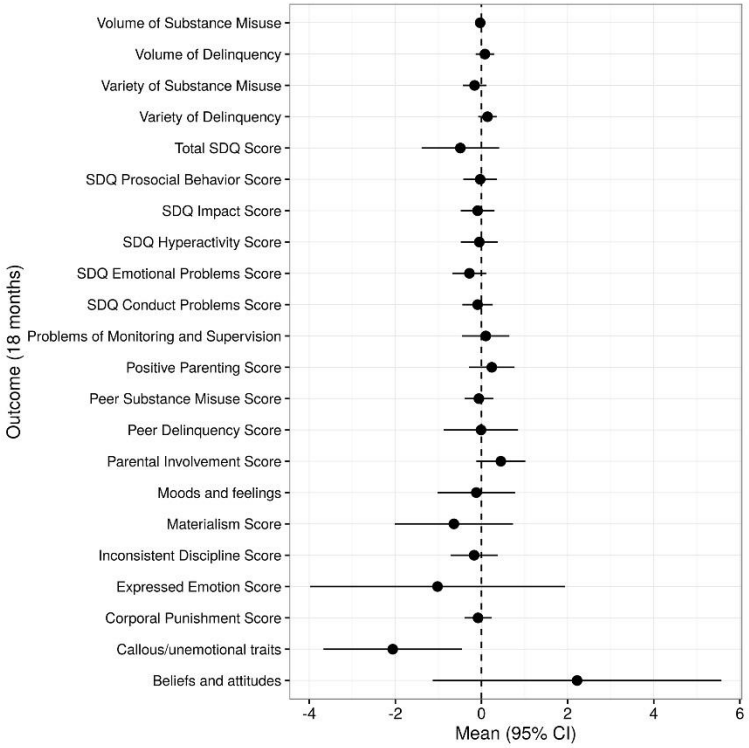


Figure A3, continued

e



f



Antisocial behaviour and attitudes

Young people's self-reported delinquency
As described above, in order the estimate the impact of failures to adhere to the assessment protocol we used multiple imputation techniques using other clinical baseline covariates, questionnaire and offending data in multiple imputations. The models used were identical to the linear mixed-effects models adopted for examining the observed data and the results from the 30 analyses were integrated using Rubin's rules. Tables A8a and b report on significance testing for the secondary outcomes reported in Table 4A and B in the main text of the paper, but here using a multiple imputation procedure. Multiple imputations confirmed the data analyses performed on observed values. Figure A4 a–d displays the results obtained for youth- and parent-reported SDQ conduct problems, as well as callous–unemotional traits as reported by the young person and parent. The ICUT completed by the young person was the only instrument in this battery to yield significant group differences at 18 months post-randomisation, while immediately post-treatment (ie, 6-month follow-up) parents' rating of the young person's callous–unemotional traits reflected greater gain following the MST intervention.

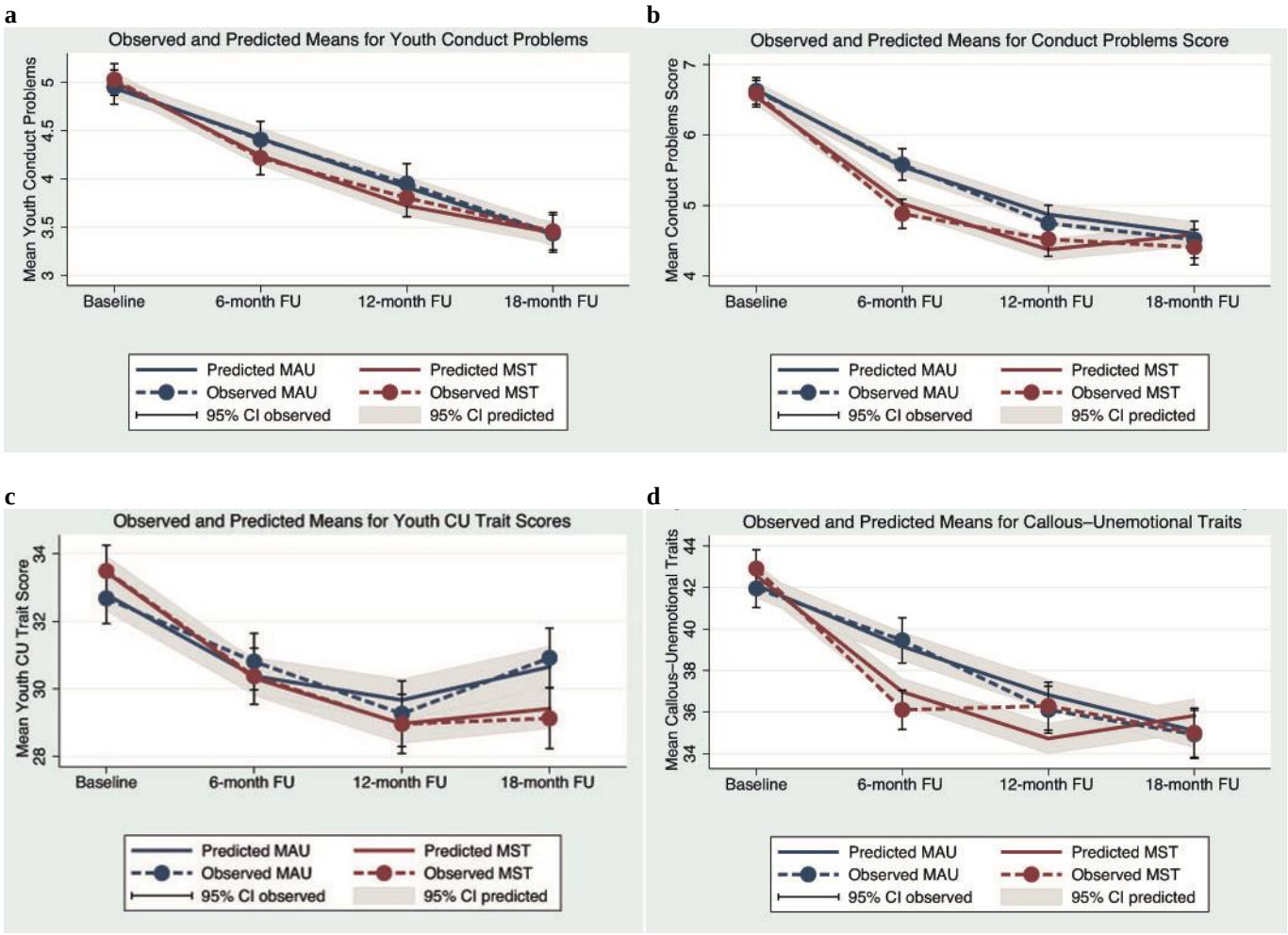
The prevalence and incidence of delinquent behaviour such as vandalism, theft, and burglary were monitored using the Self-Report Delinquency Measure,[11] which also yields a peer delinquency assessment. MST was expected to achieve decreases in associations with antisocial peers, increases in positive peer relations, and greater commitment to prosocial activities (eg, education). This prediction was consistent with the model and hypothesised mediating mechanisms proposed by the developers of MST.[3] Table A8b and Figure A5a–e display the multiply imputed results obtained by using this instrument. The MST and MAU groups were distinguished only in terms of substance misuse at the 6-month observation point, when young people in the MST group claimed to use fewer substances, and to have lower substance use, than those in MAU. No differences in terms of peer delinquency or self-reported delinquent acts emerged at any point. Non-compliance and increasingly serious forms of antisocial behaviour, together with young people's perceptions of law-abiding behaviour and institutions, were measured using the Antisocial Beliefs and Attitudes Scale.[12] Results from multiple imputations aligned with analysis of observed values and are shown in Table A8b and in Figure A5f. Measures of antisocial attitudes did not differentiate the groups at any time point. Similarly, no differences in youth materialism were evident at any point (Figure A5g).

Table A8a: Parent report and young person's self-report of antisocial behaviour and attitudes: estimates based on multiple imputation procedure

	Group (n) and between-group significance (*t*-test)	SDQ conduct problems (YP) Mean (SD) [n]	SDQ conduct problems (P) Mean (SD) [n]	ICUT (YP) Mean (SD) [n]	ICUT (P) Mean (SD) [n]
6 months–baseline	MST	5·0 (2·1) [n=340]	6·59 (2·41) [n=340]	33·5 (9·7) [n=341]	42·91 (11·58) [n=341]
	MAU	4·9 (2·3) [n=340]	6·62 (2·45) [n=340]	32·7 (9·6) [n=339]	41·96 (11·74) [n=339]
6-month follow-up	MST	4·2 (2·0) [n=290]	4·8 (2·5) [n=290]	30·3 (9·8) [n=292]	35·9 (11·3) [n=292]
	MAU	4·5 (2·2) [n=264]	5·5 (2·5) [n=268]	30·6 (9·7) [n=268]	39·3 (11·8) [n=268]
	Effect (95% CI)	–0·21 (–0·50 to 0·08)	–0·62 (–0·99 to –0·25)	–0·70 (–2·05 to 0·65)	–3·72 (–5·39 to –2·05)
	p value	0·17	<0·0001	0·31	<0·0001
12-month follow-up	MST	4·0 (2·2) [n=252]	4·6 (2·6) [n=246]	28·9 (9·3) [n=248]	36·0 (12·1) [n=248]
	MAU	3·9 (2·1) [n=237]	4·8 (2·7) [n=237]	29·3 (9·7) [n=238]	36·4 (11·7) [n=238]
	Difference (95% CI)	–0·11 (–0·42 to 0·20)	–0·25 (–0·66 to 0·16)	–1·11 (–2·54 to 0·32)	–0·64 (–2·42 to 1·14)
	p value	0·49	0·22	0·13	0·48
18-month follow-up	MST	3·4 (2·0) [n=221]	4·4 (2·5) [n=232]	29·2 (9·5) [n=234]	35·1 (11·6) [n=234]
	MAU	3·5 (1·9) [n=193]	4·6 (2·5) [n=209]	30·6 (9·2) [n=217]	35·5 (11·9) [n=217]
	Difference (95% CI)	–0·10 (–0·43 to 0·23)	–0·16 (–0·57 to 0·25)	–2·07 (–3·60 to –0·54)	–1·07 (–2·97 to 0·83)
	p value	0·57	0·46	0·0085	0·27

ICUT=Inventory of Callous and Unemotional Traits. SDQ=Strengths and Difficulties Questionnaire. P=parent report. YP=young person's report.

Figure A4: Observed and model-predicted means (based on the explicit modelling of the temporal effects) for (a) young people's self-reported and (b) parent-reported delinquency, and (c) young people's self-reported and (d) parent-reported callous–unemotional traits



Data were obtained using the Strengths and Difficulties Questionnaire and the Inventory of Callous and Unemotional Traits. CU-callous–unemotional. FU=follow-up. MAU=management as usual. MST=Multisystemic Therapy.

Table A8b: Young person's self-report of delinquent behaviour, antisocial beliefs and attitudes, and materialism: estimates based on multiple imputation procedure

	Group (n) and between-group significance (*t*-test)	SRDM Variety of delinquent acts Mean (SD) [n]	SRDM Volume of delinquent acts Mean (SD) [n]	SRDM Variety of substance misuse Mean (SD) [n]	SRDM Volume of substance misuse Mean (SD) [n]	SRDM Peer Delinquency Mean (SD) [n]	ABAS Mean (SD) [n]	Youth Materialism Scale Mean (SD) [n]
6 months–baseline	MST	4·8 (3·6) [n=337]	19·7 (18·3) [n=337]	0·8 (1·7) [n=337]	1·6 (3·7) [n=337]	5·0 (4·7) [n=337]	60·8 (23·1) [n=341]	37·0 (8·9) [n=342]
	MAU	3·1 (3·7) [n=335]	20·9 (19·0) [n=335]	0·7 (1·3) [n=335]	1·5 (3·0) [n=335]	4·9 (4·7) [n=335]	61·7 (24·4) [n=339]	37·6 (8·9) [n=341]
6-month follow-up	MST	3·9 (3·5) [n=288]	15·7 (17·1) [n=288]	0·7 (1·5) [n=288]	1·5 (3·1) [n=288]	4·7 (4·4) [n=288]	55·5 (24·0) [n=292]	–0·72 (–2·03 to 0·59)
	MAU	4·4 (3·8) [n=262]	17·6 (17·7) [n=262]	0·8 (1·5) [n=262]	1·8 (3·2) [n=262]	4·9 (4·9) [n=262]	58·0 (23·5) [n=268]	0·28
	Effect	–0·11 (–0·29 to 0·07)	–0·14 (–0·34 to 0·06)	–0·31 (–0·56 to –0·06)	–0·10 (–0·22 to 0·02)	0·25 (–0·57 to 1·07)	–0·88 (–4·07 to 2·31)	36·4 (9·4) [n=293]
	p value	0·224	0·165	0·016	0·073	0·560	0·590	37·0 (9·0) [n=263]
12-month follow-up	MST	3·3 (3·4) [n=243]	12·3 (15·6) [n=243]	0·8 (1·8) [n=243]	1·8 (3·8) [n=243]	5·0 (5·1) [n=243]	54·5 (23·5) [n=248]	–0·65 (–2·02 to 0·72)
	MAU	3·3 (3·3) [n=230]	12·6 (14·1) [n=230]	0·7 (1·3) [n=230]	1·5 (2·5) [n=230]	5·0 (4·9) [n=230]	54·7 (22·5) [n=238]	0·35
	Difference (95% CI)	–0·04 (–0·24 to 0·16)	–0·14 (–0·34 to 0·06)	–0·05 (–0·30 to 0·20)	–0·02 (–0·14 to 0·10)	0·00 (–0·82 to 0·82)	–0·04 (–3·55 to 3·47)	36·6 (9·5) [n=252]
	p value	0·672	0·165	0·736	0·761	0·991	0·982	36·9 (9·2) [n=238]
18-month follow-up	MST	2·9 (3·4) [n=231]	10·4 (14·3) [n=231]	0·7 (1·4) [n=231]	1·5 (2·7) [n=231]	4·7 (5·0) [n=231]	53·1 (23·6) [n=234]	–0·63 (–2·06 to 0·80)
	MAU	2·5 (2·6) [n=215]	9·6 (12·0) [n=215]	0·7 (1·2) [n=215]	1·4 (2·2) [n=215]	5·0 (5·3) [n=215]	52·8 (23·6) [n=217]	0·39
	Difference	0·14 (–0·08 to 0·36)	0·08 (–0·14 to 0·30)	–0·16 (–0·45 to 0·13)	–0·03 (–0·15 to 0·09)	–0·01 (–0·85 to 0·83)	2·19 (–1·24 to 5·62)	37·0 (8·9) [n=241]
	p value	0·196	0·506	0·288	0·585	0·989	0·210	37·6 (8·9) [n=211]

Data were obtained using the Self-Report Delinquency Measure (SRDM), Antisocial Beliefs and Attitudes Scale (ABAS), and Youth Materialism Scale. *Due to apparent heteroscedastic residuals, the difference between arms, CI, and p-values are from a linear mixed-effects model with a log-transform.

