BMJ Mental Health

CHILD AND ADOLESCENT MENTAL HEALTH

Sex differences in internalising and externalising symptom patterns, pathway and persistence in adolescents receiving psychiatric care: a 2-year follow-up of the MILESTONE European cohort

Miriam D'Addazio ,[1] Silvia Leone ,[1] Anna Caterina Leucci ,[2] Marta Magno ,[1] Anna Rita Atti ,[2] Stefano Calza ,[3] Martina Carnevale ,[3] Elisa Caselani ,[1] Laura Iozzino ,[1] Federica Marcolini ,[2] Donato Martella ,[1,4] Samuele Cortese ,[5,6,7,8,9] Gwendolyn Dieleman ,[10] Tomislav Franić ,[11] Athanasios Maras ,[12] Fiona McNicholas ,[13,14,15] Diane Purper-Ouakil ,[16,17] Paramala Santosh ,[18,19,20] Ulrike M E Schulze ,[21] Cathy Street ,[22] Swaran Preet Singh ,[23] Sabine Tremmery ,[24] Helena Tuomainen ,[25] Larissa van Bodegom ,[10] Dieter Wolke ,[25,26] Stefano Vicari ,[27,28] Giovanni de Girolamo ,[1] ON behalf of MILESTONE consortium[29]

ABSTRACT

Background Sex differences in adolescent mental disorders are well documented, but less is known about how these differences evolve during the transition to adulthood, particularly during the shift from Child and Adolescent Mental Health Services (CAMHS) to Adult Mental Health Services (AMHS), a critical period marked by vulnerability to discontinuity of care.

Objective To examine sex-specific differences in psychiatric symptom profiles and clinical trajectories during the CAMHS–AMHS transition using data from the European MILESTONE project, a 2-year longitudinal study.

Methods A cohort of 1004 adolescents (aged 17–19) in CAMHS was assessed at baseline (T1), 9 months (T2), 15 months (T3) and 24 months (T4). Measures included the Child Behavior Checklist/Adult Behavior Checklist, Health of the Nation Outcome Scales for Children and Adolescents, and the Specific Level of Functioning Scale. Multilevel modelling included sex, diagnosis and time point as covariates, testing three-way interactions for differential trajectories.

Findings Males showed higher symptom severity and impairment across measures, particularly in anxiety/somatic/trauma, eating disorder/obsessive-compulsive disorder (ED/OCD) and schizophrenia. Females exhibited better functioning. Significant sex differences emerged in internalising symptoms (eg, anxiety/somatic/trauma, autism spectrum disorder (ASD)), externalising symptoms (eg, attention deficit hyperactivity disorder (ADHD)), overall psychopathology (eg, ADHD, personality disorder/conduct disorder/substance use disorder) and functioning (eg, ED/OCD, schizophrenia spectrum disorder).

Conclusions Sex-related differences in symptom severity, diagnosis and functioning persist across the CAMHS–AMHS transition, with males generally more impaired and females showing better adaptive outcomes.

Clinical implications Identifying sex-specific trajectories can enhance personalised transitional care, reduce misdiagnosis and guide targeted interventions—especially for under-recognised presentations such as ADHD/ASD in females and ED in males.

Trial registration number ISRCTN83240263.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Sex differences in adolescent mental disorders are well documented, but little is known about how these differences evolve during the transition from *Child and Adolescent Mental Health Services* to *Adult Mental Health Services*.

WHAT THIS STUDY ADDS

⇒ Using longitudinal data from the European MILESTONE project, we show that males generally present greater symptom severity and functional impairment, whereas females demonstrate higher internalising symptoms but better psychosocial functioning, and these differences persist over time.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Recognition of sex-specific trajectories can inform personalised transitional care, reduce misdiagnosis and guide the development of targeted interventions for vulnerable subgroups.

INTRODUCTION

Sex differences in adolescent mental disorders are well established across diagnostic categories, affecting symptom presentation, prevalence,

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjment-2025-302260).

For numbered affiliations see end of article.

Correspondence to
Professor Stefano Vicari;
stefano.vicari@opbg.net

MD'A and SL are joint first authors.