Figure A5: Observed and model-predicted means (based on the explicit modelling of the temporal effects) for young people's (a–e) self-reported delinquency, (f) antisocial beliefs and attitudes, and (g) materialism

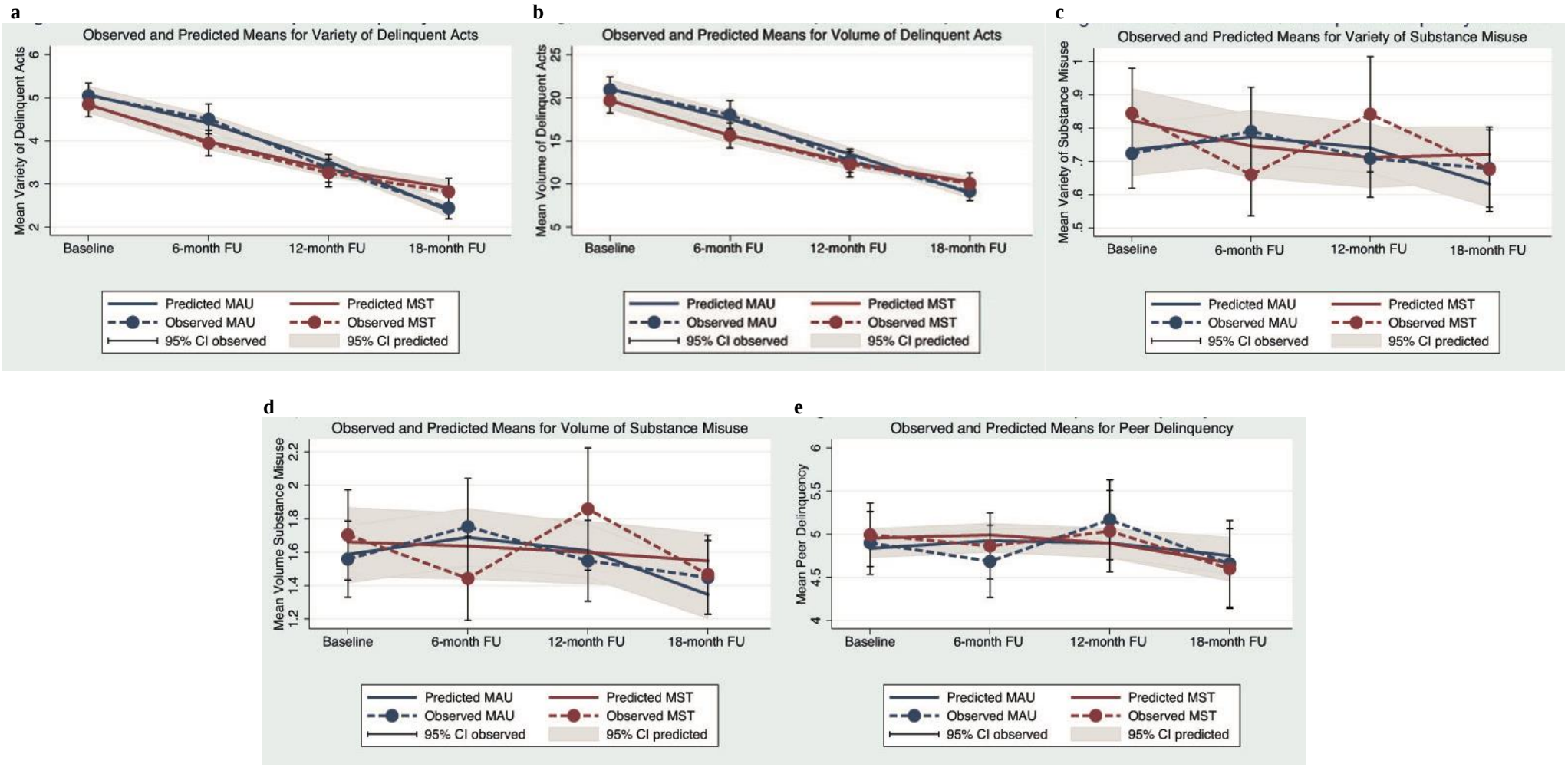
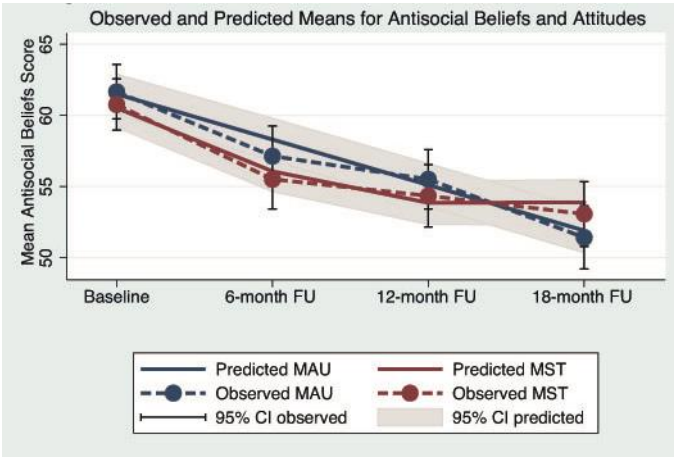
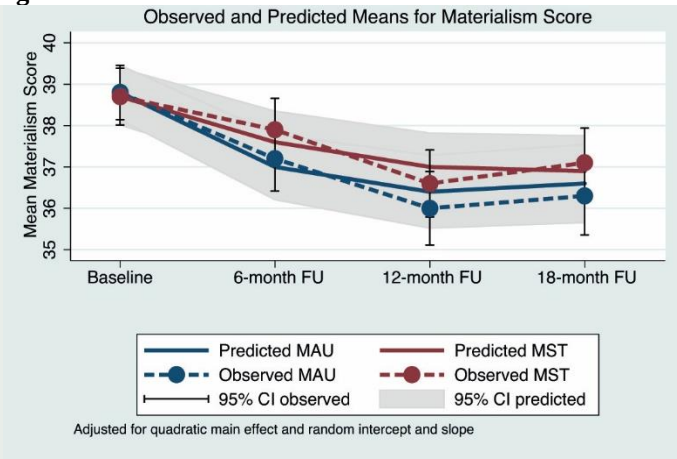


Figure A5, continued

f



g



FU=follow-up. MAU=management as usual. MST=Multisystemic Therapy.

Parenting skills and family functioning

Parents' reporting on their own parenting behaviour
Parenting skills were assessed using the Monitoring and Supervision scale from the APQ based on information provided by the parent and the young person, and the Parental Support Scale of the Loeber Caregiver Questionnaire. Table A9 displays the results from the APQ prespecified parenting variables based on multiple imputations; these are also displayed in Figure A6. Tables A11 and A12 display information from further parent-reported APQ scales for both observed and imputed datasets, respectively, followed by graphical illustrations of observed and estimated means in Figures A6b and A8a–d. Greater parental involvement and reductions in problems of monitoring and supervision were evident in the MST group at 6 months, but there was no longer a significant between-group difference at later observation points. By contrast, the difference in terms of lower levels of inconsistent discipline in the MST group persisted at all time points, including the final follow-up. Positive parenting and corporal punishment were not significantly different between the groups.

Young people's reports on their experience of their parents' parenting behaviour
The quality of parenting practices could also be evaluated from the young people's perspective using the APQ. The results are displayed in Table A9 for prespecified outcomes based on multiple imputation to complement the observed data (reported in Table 5A of the main text), and in Tables A13 and A14 for additional variables for observed and multiply imputed datasets, respectively. Figures A6a and A9a–d illustrate these findings. The results suggest that young people in the MST group noticed little change in parenting behaviour across these broad set of scales, with no recognition of increased monitoring and supervision or greater parental involvement, which were reported by their parents (see above and Tables A11 and A12). Nor were young people in the MST group aware of differences in terms of lower levels of inconsistent discipline, which parents in the MST group reported at all time points. Positive parenting and corporal punishment, which did not distinguish the two groups in terms of parents' reports, appeared not to distinguish them from the young people's perspective either.

Parents' report on family functioning and marital conflict
The imputed dataset showed Loeber parental support scores (Table A9, Figure A6c) to be significantly higher for the MST group compared with the MAU group at 6 months, but this level was not maintained, and the difference was no longer significant at the later observation points (although MST showed marginally significant superiority at 12 months when data from the multiple imputation procedure were analysed). Measures of expressed emotion did not differentiate the two groups at any time point. For the results based on observed values, see Table 5A in the main paper.

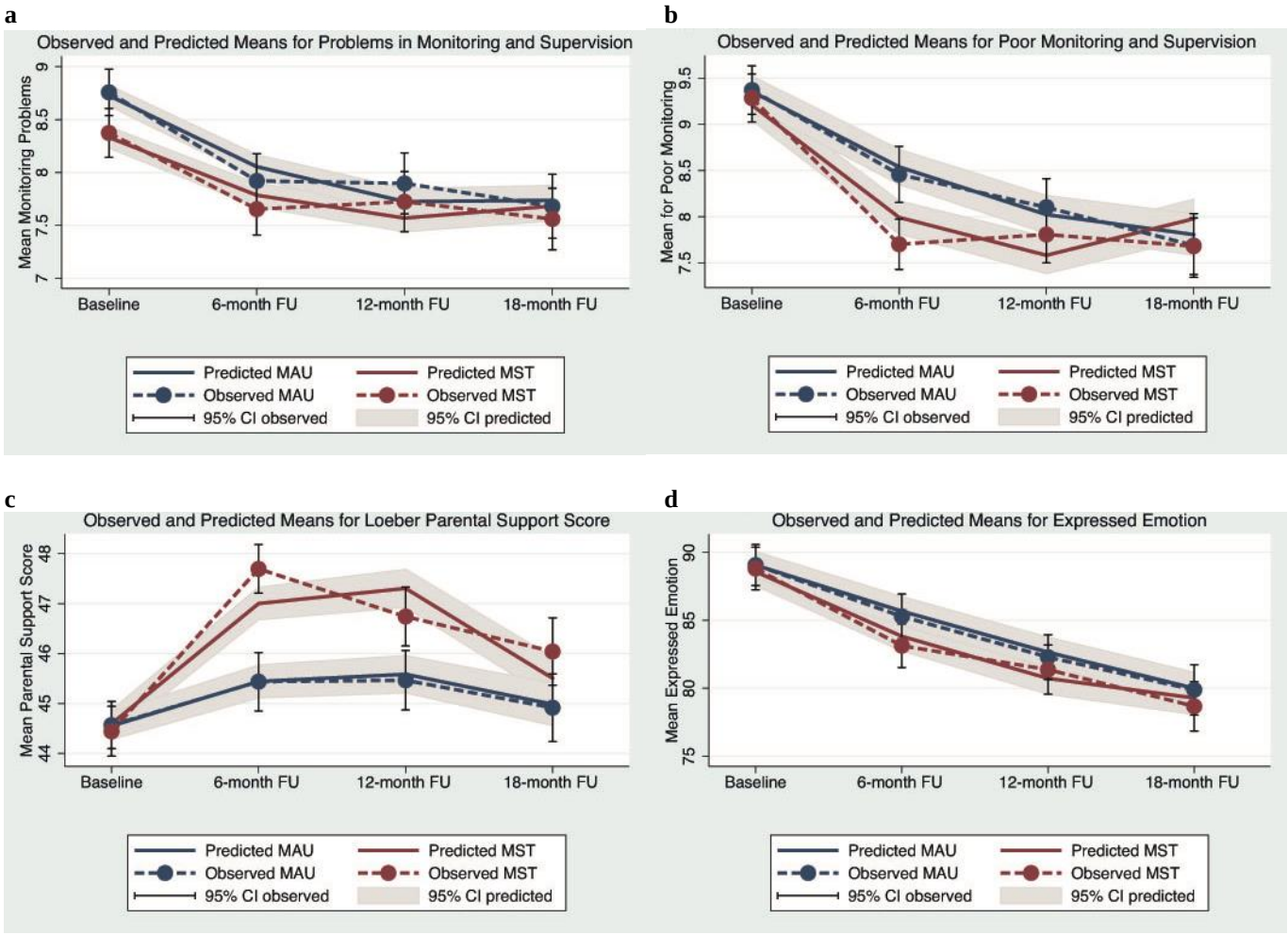
The quality of the parent–adolescent relationship, family functioning, and parenting practices were evaluated using the FACES-IV. Interparental disruption was measured using the short form of the CTS. Results from these instruments are shown in Table A10 for imputed samples and in Figure A7a–d. In line with the previous observations of parental reports, family cohesion, family communication, and family satisfaction ratings all favoured the MST group at 6 months. Family satisfaction remained superior in families assigned to MST at 12 months, but by 18 months the MAU group reached similar levels of family satisfaction. The measure of interparental conflict yielded comparable levels for the two groups but both declined markedly over the study period. For the results based on observed values for these measures, see Table 5B in the main paper.