Received 27 November 2025
Accepted 22 June 2026

To cite: D'Addazio M, Leone S, Leucci AC, *et al*. *BMJ Ment Health* 2026;**29**:1–10.

age of onset, functional impairment, prognosis and treatment response.[1 2] However, less is known about how these differences evolve during the transition from adolescence to early adulthood, a critical phase for young people moving from Child and Adolescent Mental Health Services (CAMHS) to Adult Mental Health Services (AMHS), where discontinuity of care may increase vulnerability.[3 4]

Evidence from epidemiological and clinical studies consistently shows higher rates of internalising disorders—including depression, anxiety, eating disorder (ED) and somatic symptoms—among adolescent females, whereas males more frequently present with externalising behaviours such as aggression, conduct problems and substance misuse.[1 5 6] Female adolescents also appear more vulnerable to mood, trauma-related and somatic disorders, often showing earlier onset, recurrent depressive episodes, comorbid EDs and poorer quality of life.[7 8] EDs are markedly more prevalent in girls than boys by early adulthood,[9] while females may also report somatic symptoms more frequently, potentially reflecting differences in stress responsivity and help-seeking behaviours.[8]

Sex differences extend to trauma-related and neurodevelopmental conditions. Females are more likely to develop dissociative symptoms following trauma exposure,[10] while bipolar disorder (BD) in females is more commonly associated with depressive episodes and non-suicidal self-injury, compared with more manic presentations in males.[10] Attention deficit hyperactivity disorder (ADHD) is diagnosed more frequently in males, although lower female representation may partly reflect under-recognition or diagnostic bias.[2 11] Similar concerns have been raised for autism spectrum disorder (ASD), which is diagnosed more often in males, despite evidence that females may use camouflaging or compensatory strategies that contribute to under-recognition.[12 13] Substance use disorder (SUD), conduct disorder (CD) and schizophrenia spectrum disorder (SSD) also remain more prevalent in males,[14 15] whereas findings for personality disorder (PD) and obsessive-compulsive disorder (OCD) are less consistent.[16 17]

Beyond prevalence, sex differences may also influence developmental trajectories. Anxiety symptoms tend to decline in females during adolescence but remain more stable in males, whereas depression often decreases in males but may persist or re-emerge in females.[7] Externalising behaviours generally lessen over time, particularly in males, while SSD may show poorer outcomes in males.[18] Understanding these longitudinal patterns is particularly relevant during the transition from CAMHS to AMHS, a period associated with substantial changes in care pathways and psychosocial functioning.[6]

The MILESTONE project (2014–2019), conducted across eight European countries, provided a unique opportunity to investigate sex differences in psychiatric symptoms and outcomes during this transition period.[6] Using a longitudinal design with a 2-year follow-up, the present study aimed to examine sex-specific symptom profiles, psychosocial functioning and clinical trajectories across diagnostic categories in a large European clinical cohort.

Objectives

The present study examines sex differences in psychopathological presentation and outcomes over a 2-year follow-up in a large European clinical cohort of adolescents transitioning from CAMHS to AMHS. Specifically, the objectives are to (1) examine sex differences in self-reported internalising and externalising symptom patterns across major diagnostic groups and (2) integrate sex differences in overall psychopathology severity and psychosocial functioning over time, as assessed by clinician-reported and caregiver-reported measures.

In psychiatric research, gender refers to sociocultural roles and identities, while sex refers to biological characteristics. Although both influence psychiatric disorders,[19] gender identity data were not systematically collected in the 2015–2016 fieldwork; therefore, sex is used throughout this manuscript.

METHODS

Study design and participant characteristics

The MILESTONE study examined service utilisation, mental health and related outcomes over 2 years in 1004 young people (17–19 years) transitioning out of CAMHS across eight European countries. A detailed description of study methods, eligibility criteria, data collection and assessment procedures is available in the full protocol and previous publications.[6] In this multicentre, multinational prospective cohort study, all available observations were analysed: 564 participants for the Child Behavior Checklist (CBCL) and Adult Behavior Checklist (ABCL) within the Achenbach System of Empirically Based Assessment (ASEBA),[20 21] 759 for the Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA)[22] and 513 for the Specific Level of Functioning Scale (SLOF),[23] with missing assessments excluded.