Table A9: Young people's and parents' report on parenting skills and family functioning, using estimates based on multiple imputation procedure

	Group and between-group significance	APQ Problems of monitoring and supervision (YP) Mean (SD) [n]	APQ Problems of monitoring and supervision (P) Mean (SD) [n]	Loeber parental support score (P) Mean (SD) [n]	Level of Expressed Emotion (YP) Mean (SD) [n]
6 months–baseline	MST	8·4 (3·0) [n=341]	9·29 (3·33) [n=341]	44·44 (6·40) [n=337]	88·8 (20·0) [n=341]
	MAU	8·8 (2·8) [n=339]	9·37 (3·34) [n=339]	44·57 (6·04) [n=335]	89·1 (19·1) [n=339]
6-month follow-up	MST	7·7 (3·0) [n=292]	7·7 (3·2) [n=292]	47·6 (5·7) [n=288]	83·3 (18·6) [n=292]
	MAU	8·0 (2·9) [n=261]	8·5 (3·4) [n=268]	45·5 (6·7) [n=262]	86·6 (18·9) [n=268]
	Difference (95% CI)	–0·02 (–0·51 to 0·47)	–0·60 (–1·07 to –0·13)	1·94 (0·98 to 2·90)	–1·91 (–4·65 to 0·83)
	p value	0·94	0·013	0·00019	0·17
12-month follow-up	MST	7·8 (3·0) [n=246]	7·8 (3·3) [n=248]	46·9 (6·3) [n=243]	81·9 (19·6) [n=248]
	MAU	7·8 (3·1) [n=233]	8·1 (3·2) [n=238]	45·5 (6·5) [n=230]	82·6 (17·9) [n=238]
	Difference (95% CI)	0·07 (–0·42 to 0·56)	–0·15 (–0·64 to 0·34)	0·99 (–0·01 to 1·99)	–0·70 (–3·68 to 2·28)
	p value	0·78	0·54	0·055	0·65
18-month follow-up	MST	7·9 (3·1) [n=235]	7·7 (3·3) [n=234]	45·9 (6·9) [n=231]	78·8 (19·2) [n=234]
	MAU	8·0 (3·1) [n=206]	7·9 (3·4) [n=217]	45·0 (6·8) [n=215]	80·4 (18·3) [n=217]
	Difference (95% CI)	0·12 (–0·37 to 0·61)	0·05 (–0·48 to 0·58)	0·71 (–0·35 to 1·77)	–1·22 (–4·32 to 1·88)
	p value	0·63	0·85	0·19	0·44

Data were obtained using the Alabama Parenting Questionnaire (APQ), Loeber Caregiver Questionnaire (Loeber), and Level of Expressed Emotion. MAU=management as usual. MST=Multisystemic Therapy. YP=completed by young person. P=completed by parent.

Figure A6: (a) Young people's and (b) parents' report of parenting skills (APQ); (c) parent report of parental support (Loeber); (d) young people's report of level of expressed emotion



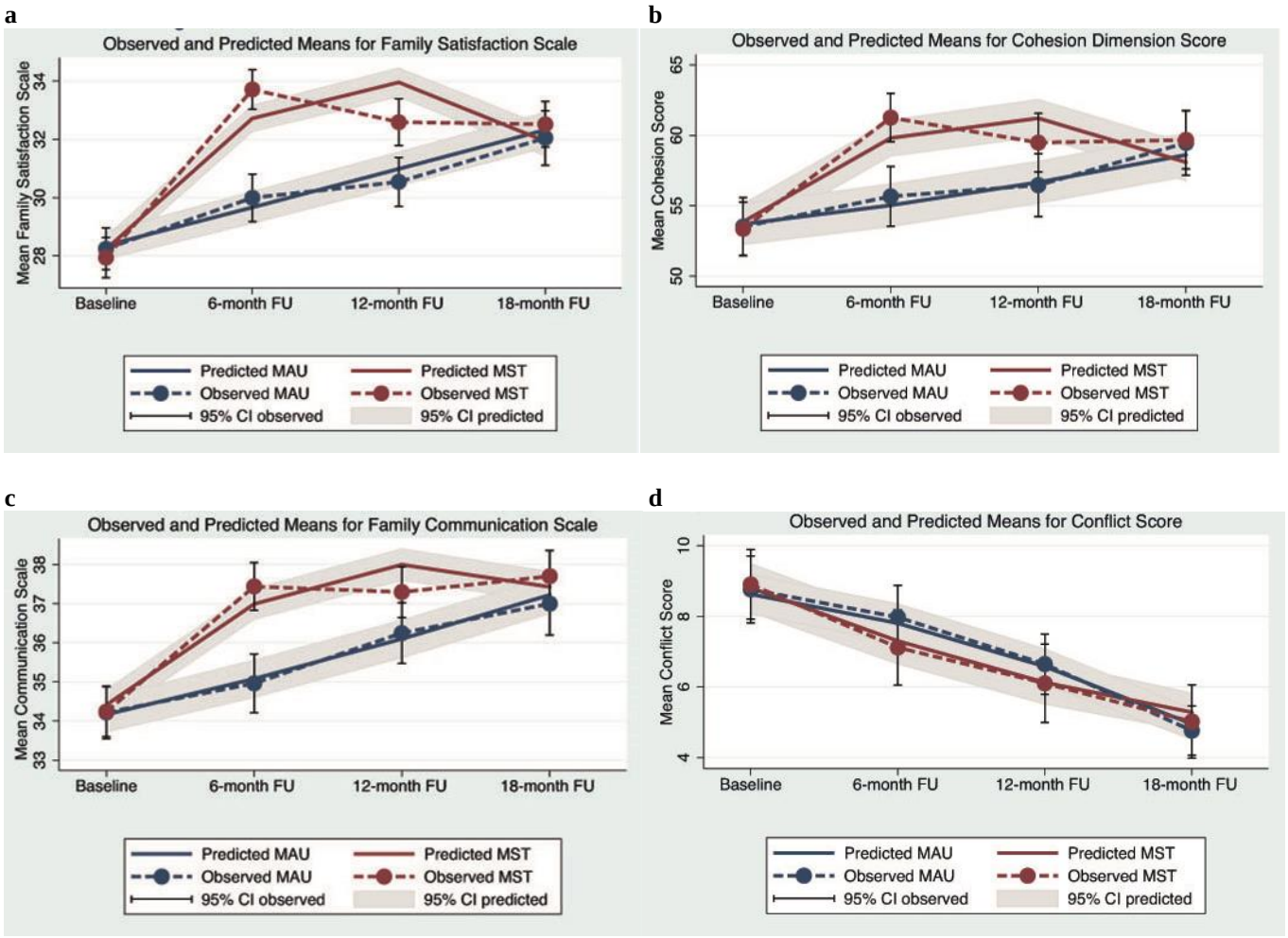
Data were obtained using the Alabama Parenting Questionnaire (APQ), Loeber Caregiver Questionnaire (Loeber), and Level of Expressed Emotion. MAU=management as usual. MST=Multisystemic Therapy.

Table A10: Parents' report on family functioning: estimates based on multiple imputation procedure

	Group (n) and between-group significance	FACES-IV family satisfaction Mean (SD)	FACES-IV cohesion Mean (SD)	FACES-IV family communication Mean (SD)	CTS Mean (SD)
6 months–baseline	MST (n=337)	27·94 (8·73)	53·36 (22·54)	34·24 (8·11)	8·90 (9·87)
	MAU (n=335)	28·24 (9·09)	53·52 (24·21)	34·22 (8·55)	8·77 (9·72)
6-month follow-up	MST (n=288)	33·5 (8·0)	61·2 (18·9)	37·4 (7·0)	7·6 (9·8)
	MAU (n=262)	30·3 (9·1)	55·8 (21·6)	35·2 (8·3)	7·1 (7·9)
	Difference (95% CI)	3·77 (2·50 to 5·04)	5·59 (2·22 to 8·96)	2·42 (1·34 to 3·50)	–0·60 (–2·11 to 0·91)
	p value	<0·0001	0·00058	<0·0001	0·44
12-month follow-up	MST (n=243)	33·2 (8·6)	60·9 (19·3)	37·5 (6·9)	6·1 (8·7)
	MAU (n=230)	30·7 (8·6)	56·3 (21·0)	36·2 (7·7)	6·8 (8·9)
	Difference (95% CI)	1·94 (0·53 to 3·35)	2·68 (–0·63 to 5·99)	1·03 (–0·15 to 2·21)	–0·30 (–2·02 to 1·42)
	p value	0·0022	0·11	0·086	0·74
18-month follow-up	MST (n=231)	32·6 (8·0)	59·4 (19·3)	38·0 (6·8)	4·9 (7·8)
	MAU (n=215)	32·3 (9·1)	58·4 (20·3)	37·4 (8·0)	5·9 (7·9)
	Difference (95% CI)	0·45 (–0·86 to 1·76)	1·16 (–2·39 to 4·71)	0·62 (–0·60 to 1·84)	0·15 (–1·50 to 1·80)
	p value	0·50	0·52	0·32	0·86

CTS=Conflict Tactics Scale-Short Form. FACES-IV=Family Adaptability and Cohesion Evaluation Scales. Loeber=Loeber Caregiver Questionnaire. MAU=management as usual. MST=Multisystemic Therapy.

Figure A7: Observed and model-predicted means (based on the explicit modelling of the temporal effects) for (a–c) parents' reports on family functioning (FACES-IV) and (d) family conflict (CTS)



CTS=Conflict Tactics Scale-Short Form. FACES-IV=Family Adaptability and Cohesion Evaluation Scales. FU=follow-up. MAU=management as usual. MST=Multisystemic Therapy.

Table A11: Parents' self-report on their parenting behaviour

	Group (n) and between-group significance	APQ Parent involvement Mean (SD)	APQ Positive parenting Mean (SD)	APQ Corporal punishment Mean (SD)	APQ Inconsistent discipline Mean (SD)
6 months–baseline	MST (n=341)	9·81 (2·52)	12·64 (2·22)	4·02 (1·57)	9·46 (2·83)
	MAU (n=339)	9·57 (2·40)	12·66 (2·42)	3·86 (1·33)	9·30 (2·56)
6-month follow-up	MST (n=292)	10·4 (2·3)	13·0 (2·2)	3·4 (1·0)	8·4 (2·8)
	MAU (n=268)	9·7 (2·5)	12·8 (2·3)	3·5 (1·1)	9·0 (2·8)
	Difference (95% CI)	0·52 (0·15 to 0·89)	0·20 (–0·13 to 0·53)	–0·13 (–0·29 to 0·03)	–0·60 (–1·01 to –0·19)
	p value	0·0066	0·22	0·12	0·0052
12-month follow-up	MST (n=248)	10·2 (2·5)	13·0 (2·1)	3·4 (0·9)	8·5 (2·6)
	MAU (n=238)	10·0 (2·3)	12·8 (2·3)	3·5 (1·0)	8·9 (2·6)
	Difference (95% CI)	–0·08 (–0·47 to 0·31)	0·00 (–0·35 to 0·35)	–0·16 (–0·34 to 0·02)	–0·50 (–0·95 to –0·05)
	p value	0·69	0·99	0·067	0·030
18-month follow-up	MST (n=234)	10·5 (2·5)	12·9 (2·2)	3·4 (1·0)	8·4 (2·5)
	MAU (n=217)	10·0 (2·5)	12·8 (2·4)	3·4 (1·0)	9·0 (2·6)
	Difference (95% CI)	0·06 (–0·35 to 0·47)	–0·03 (–0·38 to 0·32)	–0·02 (–0·20 to 0·16)	–0·53 (–1·00 to –0·06)
	p value	0·79	0·87	0·81	0·029

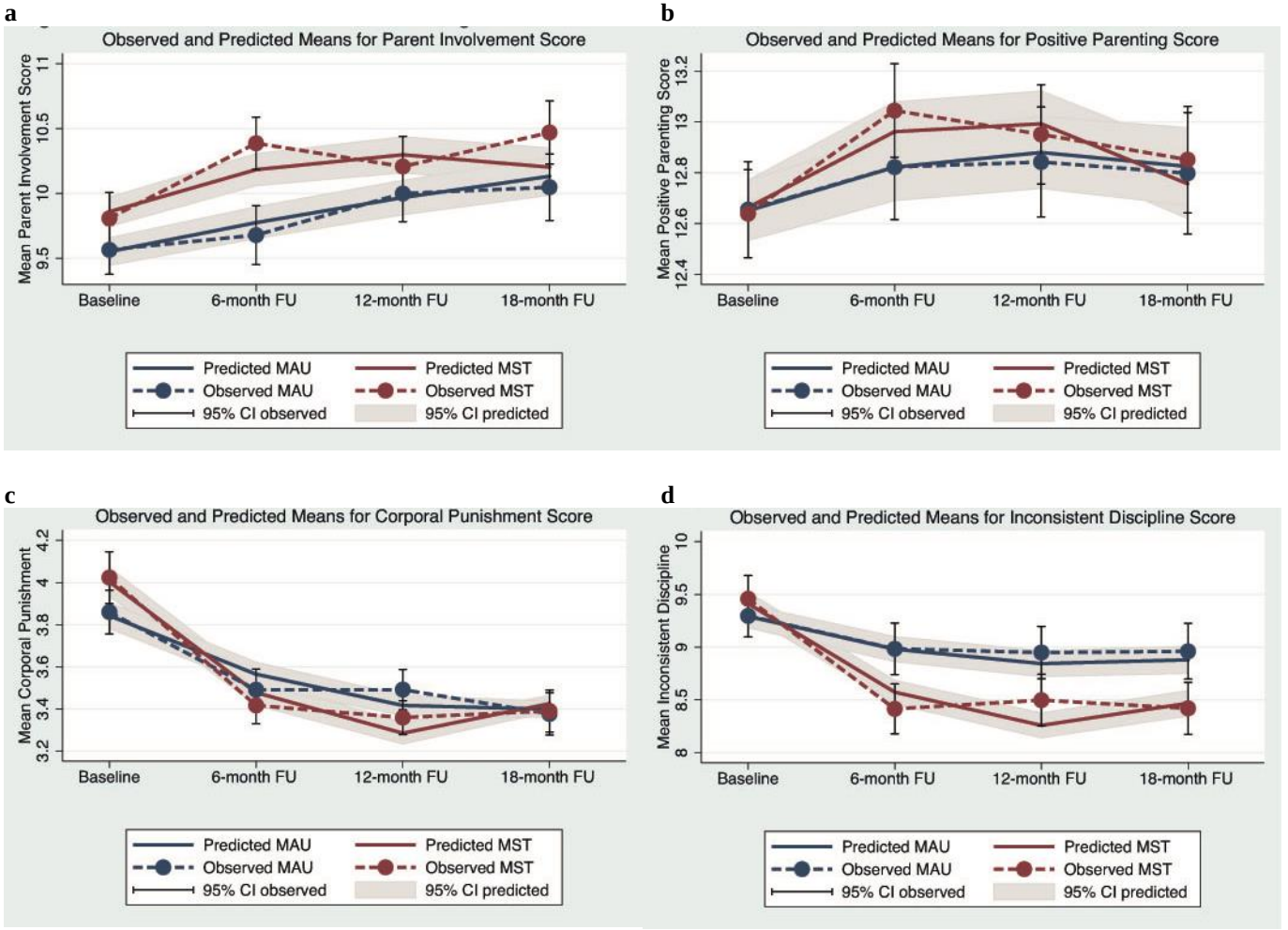
Data were obtained using the Alabama Parenting Questionnaire (APQ). MAU=management as usual. MST=Multisystemic Therapy.

Table A12. Parents' self-report on their parenting behaviour: estimates based on multiple imputation procedure

	Group (n) and between-group significance	APQ Parent involvement Mean (SD)	APQ Positive parenting Mean (SD)	APQ Corporal punishment Mean (SD)	APQ Inconsistent discipline Mean (SD)
6 months–baseline	MST (n=341)	9·81 (2·52)	12·64 (2·22)	4·02 (1·57)	9·46 (2·83)
	MAU (n=339)	9·57 (2·40)	12·66 (2·42)	3·86 (1·33)	9·30 (2·56)
6-month follow-up	MST (n=292)	10·4 (2·3)	13·1 (2·2)	3·4 (1·0)	8·4 (2·8)
	MAU (n=268)	9·6 (2·5)	12·8 (2·4)	3·5 (1·2)	9·0 (2·7)
	Difference (95% CI)	0·48 (0·11 to 0·85)	0·21 (–0·12 to 0·54)	–0·10 (–0·26 to 0·06)	–0·60 (–1·05 to –0·15)
	p value	0·014	0·219	0·197	0·008
12-month follow-up	MST (n=248)	10·2 (2·5)	13·0 (2·2)	3·4 (1·0)	8·4 (2·7)
	MAU (n=238)	9·8 (2·4)	12·9 (2·3)	3·4 (1·0)	8·9 (2·6)
	Difference (95% CI)	–0·07 (–0·46 to 0·32)	0·05 (–0·30 to 0·40)	–0·12 (–0·30 to 0·06)	–0·51 (–0·96 to –0·06)
	p value	0·707	0·789	0·179	0·024
18-month follow-up	MST (n=234)	10·2 (2·4)	12·8 (2·3)	3·4 (1·0)	8·4 (2·5)
	MAU (n=217)	10·0 (2·5)	12·8 (2·3)	3·4 (0·9)	8·7 (2·6)
	Difference (95% CI)	0·12 (–0·31 to 0·55)	0·02 (–0·33 to 0·37)	–0·02 (–0·20 to 0·16)	–0·58 (–1·09 to –0·07)
	p value	0·588	0·908	0·832	0·028

Data were obtained using the Alabama Parenting Questionnaire (APQ). MAU=management as usual. MST=Multisystemic Therapy.

Figure A8: (a–d) Observed and model-predicted means (based on the explicit modelling of the temporal effects) for parents' self-report on their parenting behaviour using the Alabama Parenting Questionnaire



FU=follow-up. MAU=management as usual. MST=Multisystemic Therapy.

Table A13: Young people's report on parenting behaviour

	Group (n) and between-group significance	APQ Parent involvement Mean (SD)	APQ Positive parenting Mean (SD)	APQ Corporal punishment Mean (SD)	APQ Inconsistent discipline Mean (SD)
6 months–baseline	MST (n=341)	7·9 (3·2)	10·0 (3·3)	4·3 (2·3)	8·4 (3·0)
	MAU (n=339)	7·9 (3·1)	10·3 (3·2)	4·3 (2·3)	8·2 (2·9)
6-month follow-up	MST (n=292)	8·7 (3·1)	10·9 (3·1)	3·9 (2·1)	8·0 (3·0)
	MAU (n=261)	8·4 (3·1)	10·7 (3·2)	3·9 (1·8)	8·2 (2·9)
	Difference (95% CI)	0·34 (–0·15 to 0·83)	0·30 (–0·19 to 0·79)	–0·02 (–0·29 to 0·25)	–0·28 (–0·75 to 0·19)
	p value	0·174	0·239	0·895	0·246
12-month follow-up	MST (n=246)	8·6 (3·1)	11·1 (3·2)	3·7 (1·7)	7·6 (3·0)
	MAU (n=233)	8·7 (3·2)	11·0 (3·3)	3·7 (1·7)	7·9 (2·9)
	Difference (95% CI)	–0·14 (–0·67 to 0·39)	0·21 (–0·32 to 0·74)	–0·03 (–0·32 to 0·26)	–0·41 (–0·92 to 0·10)
	p value	0·601	0·433	0·849	0·112
18-month follow-up	MST (n=235)	9·0 (3·4)	11·2 (3·1)	3·5 (1·3)	7·7 (3·3)
	MAU (n=206)	8·5 (3·3)	11·1 (3·2)	3·6 (1·6)	7·9 (3·1)
	Difference (95% CI)	0·47 (–0·08 to 1·02)	0·21 (–0·34 to 0·76)	–0·08 (–0·39 to 0·23)	–0·28 (–0·81 to 0·25)
	p value	0·091	0·451	0·632	0·306

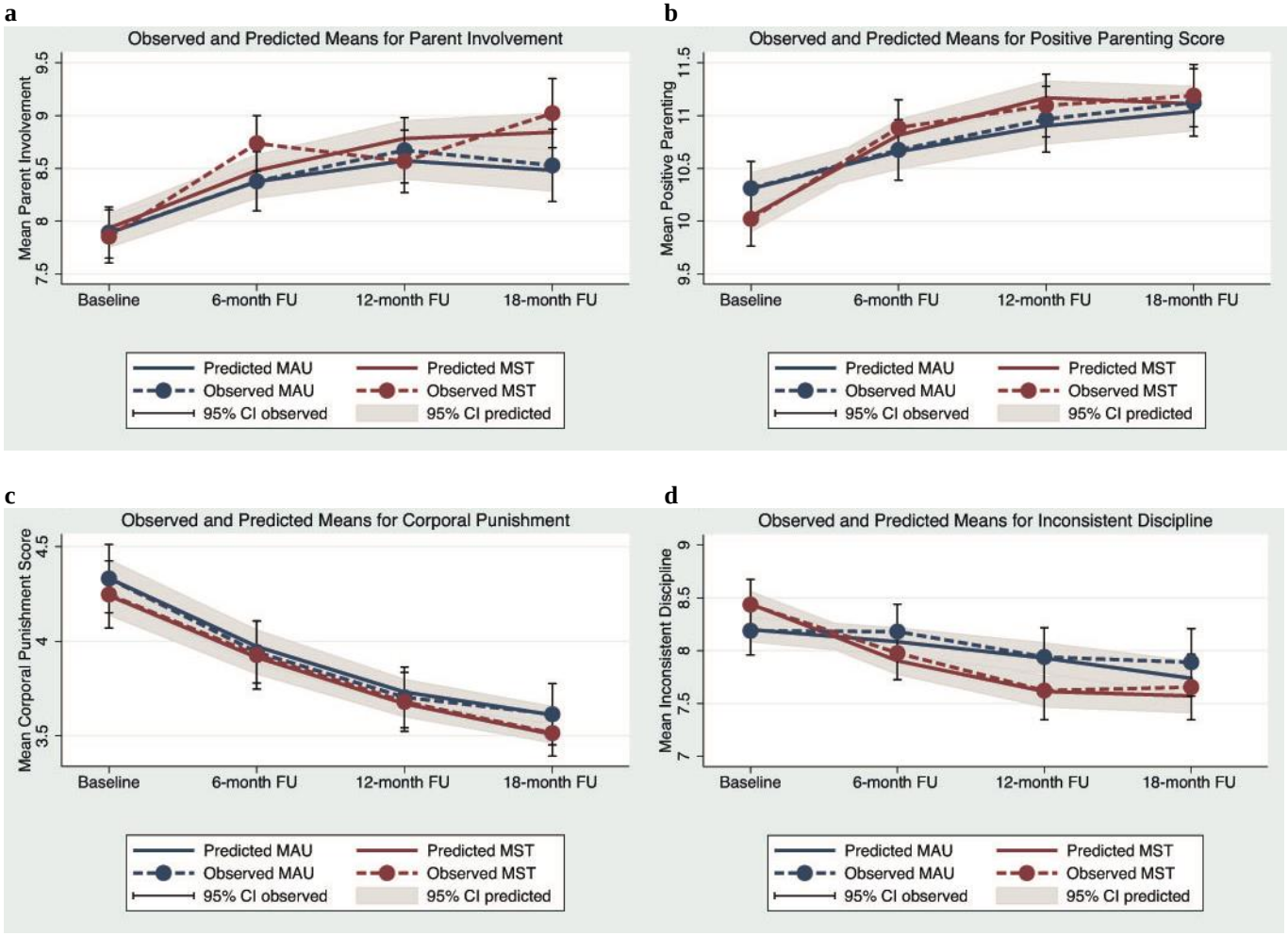
Data were obtained using the Alabama Parenting Questionnaire (APQ). MAU=management as usual. MST=Multisystemic Therapy.

Table A14: Young people's report on parenting behaviour: estimates based on multiple imputation procedure

	Group (n) and between-group significance	APQ Parent involvement Mean (SD)	APQ Positive parenting Mean (SD)	APQ Corporal punishment Mean (SD)	APQ Inconsistent discipline Mean (SD)
6 months–baseline	MST (n=341)	7·9 (3·2)	10·0 (3·3)	4·3 (2·3)	8·4 (3·0)
	MAU (n=339)	7·9 (3·1)	10·3 (3·2)	4·3 (2·3)	8·2 (2·9)
6-month follow-up	MST (n=292)	8·6 (3·1)	10·9 (3·1)	3·9 (2·1)	7·9 (3·0)
	MAU (n=261)	8·2 (3·2)	10·6 (3·2)	4·0 (1·9)	8·1 (2·9)
	Difference (95% CI)	0·30 (–0·17 to 0·77)	0·27 (–0·22 to 0·76)	–0·02 (–0·29 to 0·25)	–0·22 (–0·67 to 0·23)
	p value	0·210	0·289	0·885	0·344
12-month follow-up	MST (n=246)	8·4 (3·2)	11·0 (3·1)	3·8 (1·9)	7·5 (3·0)
	MAU (n=233)	8·4 (3·2)	10·9 (3·3)	3·7 (1·7)	7·6 (2·9)
	Difference (95% CI)	–0·11 (–0·62 to 0·40)	0·24 (–0·25 to 0·73)	–0·01 (–0·30 to 0·28)	–0·35 (–0·84 to 0·14)
	p value	0·686	0·351	0·944	0·161
18-month follow-up	MST (n=235)	8·8 (3·2)	11·0 (3·2)	3·5 (1·3)	7·6 (3·2)
	MAU (n=206)	8·5 (3·2)	11·0 (3·3)	3·6 (1·6)	7·8 (3·1)
	Difference (95% CI)	0·46 (–0·13 to 1·05)	0·23 (–0·32 to 0·78)	–0·10 (–0·37 to 0·17)	–0·19 (–0·70 to 0·32)
	p value	0·129	0·414	0·474	0·475

Data were obtained using the Alabama Parenting Questionnaire (APQ). MAU=management as usual. MST=Multisystemic Therapy.