Participants completed four assessments over 24 months: T1 (3 months pre-transition from CAMHS to AMHS), T2 (9 months post-T1), T3 (15 months post-T1) and T4 (24 months post-T1). Baseline clinical diagnoses (Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV), DSM-5, International Classification of Diseases-10th Revision) were obtained from CAMHS records or clinicians and only baseline classifications were used. Participants were initially grouped into 14 diagnostic categories, which were subsequently reduced to seven clinically coherent clusters to improve analytical feasibility and statistical robustness. This regrouping reflected established overlap in symptomatology, mechanisms and treatment approaches across disorders. Anxiety, trauma-stress and somatic symptom disorders were combined into an ‘anxiety-related’ cluster; depression and BD were merged as affective disorders; ED and OCD were grouped based on shared compulsive and rigid cognitive features; and CD, PD and SUD were combined as externalising conditions. ADHD, ASD and SSD were retained as distinct categories due to their specific clinical profiles. This approach reduced heterogeneity while preserving clinical interpretability. Participants could present comorbid diagnoses, but only the primary diagnosis was considered.

Interpretation of diagnosis-specific trajectories warrants caution. Diagnostic categories are fixed at baseline, regrouped into broad clusters and derived from routine clinical diagnoses obtained from CAMHS records or treating clinicians using heterogeneous classification systems. No standardised diagnostic assessment or systematic diagnostic revision during follow-up is reported.

Clinical characteristics

Assessment of internalising and externalising spectrum

Internalising and externalising symptoms were assessed using the CBCL and ABCL, depending on the adolescents’ age at the time of assessment, and were self-completed by the participants as part of the ASEBA measures.[20 21] These instruments include 113 items rated on a 3-point ordinal response format (0–2), with higher scores indicating greater symptom severity. Items are grouped into eight syndrome subscales, forming internalising (anxious/

depressed, withdrawn/depressed, somatic complaints) and externalising (rule-breaking, aggressive behaviour) dimensions.

Assessment of general health and social functioning

The HoNOSCA (range 0–52) is a clinician-rated tool assessing adolescent mental health across behavioural, emotional, social and school/interpersonal domains; higher scores indicate greater psychopathology.[22] The SLOF (range 43–215) is a parent-reported measure of daily functioning, including physical abilities, personal care, social skills, interpersonal relationships, social awareness and daily activities; higher scores reflect better overall functioning and competence.[23]

Statistical analysis

Sociodemographic and clinical characteristics were described using mean and SD for quantitative variables and counts/percentages for qualitative variables. Baseline characteristics are presented for each outcome separately to maximise data utilisation for each clinical scale (ASEBA, HoNOSCA and SLOF). CBCL and ABCL raw scores were analysed longitudinally to capture symptom patterns across the transition to adulthood, acknowledging that these instruments, while conceptually aligned within the ASEBA system, are not metrically identical. This decision was primarily based on the analytical objectives of our work, which focused on modelling symptom trajectories and functional outcomes over time within a clinical population, rather than comparing individuals or groups to a normative reference sample. While T-scores offer the advantage of standardisation and facilitate interpretation against normative thresholds, they are derived from general population norms that may not reflect the characteristics of clinical samples, particularly in terms of symptom severity and diagnostic complexity. The use of raw scores allows for greater sensitivity in capturing within-sample variation, which is critical in longitudinal models that assess changes over time and interactions between sex, diagnosis and follow-up. Due to the longitudinal nature of the study, a multilevel approach was adopted to analyse four outcomes: ASEBA internalising problems (IP), ASEBA externalising problems (EP), HoNOSCA and SLOF total scores (at T2, all the observations were missing), with random intercepts at patient and country levels to account for within-individual and cross-country variability. Sex, diagnosis and time point were included as covariates, with a three-way interaction (sex×diagnosis×time point) to capture differential symptom trajectories. Given the distributional characteristics of the scales, ASEBA IP/EP, HoNOSCA and SLOF scores were rescaled to (0–1) and modelled using ordered-beta hierarchical regression.[24] The Clinical Global Impression (CGI) scale, being an ordinal measure, was modelled with a cumulative link mixed model (probit).[25] The best model was selected using likelihood ratio tests (LRT) for term addition. P values were validated using parametric bootstrap (B=200).