Figure A9: (a–d) Observed and model-predicted means (based on the explicit modelling of the temporal effects) for young people's report on parenting behaviour using the Alabama Parenting Questionnaire



FU=follow-up. MAU=management as usual. MST=Multisystemic Therapy.

Wellbeing and adjustment

Youth self-report measures of wellbeing and attitudes
The analysis based on multiple imputations supported the observation that, in marked contrast to the SDQs completed by the parents (see below), the SDQs completed by the young people failed to show substantial differences between the MST and MAU groups (Table A15, Figure A10a–e). Surprisingly, only the reduction of emotional problems score indicated an advantage for those in the MST group, particularly at the 12-month follow-up point. By 18-month follow-up the reduction relative to MAU was no longer significant. Similarly, less depression was reported on the MFQ by the MST group at 6- and 12-month follow-up (see Figure A10f).

Parents' ratings of young people's behaviour and emotional wellbeing, and of their own wellbeing
Table A16 displays parental ratings on the SDQ at 6, 12 and 18 months for multiply imputed values. Observed means and fitted values are shown in Figure A11a–e. Overall, at 6 months the young people assigned to MST were rated lower in terms of conduct problems, emotional problems, and overall impact, and higher in terms of prosocial behaviour. At 12 months, only emotional problem ratings favoured the MST group, and by 18 months none of the scales distinguished the groups.

Similar results were yielded by the multiple imputation procedure in the analysis of the Conners ADHD scales, and parents' own wellbeing as reflected in the GHQ responses (see Table A16 and Figure A11f for Conners ADHD and Figure A11g for GHQ). At 6 months, parents in the MST group rated the young people's Language and Learning Problems scores as being lower (Conners; Table A17), as was also seen with observed and imputed ADHD scores (Conners; see Table 6B in the main text for observed data and Table A16 for imputed values). These advantages disappeared by 12 and 18 months. Further scales on the Conners, based on both multiply imputed teacher ratings, are shown in Table A18.

Data derived from multiple imputations (Table A16) closely mirrored the findings for parents' self-reported wellbeing score on the GHQ (see Table 6B in the main text), indicating significant advantages for the MST group that were maintained up to and including the 18-month assessment.

Teachers' reports on young people
The Conners Comprehensive Behaviour Rating Scale – Teacher report form provided further educational outcomes and included an evaluation of participants' emotional and behavioural functioning in the classroom.[27] As the rate of completion of these forms was relatively poor, with more than 33% of data missing, multiple imputations were used to estimate the impact of the intervention on classroom behaviour. The mean ratings are shown in Table A18. There was no evidence that teachers were able to identify behavioural benefits of MST in terms of reduced disruptive behaviour, improved learning, reduced mood problems, reduced anxiety problems, or (as noted above in relation to wellbeing) reduced difficulties with attention and hyperactivity (Table A16). Interestingly, unlike most other measures used in this study, these measures appeared to provide little evidence of any improvement in either intervention group across observations.

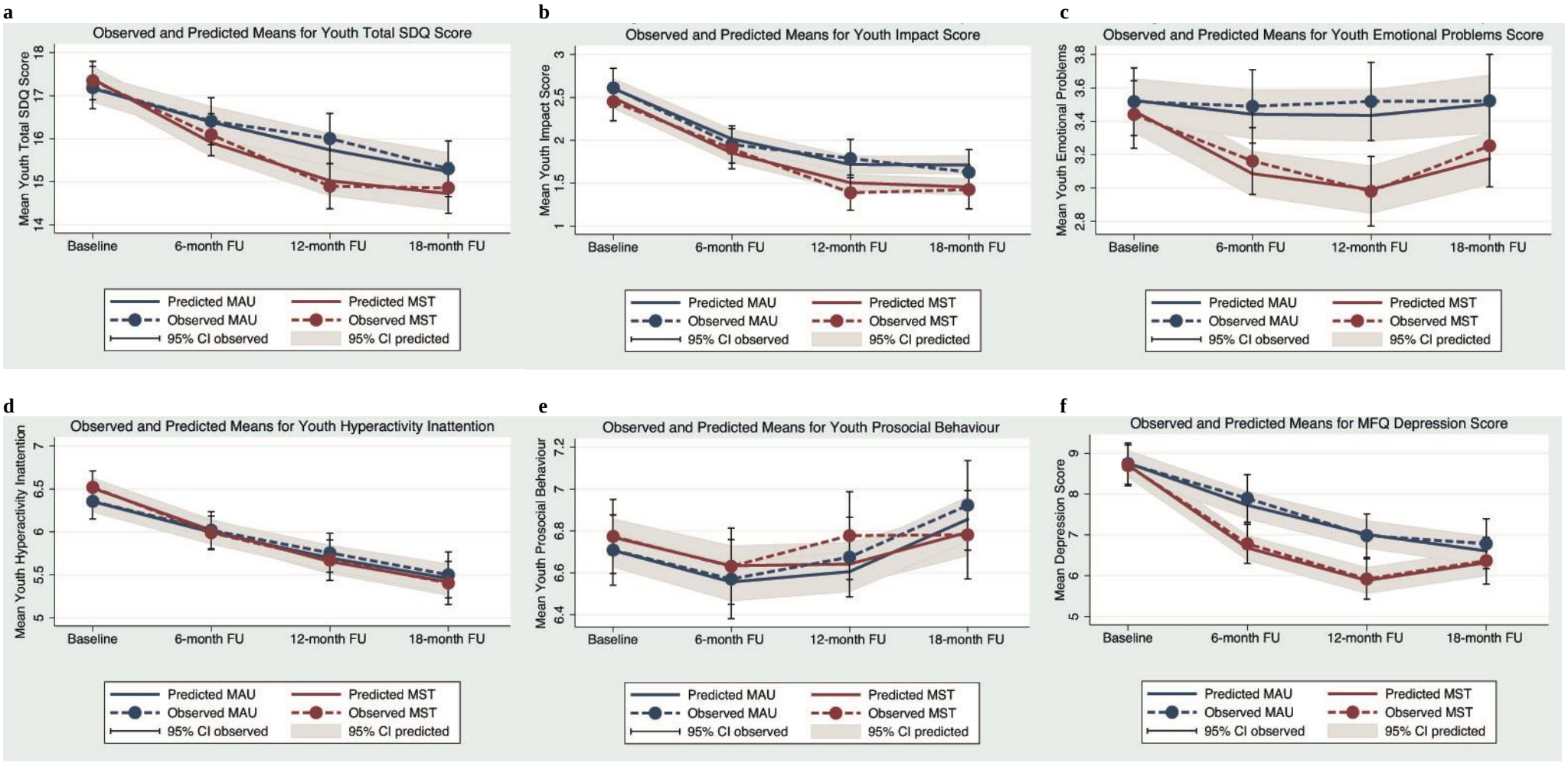
Psychiatric disorders
Psychiatric disorders were identified and a psychosis screen provided by use of the DAWBA. This computerised structured interview measure was administered to both the parents and young people at baseline and at 12 months; 72% of the sample was assessed. The clinician-rated mental health outcomes on the DAWBA using multiple imputation with baseline educational outcomes and demographic covariates are shown in Table 6C in the main paper. At intake, all but 15% of the sample had one or more psychiatric diagnoses. By 12 months, 40% were without a diagnosis. The prevalence of conduct disorder diagnosis was over 80% at baseline, and decreased to less than 46% at 12 months. However, there was no evidence that diagnostic status in any of the major categories was linked to either intervention.

Table A15: Young people's self-report of their wellbeing and behaviour: estimates based on multiple imputation procedure

	Group (n) and between-group significance	Total SDQ score Mean (SD) [n]	SDQ impact score Mean (SD) [n]	SDQ emotional problems score Mean (SD) [n]	SDQ hyperactivity/ inattention Mean (SD) [n]	SDQ prosocial behaviour Mean (SD) [n]	MFQ Mean (SD) [n]
6 months–baseline	MST	17·4 (5·7) [n=340]	2·5 (2·8) [n=340]	3·4 (2·6) [n=340]	6·5 (2·5) [n=340]	6·8 (2·3) [n=340]	8·7 (6·4) [n=-341]
	MAU	17·2 (6·3) [n=340]	2·6 (2·9) [n=340]	3·5 (2·6) [n=340]	6·4 (2·6) [n=340]	6·7 (2·1) [n=340]	8·7 (6·4) [n=339]
6-month follow-up	MST	16·0 (5·6) [n=290]	1·8 (2·5) [n=290]	3·0 (2·3) [n=290]	6·0 (2·3) [n=290]	6·6 (2·2) [n=290]	6·7 (5·6) [n=292]
	MAU	16·3 (6·0) [n=264]	1·9 (2·4) [n=264]	3·4 (2·4) [n=264]	6·0 (2·3) [n=264]	6·5 (2·2) [n=264]	7·5 (6·4) [n=268]
	Difference (95% CI)	–0·19 (–0·99 to 0·61)	0·00 (–0·37 to 0·37)	–0·24 (–0·59 to 0·11)	0·00 (–0·37 to 0·37)	0·05 (–0·28 to 0·38)	–0·80 (–1·64 to 0·04)
	p value	0·641	0·982	0·167	0·998	0·759	0·060
12-month follow-up	MST	15·3 (5·4) [n=252]	1·4 (2·2) [n=252]	3·0 (2·3) [n=252]	5·8 (2·5) [n=252]	6·8 (2·3) [n=252]	6·1 (5·5) [n=248]
	MAU	15·9 (5·8) [n=237]	1·7 (2·3) [n=237]	3·5 (2·5) [n=237]	5·7 (2·3) [n=237]	6·6 (2·1) [n=237]	6·7 (5·6) [n=238]
	Difference (95% CI)	–0·81 (–1·67 to 0·05)	–0·26 (–0·63 to 0·11)	–0·42 (–0·77 to –0·07)	0·00 (–0·37 to 0·37)	0·05 (–0·28 to 0·38)	–0·93 (–1·81 to –0·05)
	p value	0·067	0·185	0·024	0·996	0·760	0·038
18-month follow-up	MST	14·6 (5·8) [n=221]	1·6 (2·2) [n=221]	3·2 (2·5) [n=221]	5·3 (2·5) [n=221]	6·8 (2·0) [n=221]	6·4 (6·1) [n=234]
	MAU	15·5 (5·7) [n=193]	1·7 (2·5) [n=193]	3·6 (2·6) [n=193]	5·4 (2·5) [n=193]	6·8 (2·2) [n=193]	6·6 (5·8) [n=217]
	Difference (95% CI)	–0·58 (–1·52 to 0·36)	–0·12 (–0·55 to 0·31)	–0·28 (–0·67 to 0·11)	–0·08 (–0·47 to 0·31)	–0·05 (–0·42 to 0·32)	–0·22 (–1·10 to 0·66)
	p value	0·224	0·587	0·171	0·691	0·794	0·630

Data were obtained using the Strengths and Difficulties Questionnaire (SDQ) and the Short Mood and Feelings Questionnaire (MFQ). MAU=management as usual. MST = Multisystemic Therapy.

Figure A10: Observed and model-predicted means (based on the explicit modelling of the temporal effects) for young people's self-report of their wellbeing and behaviour on (a–e) the SDQ and (f) the MFQ



FU=follow-up. MAU=management as usual. MFQ=Mood and Feelings Questionnaire. MST=Multisystemic Therapy. SDQ=Strengths and Difficulties Questionnaire.

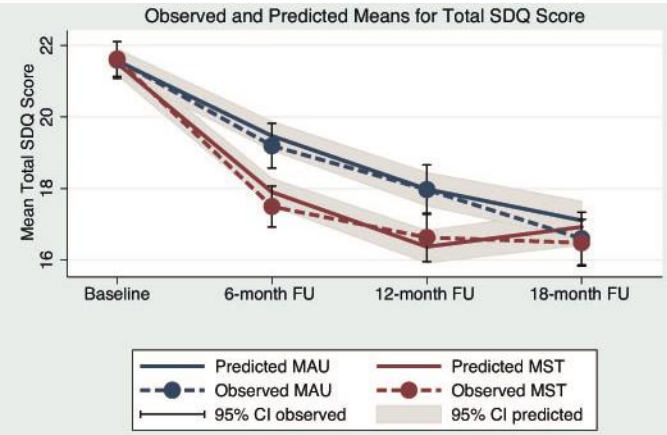
Table A16: Parents' and teachers' report of young people's wellbeing and behaviour and parents' own wellbeing: estimates based on multiple imputation procedure

	Group (n) and between-group significance	Total SDQ score (P) Mean (SD) [n]	SDQ impact score (P) Mean (SD) [n]	SDQ emotional problems score (P) Mean (SD) [n]	SDQ hyperactivity/ inattention (P) Mean (SD) [n]	SDQ prosocial behaviour (P) Mean (SD) [n]	Conners ADHD (P) Mean (SD) [n]	Conners ADHD (T) Mean (SD) [n]	GHQ Mean (SD) [n]
6 months–baseline	MST	21·6 (6·2) [n=340]	5·30 (2·73) [n=340]	4·21 (2·75) [n=340]	7·60 (2·38) [n=340]	5·25 (2·51) [n=340]	80·2 (12·3) [n=341]	74.2 (12.9) [n=213]	64·07 (16·46) [n=341]
	MAU	21·6 (6·5) [n=340]	5·29 (2·95) [n=340]	4·22 (2·64) [n=340]	7·56 (2·53) [n=340]	5·38 (2·50) [n=340]	79·0 (13·2) [n=339]	73.7 (12.8) [n=217]	62·29 (18·34) [n=339]
6-month follow-up	MST	17·3 (6·7) [n=290]	3·4 (3·0) [n=290]	3·3 (2·6) [n=290]	6·3 (2·5) [n=290]	5·9 (2·4) [n=290]	71·7 (15·2) [n=292]	69·3 (16·2) [n=150]	52·2 (15·0) [n=292]
	MAU	18·8 (6·9) [n=268]	3·9 (3·1) [n=268]	3·7 (2·7) [n=268]	6·6 (2·6) [n=268]	5·6 (2·4) [n=268]	75·9 (15·3) [n=268]	69·1 (16·6) [n=155]	58·6 (18·0) [n=268]
	Difference (95% CI)	–1·46 (–2·44 to –0·48)	–0·51 (–0·98 to –0·04)	–0·48 (–0·85 to –0·11)	–0·29 (–0·66 to 0·08)	0·44 (0·09 to 0·79)	–4·42 (–6·79 to –2·05)	0·27 (–1·63 to 2·17)	–6·52 (–8·97 to –4·07)
	p value	0·004	0·032	0·013	0·135	0·013	0·000	0·78	0·000
12-month follow-up	MST	16·9 (6·9) [n=246]	3·4 (3·0) [n=246]	3·1 (2·5) [n=246]	6·0 (2·7) [n=246]	5·8 (2·5) [n=246]	72·0 (15·2) [n=248]	67·5 (17·2) [n=134]	54·0 (16·5) [n=248]
	MAU	17·8 (6·9) [n=237]	3·7 (3·0) [n=237]	3·6 (2·6) [n=237]	6·4 (2·7) [n=237]	6·1 (2·5) [n=237]	72·8 (15·6) [n=238]	68·4 (16·5) [n=123]	57·3 (17·8) [n=238]
	Difference (95% CI)	–1·13 (–2·13 to –0·13)	–0·34 (–0·79 to 0·11)	–0·51 (–0·90 to –0·12)	–0·35 (–0·74 to 0·04)	–0·31 (–0·70 to 0·08)	–1·60 (–3·85 to 0·65)	–0·64 (–2·74 to 1·46)	–3·11 (–5·83 to –0·39)
	p value	0·028	0·142	0·010	0·081	0·122	0·167	0·55	0·027
18-month follow-up	MST	16·5 (6·5) [n=232]	3·2 (3·0) [n=232]	3·1 (2·5) [n=232]	6·1 (2·5) [n=232]	5·8 (2·4) [n=232]	69·1 (16·3) [n=234]	68·6 (17·0) [n=87]	53·1 (16·3) [n=234]
	MAU	17·0 (6·9) [n=209]	3·5 (3·1) [n=209]	3·6 (2·8) [n=209]	5·9 (2·7) [n=209]	6·1 (2·5) [n=209]	70·9 (16·1) [n=217]	68·7 (16·7) [n=90]	56·6 (17·9) [n=217]
	Difference (95% CI)	–0·29 (–1·37 to 0·79)	–0·15 (–0·72 to 0·42)	–0·34 (–0·77 to 0·09)	0·10 (–0·29 to 0·49)	–0·24 (–0·61 to 0·13)	–1·06 (–3·76 to 1·64)	–0·05 (–1·95 to 1·85)	–3·00 (–5·78 to –0·22)
	p value	0·60	0·59	0·12	0·63	0·21	0·44	0·96	0·036

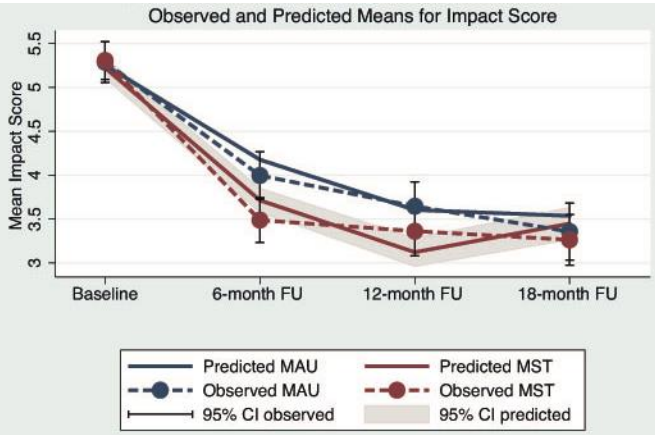
Data were obtained using the Strengths and Difficulties Questionnaire (SDQ), Conners ADHD Rating Scale – Parent and Teacher form (Conners Comprehensive Behaviour Rating Scale), and General Health Questionnaire (GHQ). MAU=management as usual. MST=Multisystemic Therapy. P=completed by parent. T=completed by teacher. There was a high proportion of missing data for the Conners ADHD (T) outcomes at both baseline and follow-up, so we used multiple imputation (without post-baseline offending data) with 30 replicates for the primary analysis of these outcomes.

Figure A11: Observed and model-predicted means (based on the explicit modelling of the temporal effects) for (a–e) parents' report of young people's wellbeing and behaviour on the SDQ, (f) parents' report of young people's behaviour on the Conners ADHD Rating Scale, and (g) parents' own wellbeing on the GHQ|.

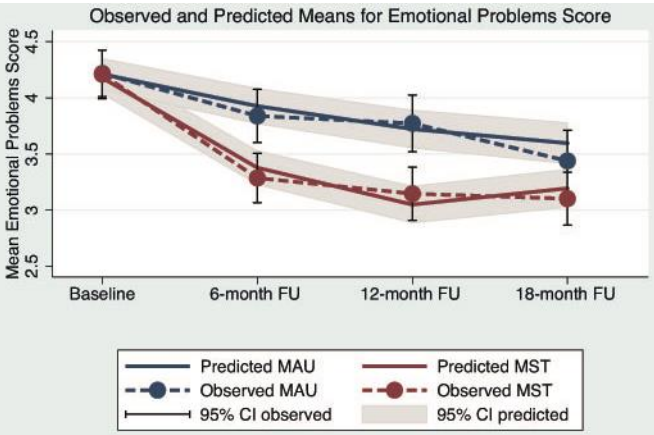
a



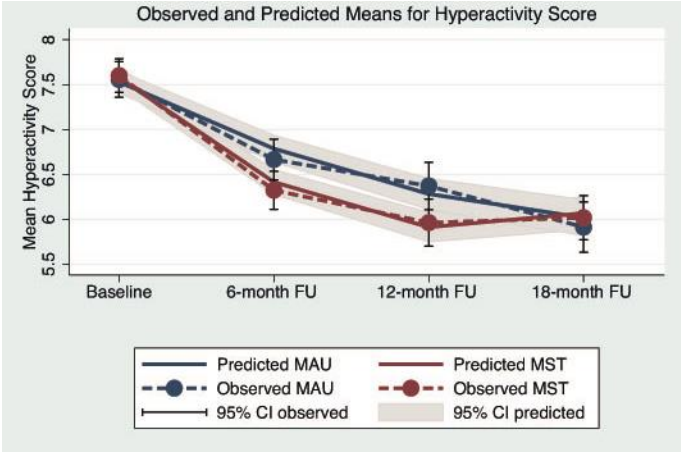
b



c



d



e

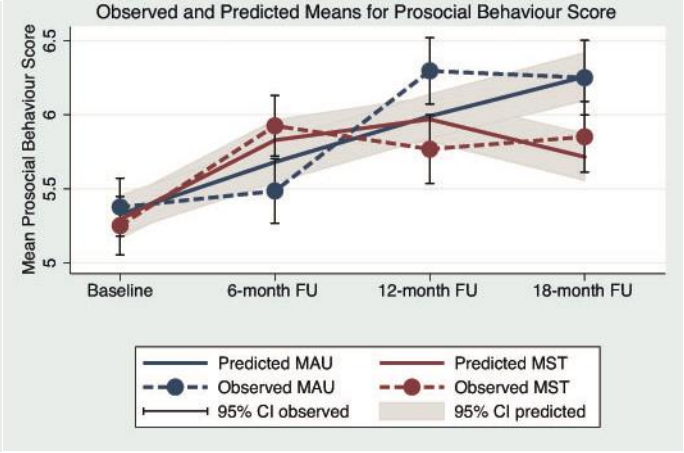
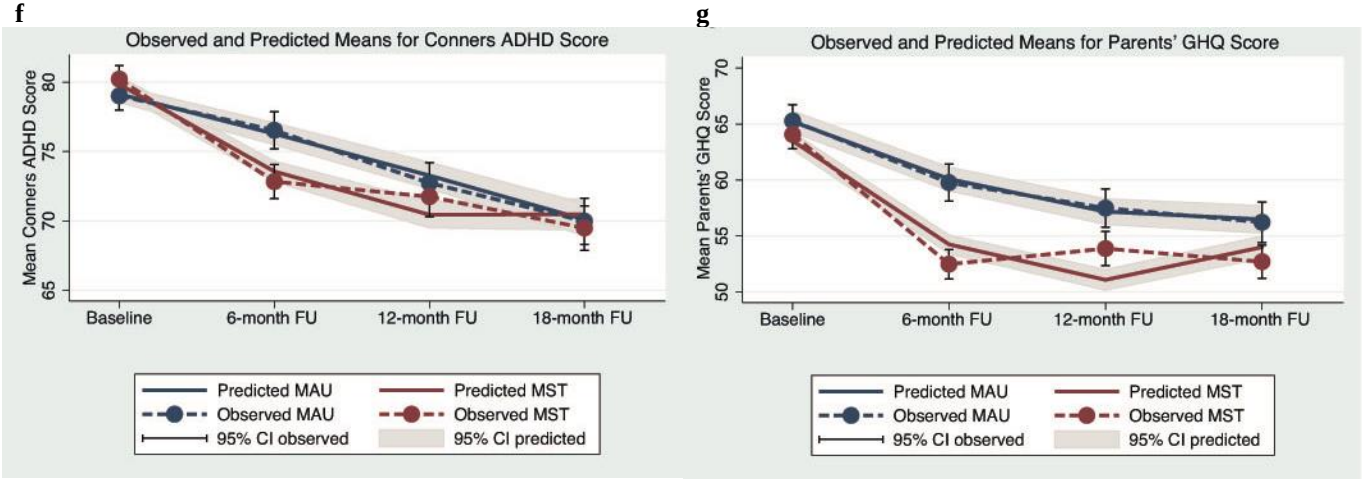


Figure A11, continued



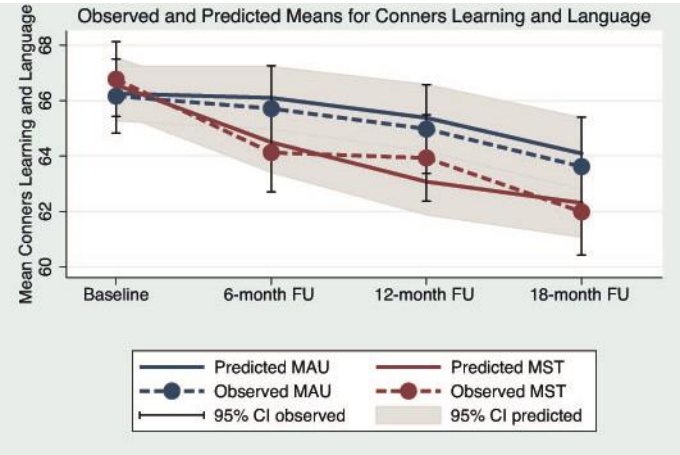
ADHD=Attention deficit and hyperactivity disorder. FU=follow-up. GHQ=General Health Questionnaire. MAU=management as usual. MST=Multisystemic Therapy. SDQ=Strengths and Difficulties Questionnaire.

Table A17. Parents' report of language and learning outcomes: observed data and estimates based on multiple imputation procedure

	Group (n) and between-group significance	Conners L&L observed Mean (SD)	Conners L&L estimated Mean (SD)
6 months–baseline	MST (n=341)	66·78 (17·25)	66·78 (17·25)
	MAU (n=339)	66·16 (16·97)	66·16 (16·97)
6-month follow-up	MST (n=292)	64·1 (16·7)	63·4 (16·8)
	MAU (n=268)	65·7 (17·3)	65·8 (17·0)
	Difference (95% CI)	–2·45 (–4·65 to –0·25)	–2·33 (–4·58 to –0·08)
	p value	0·028	0·043
12-month follow-up	MST (n=248)	63·9 (16·9)	63·8 (16·6)
	MAU (n=238)	65·1 (17·0)	65·4 (17·2)
	Difference (95% CI)	–1·82 (–4·13 to 0·49)	–1·90 (–4·13 to 0·33)
	p value	0·122	0·097
18-month follow-up	MST (n=234)	62·0 (16·5)	62·1 (16·6)
	MAU (n=217)	63·6 (17·8)	64·6 (17·4)
	Difference (95% CI)	–2·31 (–4·70 to 0·08)	–2·22 (–4·57 to 0·13)
	p value	0·058	0·065

L&L=Conners Language and Learning T-score. MAU=management as usual. MST=Multisystemic Therapy.