We evaluate a potential differential attrition among relevant clinical groups (gender, diagnosis, country) using linear models with logit-transformed proportions, modelling the percentage of response (number of responses over the number of enrolled patients at baseline) along time and its interaction with the group levels. The LRT p value for the interaction term was used to evaluate the statistical significance of a potential differential trend.

To account for the presence of missing values in score measurements, we performed a sensitivity analysis (see online supplemental table 9S), performing the same analysis adopted for complete data on a multiple imputed dataset. We used multiple imputation by chained equations (MICE) using *mice* and *miceadds* R packages. Multiple imputation was performed using a two-level algorithm using subject as clustering factor, using the predictive mean matching method, with 55 imputations and 15 iterations at max. The imputation procedure included the main clinical characteristics such as gender, age, diagnosis, centre and visit timing, in addition to HoNOSCA, SLOF, CBCL and CGI.

Moreover, we fitted a pattern-mixture model for each score. As the missing pattern is intermittent, we had overall 12 different patterns, some with very little observations. Therefore, we used only those missing patterns that had at least 20 observations and classified the remaining as 'other'. The pattern effect tested using an LRT is reported in online supplemental table 15S. If a marginal effect was present, we also tested a potential interaction with gender, which is the main focus of the paper. Please see online supplemental table 11S for the number of patients at each time point, as well as the same information stratified by diagnosis (online supplemental table 12S), gender (online supplemental table 13S) and country (online supplemental table 14S).

Results are reported as contrast estimates and corresponding 95% CIs. Individual group (diagnosis and gender at different times) estimates from the models are reported in the original scale via a back transformation. All analyses were conducted using R software V.4.5.2, assuming a 5% two-sided significance level. Models were fitted using generalised linear mixed models using Template Model Builder (*glmmTM*) and *ordinal* packages.

RESULTS

Sociodemographic and clinical characteristics of the sample

Table 1 presents sociodemographic and clinical characteristics for ASEBA (n=564), HoNOSCA (n=759) and SLOF (n=513). Males predominated in ASEBA (60.5%) and HoNOSCA (61.3%), while females in SLOF (61.0%). Italy contributed the largest proportion of participants across all measures. Anxiety-related disorders and depression/BD were the most common diagnoses, SSD the least frequent. Mean scores were ASEBA EP 10.4 (SD=9.4) and ASEBA IP 17.6 (SD=10.6); HoNOSCA 13.3 (SD=7.5); and SLOF 188.4 (SD=19.5).

In the overall study, we observed an overall 25% attrition rate along the follow-up. In order to evaluate a potential informative attrition, we modelled the number of evaluated patients at each time point as a function of time and relevant clinical groups. Both gender (p=0.58) and diagnosis (p=0.17) showed a non-significant trend in attrition stratified by group levels. On the other hand, there was a significant effect of country, with a significant interaction term (p=0.006), suggesting a differential loss to follow-up in different centres.

Multilevel models

Table 2 shows the analysis of variance table based on LRT, with Wald p values for variable effects on the clinical scores. Model selection was performed based on LRT and Akaike information criterion; terms not included in the model are not reported in the table. Online supplemental tables 1S–4S show total scores for ASEBA IP/EP, HoNOSCA and SLOF by sex and diagnosis, including median, minimum and maximum values. Self-reported ASEBA IP predicted scores (figure 1, online supplemental table 5S) were higher in females across all diagnostic groups, especially at baseline, with marked differences in ADHD (T1: F 16.17, 95% CI 12.98 to 19.36 vs M 12.99, 95% CI 11.01 to 14.98), ASD (T1: F 24.48, 95% CI 21.01 to 27.95 vs M 17.86, 95% CI 15.40 to 20.32), depression/bipolar (T1: F 19.72, 95% CI 17.80 to 21.65 vs M 17.72, 95% CI 14.32 to 21.12) and PD/CD/SUD (T1: F 20.15, 95% CI 16.95 to 23.35 vs M 15.55, 95% CI 12.02