Figure A12: Observed and model-predicted means (based on the explicit modelling of the temporal effects) for Conners Learning and Language



FU=follow-up. MAU=management as usual. MST=Multisystemic Therapy.

Table A18: Teacher-rated education outcomes using multiple imputation with baseline educational outcomes and demographic covariates

	Group (n) and between-group significance	Disruptive Behaviour Disorder T-score	Learning & Language Disorder T-score	Mood Disorder T-score	Anxiety Disorder T-score
6-month follow-up	MST (n=150)	70·7 (21·2)	67·5 (12·9)	75·3 (21·4)	71·3 (27·8)
	MAU (n=155)	70·7 (22·0)	66·1 (13·1)	73·8 (22·2)	75·2 (25·1)
	Difference (95% CI)	0·36 (–1·95 to 2·67)	0·56 (–1·01 to 2·13)	1·13 (–1·28 to 3·54)	–1·91 (–4·44 to 0·62)
	p value	0·76	0·49	0·36	0·14
12-month follow-up	MST (n=134)	68·0 (24·5)	65·4 (13·0)	70·2 (26·1)	70·7 (28·5)
	MAU (n=123)	70·3 (22·4)	67·1 (13·1)	72·8 (23·7)	74·7 (25·6)
	Difference (95% CI)	–2·56 (–4·77 to –0·35)	–0·88 (–2·33 to 0·57)	–1·65 (–4·30 to 1·00)	–2·39 (–5·23 to 0·45)
	p value	0·025	0·24	0·22	0·10
18-month follow-up	MST (n=87)	72·4 (19·6)	67·4 (13·1)	73·7 (23·7)	75·3 (25·6)
	MAU (n=90)	70·8 (21·9)	67·0 (13·2)	73·6 (23·0)	74·1 (25·9)
	Difference (95% CI)	0·18 (–2·23 to 2·59)	0·37 (–1·02 to 1·76)	1·04 (–1·94 to 4·02)	1·70 (–1·08 to 4·48)
	p value	0·88	0·61	0·50	0·23

Data were obtained using the Conners Rating Scales-Revised (teacher report form). MAU=Management as usual. MST=Multisystemic Therapy.

Appendix iii: Economic Data Supplement

Methods

Health economic analysis was conducted by King's Health Economics at King's College London. Economic evaluation techniques were used to explore the relative costs and cost-effectiveness of the alternative management strategies—that is, MST and MAU. The evaluation took a broad perspective, including all health, social, education, and voluntary sector services, plus costs falling on the criminal justice sector, costs resulting from crimes committed, and out-of-pocket expenses to the young people and their families.

Method of economic evaluation

The a priori primary economic evaluation, as stated in the application for funding, was a cost-effectiveness analysis using the primary clinical outcome measure (out-of-home placement). In addition, a secondary analysis was proposed assessing cost-effectiveness in terms of quality-adjusted life years (QALYs), using the EQ-5D-3L measure of health-related quality of life (HRQoL).[28,29] However, an administrative error at the start of the trial meant that the EQ-5D was excluded from the outcome pack, resulting in extensive missing data (68% at baseline, 49% at 6 months, 37% at 12 months, and 33% at 18 months). The available data were too limited to have any confidence in and so this analysis had to be abandoned; no appropriate mapping studies were identified at the time the analysis was undertaken to derive QALYs from an alternative measure of outcome.[30,31]

Given that HRQoL is considered the most appropriate measure of outcome for health economic evaluations in the UK, this is an important limitation of the study. However, the EQ-5D was deliberately selected to be a secondary economic analysis because of a number of concerns with the relevance of HRQoL and the EQ-5D to the current population. Firstly, the young people in the present study cannot all be considered to be 'unwell', given the focus on antisocial behaviour, rather than necessarily a clinical diagnosis (e.g. of conduct disorder). The ability of a measure of HRQoL to capture change in such a population may therefore be limited. Secondly, at the time the study was designed (2008), there was little evidence to support the validity of the EQ-5D in mental health populations, particularly for young people.

Unit costs applied to economic data

For each participant, a unit cost was applied to each item of service use reported to calculate the total cost for the duration of the trial. The cost of the MST intervention was calculated using a standard micro-costing approach.[29,30] This involved estimation of indirect time spent on individual cases, including preparation, meetings, telephone calls and attending supervision, as well as detailed recording of the total duration of direct face-to-face contacts. A unit cost per hour of face-to-face contact between families and an MST therapist was calculated using data on salaries, employer on-costs (National Insurance and superannuation), conditions of service, and appropriate administrative, managerial, and capital overheads.[31] The cost of contributions from MST Services (the organisation licensed to disseminate MST technologies), which included MST training, provision of MST supervision and the MST licence, was provided as a total cost for all sites in the study and was allocated equally across all participants. A costing schema for MST intervention is presented in Table A19.

Table A19: Costing schema for MST intervention

Cost of MST therapists per hour	Unit cost 2012–2013	Notes
Salary plus on-costs	£47,692·00	Salaries inclusive of pension and employer's National Insurance
Overheads	£18,000·00	Comprises direct and indirect overheads
Capital overheads	£2,180·00	Based on the new-build and land requirements of NHS facilities, but adjusted to reflect shared used of both treatment and non-treatment space. Capital costs have been annuitised over 60 years at a discount rate of 3.5%.
Working time	1605 hours per year	
Face-to-face time	1:1·62	The direct:indirect ratio was based on a survey of MST therapists who took part in the trial.
Length of sessions	60 minutes	
Cost per hour	£42·00	
Cost per hour face-to-face	**£69·00**	
Cost of MST Services per client		
Training and supervision	£235·00	Per client
Licence	£27·69	Per client
Cost per client	**£262·69**	

Nationally applicable unit costs were applied to all other services, including MAU. The unit costs for education services were taken from national statistics of school income and expenditure for local authority maintained schools in England for 2011-12 and 2012-13.[32] Unit costs for hospital services were taken from the National Schedule of NHS Reference costs 2012.[33] Costs contained in the annual unit costs of health and social care publication were used to calculate costs of accommodation, community-based health, social, and voluntary services.[31] The cost of medication was calculated on the basis of averages listed in the British National Formulary[34] for the generic drug and using daily dose information collected using the CA-SUS. Unit costs for criminal justice services were taken from the unit costs in criminal justice publication[35] and reports from the Home Office on the cost of criminal justice.[36-38] Out-of-pocket expenses were excluded from the analysis as a result of the poor quality of reporting. Only 23% of the sample reported out-of-pocket expenses and only 20% of these (n=31) provided adequate data to enable these expenses to be costed.

Cost-effectiveness analysis

Cost-effectiveness was explored using incremental cost-effectiveness ratios (ICER), that is, the difference in mean cost divided by the difference in mean effect,[33] with effects measured in terms of the proportion of participants requiring out of home placement. Statistical uncertainty of the ICER was accounted for by generating 1000 bootstrapped resamples and these were then used to calculate the probability that MST is the optimal choice, for different values a decision-maker may be willing to pay for a unit improvement in outcome (the ceiling ratio, λ). Cost-effectiveness acceptability curves (Figure A13) are generated by plotting these probabilities for a range of possible values of λ to explore the uncertainty that exists around estimates of mean costs and effects, and to show the probability that MST is cost-effective compared with MAU.

Complete case analysis was used for the economic evaluation and controlled for the following covariates: treatment centre, number of past convictions, sex, age at onset of criminal behaviour, and baseline measurement of the variables of interest. Additionally, data were truncated to exclude influential outliers, that is, cases with total costs in the 99th percentile that make a significant difference to the results.[42]

Results

A summary of service use over the 18-month follow-up period is provided in Tables A20 and A21.

Table A20: Service use (unit) over 18 months of follow-up

	MAU (n=209)			MST (n=226)		
	Mean (SD)	**Range**	**% using**	**Mean (SD)**	**Range**	**% using**
MST						
MST (hours of direct contact)	0 (0)	0–0	0	35·79 (24·11)	0–114	81
Accommodation						
Foster care (days)	1·90 (11·08)	0–90	4	3·84 (23·37)	0–278	6
Residential care (days)	3·45 (24·13)	0–233	3	1·56 (13·35)	0–166	2
Staffed accommodation (days)	0·75 (6·24)	0–60	1	0·14 (2·06)	0–31	0·4
Other (days)	2·51 (17·83)	0–176	3	0·91 (8·06)	0–90	2
Education						
Mainstream school (hours)	1005·64 (962·81)	0–3035	71	1082·47 (965·77)	0–3088	72
Specialist school (hours)	223·15 (493·96)	0–2470	27	235·37 (515·66)	0–2535	26
Residential school (hours)	5·26 (49·88)	0–630	1	5·91 (47·13)	0–550	2
Hospital school (hours)	2·90 (29·28)	0–390	1	0	0	0
Pupil Referral Unit (hours)	137·18 (287·02)	0–1430	27	192·90 (369·80)	0–1820	32
Home tuition (hours)	13·90 (74·49)	0–780	6	17·40 (108·96)	0–1430	7
Further education (hours)	174·24 (364·29)	0–2028	32	156·43 (367·20)	0–2080	28
Secondary health care						
Inpatient stay (nights)	1·00 (7·95)	0–109	13	0·44 (3·09)	0–44	10
Outpatient appointments (contacts)	1·03 (2·63)	0–20	28	1·12 (3·12)	0–20	27
Accident and emergency (contacts)	1·66 (4·92)	0–58	50	0·99 (1·85)	0–17	46
Community based						
Counsellor (contacts)	1·52 (5·63)	0–39	14	1·23 (6·59)	0–63	9
Family therapist (contacts)	0·90 (3·57)	0–30	11	0·51 (4·11)	0–50	3
Art/drama/music/occupational therapy (contacts)	0·23 (1·92)	0–26	3	0·04 (0·48)	0–6	1
Social worker (contacts)	7·95 (12·95)	0–64	58	6·93 (15·40)	0–117	42
Family support worker (contacts)	5·08 (15·75)	0–130	23	2·66 (12·45)	0–140	14
Social services youth worker (contacts)	1·08 (4·94)	0–50	10	0·37 (2·63)	0–28	4
Accommodation key worker (contacts)	0·44 (2·80)	0–30	33	0·54 (3·10)	0–26	5
Educational psychologist (contacts)	0·41 (2·57)	0–26	7	0·39 (2·52)	0–26	8
Education welfare officer (contacts)	0·59 (2·69)	0–24	12	1·88 (9·81)	0–98	16
Connexions worker (contacts)	2·81 (8·92)	0–78	27	1·56 (6·51)	0–78	25
Mentor (contacts)	5·61 (20·15)	0–150	18	7·15 (29·17)	0–206	15
Drug/alcohol support worker (contacts)	1·38 (5·76)	0–53	12	1·69 (6·70)	0–52	10
Advice service, eg, Citizens' Advice Bureau, housing association, careers advice (contacts)	0·02 (0·21)	0–2	1	0·06 (0·60)	0–8	1
Helpline (contacts)	0	0	0	0·02 (0·16)	0–2	1
Complementary therapist (contacts)	0·44 (6·23)	0–90	1	0·01 (0·15)	0–2	1
FIP (contacts)	1·28 (6·05)	0–42	6	0·35 (4·26)	0–62	1
Other (contacts)	10·21 (29·21)	0–234	34	4·67 (16·24)	0–182	27
Criminal justice system						
Police custody (days)	0·91(3·51)	0–44	25	0·50 (1·36)	0–9	20
Youth custody (days)	2·67 (18·27)	0–197	6	2·71 (17·74)	0–150	5
Probation officer (contacts)	1·15 (7·30)	0–80	6	0·04 (0·36)	0–5	2
Youth offending team worker (contacts)	13·29 (27·44)	0–154	39	10·00 (21·61)	0–152	35
Police (contacts)	11·76 (52·53)	0–675	72	8·10 (18·94)	0–133	62
Solicitor (contacts)	1·51 (4·16)	0–39	27	0·97 (2·88)	0–31	28
Court appearance as victim (number)	0·02 (0·15)	0–1	2	0·04 (0·36)	0–5	2
Court appearance as defendant (number)	0·43 (1·32)	0–8	16	0·30 (0·91)	0–8	16

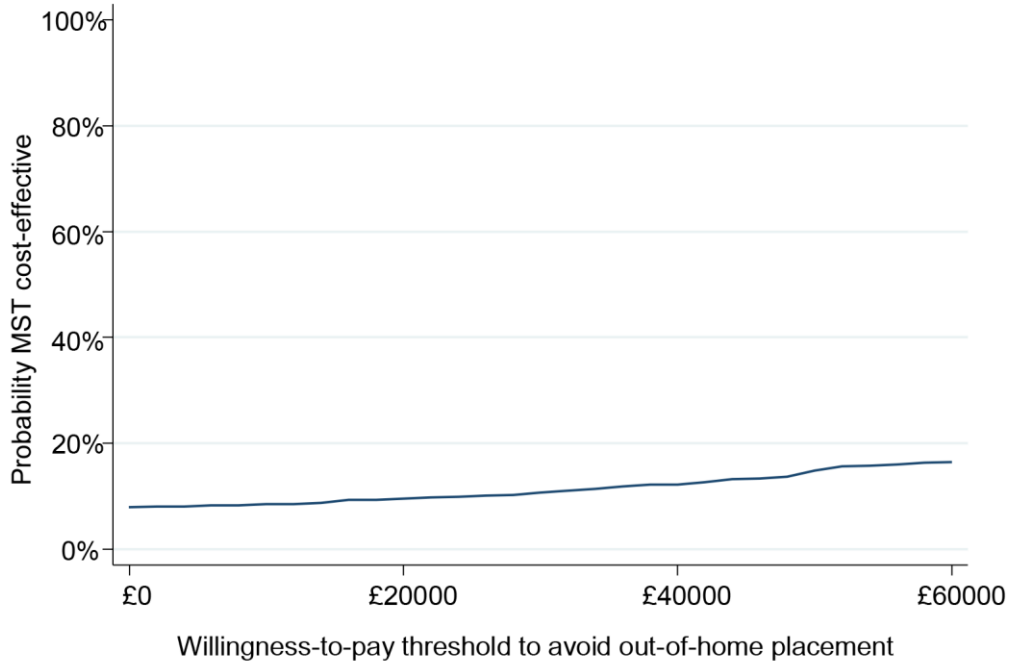
Table A21: Use of medication, baseline to 18-month follow-up

Type of medication	MAU (n=209)	MST (n=226)
Antidepressants	6	5
ADHD	12	11
Benzodiazepines	0	0
Sleep disturbance	6	3
Antipsychotics	3	1
Antiepileptics	1	0

Data are presented as the percentage of the sample using each type of medication. ADHD=attention deficit hyperactivity disorder. MAU=management as usual. MST=Multisystemic Therapy.

The cost-effectiveness acceptability curve showing the probability of MST being cost-effective compared with MAU is low and does not rise above 18% for a range of willingness-to-pay thresholds (Figure A13).

Figure A13: Cost-effectiveness acceptability curve showing the probability that MST is cost-effective compared with MAU for different values of willingness to pay to avoid out-of-home placement



Sensitivity analysis
The impact of missing data, considered using multiple imputation by chained equations, was explored as a sensitivity analysis. The results are presented in Table A22.

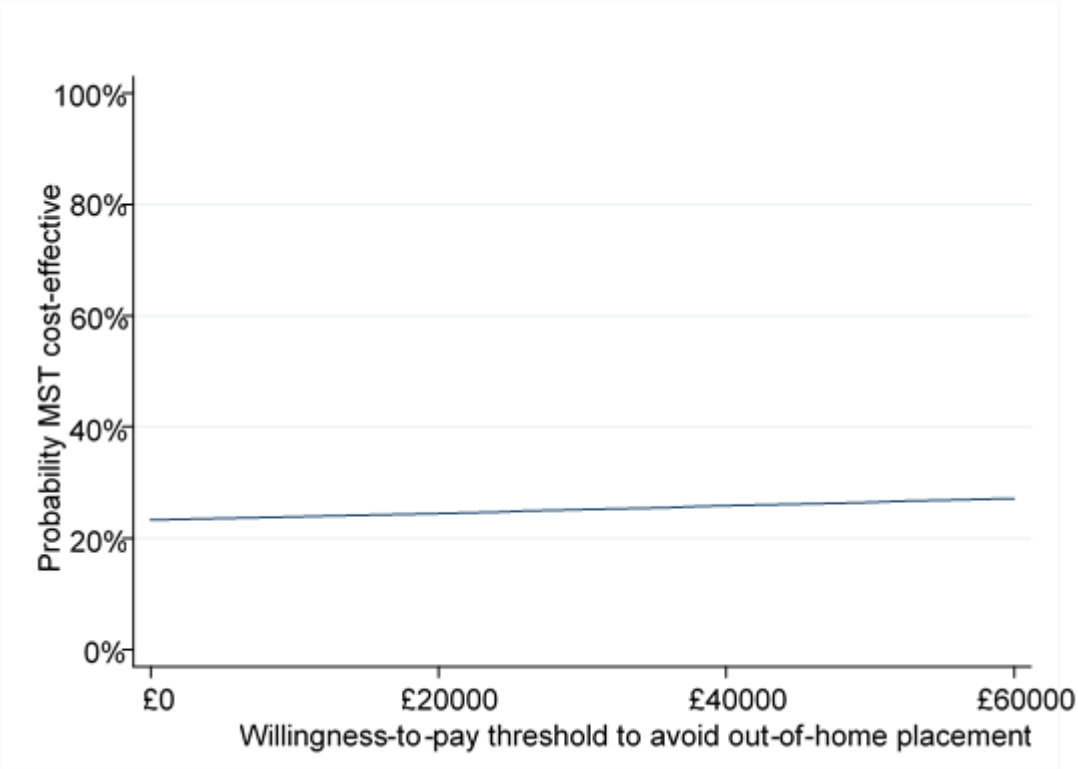
Table A22: Differences in costs per participant over the 18-month follow-up period

	MST Mean (SD)	MAU Mean (SD)	Mean difference*	95% CI*	p value*
Main analysis	(n=226)	(n=209)			
Total costs	28678·32 (34175·21)	30927·68 (36106·37)	–1622·94	–7684·45 to 4438·57	0·60
Out-of-home placement	9·73	8·17	1·56		
Sensitivity analysis	(n=342)	(n=341)			
Total costs	38105·35 (7486·00)	46169·83 (10706·92)	7534·93	–13542·04 to 28611·91	0·48
Out-of-home placement	12·62	10·79	1·83		

MAU=management as usual. MST=Multisystemic Therapy. *Adjusted for stratification variable

Imputation of missing data increased the difference in total cost between the trial arms (£8064.48 compared to £2249.36 in the main analysis) but remained higher for MAU (£46169.83 compared with £38105.35 for MST). Figure A14 shows the cost-effectiveness acceptability curve for the sensitivity analysis, which supports that the probability of MST being cost-effective compared with MAU remains low and does not rise above 28% for a range of willingness-to-pay thresholds.

Figure A14: Cost-effectiveness acceptability curve for the sensitivity analysis showing the probability that MST is cost-effective compared with MAU for different values of willingness to pay to avoid out-of-home placement



References

1. Andrews DA, Bonta J. Rehabilitating criminal justice policy and practice. *Psychol Public Policy Law* 2010; **16**(1): 39-55.
2. Henggeler SW, Schoenwald SK, Borduin CM, Rowland MD, Cunningham PB. Multisystemic treatment of antisocial behavior in children and adolescents. New York: Guilford Press; 1998.
3. Huey SJ, Jr., Henggeler SW, Brondino MJ, Pickrel SG. Mechanisms of change in multisystemic therapy: reducing delinquent behavior through therapist adherence and improved family and peer functioning. *J Consult Clin Psychol* 2000; **68**(3): 451-67.
4. Clark LA, Watson D. Constructing validity: basic issues in objective scale development. *Psychol Assess* 1995; **7**(3): 309-19.
5. Youth Justice Board. Annual Statistics 2005/06. London: HMSO, 2007.
6. Butler S, Baruch G, Hickey N, Fonagy P. A randomized controlled trial of multisystemic therapy and a statutory therapeutic intervention for young offenders. *J Am Acad Child Adolesc Psychiatry* 2011; **50**(12): 1220-35.
7. Barrett B, Byford S, Chitsabesan P, Kenning C. Mental health provision for young offenders: service use and cost. *Br J Psychiatry* 2006; **188**(6): 541-6.
8. Byford S, Barrett B, Roberts C, et al. Cost-effectiveness of selective serotonin reuptake inhibitors and routine specialist care with and without cognitive-behavioural therapy in adolescents with major depression. *Br J Psychiatry* 2007; **191**(6): 521-7.
9. Clark A, O'Malley A, Woodham A, Barrett B, Byford S. Children with complex mental health problems: needs, costs and predictors over one year. *Child Adolesc Ment Health* 2005; **10**(4): 170-8.
10. Harrington R, Peters S, Green J, Byford S, Woods J, McGowan R. Randomised comparison of the effectiveness and costs of community and hospital based mental health services for children with behavioural disorders. *BMJ* 2000; **321**(7268): 1047-50.
11. Smith DJ, McVie S. Theory and method in the Edinburgh study of youth transitions and crime. *Br J Criminol* 2003; **43**(1): 169-95.
12. Butler SM, Leschied AW, Fearon P. Antisocial beliefs and attitudes in pre-adolescent and adolescent youth: the development of the antisocial beliefs and attitudes scales (ABAS). *J Youth Adolesc* 2007; **36**(8): 1058-71.
13. Mann BJ, Borduin CM, Henggeler SW, Blaske DM. An investigation of systemic conceptualizations of parent-child coalitions and symptom change. *J Consult Clin Psychol* 1990; **58**(3): 336-44.
14. Henggeler SW, Rodick JD, Hanson CL, Watson SM, Borduin CM, Urey JR. Multisystemic treatment of juvenile offenders: effects on adolescent behavior and family interaction. *Dev Psychol* 1986; **22**(1): 132-41.
15. Henggeler SW, Letourneau EJ, Chapman JE, Borduin CM, Schewe PA, McCart MR. Mediators of change for multisystemic therapy with juvenile sexual offenders. *J Consult Clin Psychol* 2009; **77**(3): 451-62.
16. Olson D. FACES IV and the Circumplex Model: validation study. *J Marital Fam Ther* 2011; **37**(1): 64-80.
17. Scott S, Briskman J, Dadds MR. Measuring parenting in community and public health research using brief child and parent reports. *J Child Fam Stud* 2011; **20**(3): 343-52.
18. Jaffee SR, Moffitt TE, Caspi A, Taylor A, Arseneault L. Influence of adult domestic violence on children's internalizing and externalizing problems: an environmentally informative twin study. *J Am Acad Child Adolesc Psychiatry* 2002; **41**(9): 1095-103.
19. Gerlsma C, Hale WW, 3rd. Predictive power and construct validity of the Level of Expressed Emotion (LEE) scale. Depressed out-patients and couples from the general community. *Br J Psychiatry* 1997; **170**(6): 520-5.
20. Hale WW, 3rd, Raaijmakers QAW, Gerlsma C, Meeus W. Does the Level of Expressed Emotion (LEE) questionnaire have the same factor structure for adolescents as it has for adults? *Soc Psychiatry Psychiatr Epidemiol* 2007; **42**(3): 215-20.
21. Goodman R, Scott S. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? *J Abnorm Child Psychol* 1999; **27**(1): 17-24.
22. Messer SC, Angold A, Costello EJ, Loeber R, van Kammen W, Stouthamer-Loeber M. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents: factor composition and structure across development. *Int J Methods Psychiatr Res* 1995; **5**(4): 251-62.
23. Goldberg DP, Gater R, Sartorius N, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Med* 1997; **27**(1): 191-7.
24. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: Description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry* 2000; **41**(5): 645-55.
25. Wechsler D. Wechsler Abbreviated Scale of Intelligence (WASI). Oxford: Harcourt Assessment; 1999.

26. Essau CA, Sasagawa S, Frick PJ. Callous-unemotional traits in a community sample of adolescents. *Assessment* 2006; **13**(4): 454-69.
27. Conners CK. Conners' Rating Scales-Revised: technical manual. Toronto: Multi-Health Systems; 1995.
28. Brooks R. EuroQol: the current state of play. *Health Policy* 1996; **37**(1): 53-72.
29. Byford S. The validity and responsiveness of the EQ-5D measure of health-related quality of life in an adolescent population with persistent major depression. *J Ment Health* 2013; **22**(2): 101-10.
30. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. *Health and Quality of Life Outcomes* 2013; **11**: 151.
31. HERC. HERC database of mapping studies, version 5.0. May 16 2016. http://www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies (accessed September 4 2017).
32. Beecham JK, Knapp MRJ. Costing psychiatric interventions. In: Thornicroft G, Brewin C, Wing JK, eds. Measuring mental health needs. London: Gaskell/Royal College of Psychiatrists; 1992: 163-83.
33. Drummond MF, Sculpher J, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2015.
34. Curtis L. Unit costs of health and social care 2013. Canterbury: PSSRU, University of Kent, 2013.
35. Department for Education. Detailed school income and expenditure statistics for local authority maintained schools in England by phase of education, 2011-12 and 2012-13. London: Department for Education, 2013.
36. Department of Health. NHS Reference Costs 2012-13. London: Department of Health, 2013.
37. Royal Pharmaceutical Society of Great Britain. British National Formulary. 2016. https://www.medicinescomplete.com/mc/bnf/current/.
38. Brookes N, Barrett B, Netten A, Knapp E. Unit costs in criminal justice (UCCJ). Canterbury: PSSRU, University of Kent, 2013.
39. Harries R. The cost of criminal justice. Home Office Research Findings No. 103. London: Home Office, 1999.
40. Brand S, Price R. The economic and social costs of crime. Home Office Research Study 217. London: Home Office, 2000.
41. Home Office. The economic and social costs of crime against individuals and households 2003/04. London: Home Office, 2005.
42. Weichle T, Hynes DM, Durazo-Arvizu R, Tarlov E, Zhang Q. Impact of alternative approaches to assess outlying and influential observations on health care costs. *SpringerPlus* 2013; **2**: 614.