Table 1 Sociodemographic and clinical characteristics of the sample at baseline

Variable	ASEBA (n=564)	HoNOSCA (n=759)	SLOF (n=513)
	Mean (SD), n (%)		
Sex, n (%)			
Male	341 (60.5%)	465 (61.3%)	200 (39.0%)
Female	223 (39.5%)	294 (38.7%)	313 (61.0%)
Country, n (%)			
Belgium	60 (10.6%)	88 (11.6%)	58 (11.3%)
Croatia	47 (8.3%)	49 (6.5%)	47 (9.2%)
France	69 (12.2%)	67 (8.8%)	60 (11.7%)
Germany	47 (8.3%)	88 (11.6%)	38 (7.4%)
Ireland	31 (5.6%)	43 (5.7%)	16 (3.1%)
Italy	142 (25.2%)	162 (21.3%)	135 (26.3%)
Netherlands	82 (14.5%)	96 (12.6%)	83 (16.2%)
UK London	13 (2.4%)	64 (8.5%)	11 (2.1%)
UK West Midlands	73 (12.9%)	102 (13.4%)	65 (12.7)
Diagnosis, n (%)			
ADHD	98 (17.4%)	110 (14.5%)	85 (16.6%)
Anxiety/somatic/trauma	141 (25.0%)	182 (24.0%)	126 (24.6%)
ASD	78 (13.8%)	97 (12.8%)	73 (14.2%)
Depression/bipolar disorder	109 (19.3%)	168 (22.1%)	103 (20.1%)
ED/OCD	70 (12.4%)	96 (12.6%)	61 (11.9%)
PD/CD/SUD	49 (8.7%)	81 (10.7%)	46 (9.0%)
SSD	19 (3.4%)	25 (3.3%)	19 (3.6%)
ABCL-CBCL externalising	10.4 (9.4)	–	–
ABCL-CBCL internalising	17.6 (10.6)	–	–
HoNOSCA	–	13.3 (7.5)	–
SLOF	–	–	188.4 (19.5)

ABCL, Adult Behavior Checklist; ADHD, attention deficit hyperactivity disorder; ASD, autism spectrum disorder; ASEBA, Achenbach System of Empirically Based Assessment; CBCL, Child Behavior Checklist; CD, conduct disorder; ED, eating disorder; HoNOSCA, Health of the Nation Outcome Scales for Children and Adolescents; OCD, obsessive-compulsive disorder; PD, personality disorder; SLOF, Specific Level of Functioning Scale; SSD, schizophrenia spectrum disorder; SUD, substance use disorder.

to 19.08). A significant time-gender interaction was observed. At all time points, there was a substantial difference between diagnoses, with no differential trend between genders along time points.

Table 2 Likelihood ratio test p values for fixed effects in multilevel models

Variable	ASEBA IP	ASEBA EP	HoNOSCA	SLOF
Time	<0.001	<0.001	<0.001	<0.001
Gender	0.004	0.11	0.88	<0.001
Diagnosis	<0.001	<0.001	<0.001	<0.001
Time:gender	0.005	0.36	–	0.086
Diagnosis:gender	–	0.44	0.015	–
Time:diagnosis	–	0.01	–	–
Time:diagnosis:gender	–	0.057	–	–

ASEBA, Achenbach System of Empirically Based Assessment; EP, externalising problems; HoNOSCA, Health of the Nation Outcome Scales for Children and Adolescents; IP, internalising problems; SLOF, Specific Level of Functioning Scale.

Self-reported ASEBA EP predicted scores (figure 1, online supplemental table 6S) showed a more complex pattern of variation, with a marginally not significant three-way interaction term, suggesting a differential time trend among gender conditional on diagnosis. Most notably, we observed an increasing trend in males compared with females, both in ADHD (T1: M 11.98, 95% CI 10.04 to 13.91; T4: M 16.95, 95% CI 14.41 to 19.49 vs T1: F 12.50, 95% CI 9.66 to 15.34; T4: F 11.52, 95% CI 8.55 to 14.49) and ASD (T1: M 11.77, 95% CI 9.68 to 13.86; T4: M 16.60, 95% CI 13.76 to 19.43 vs T1: F 11.80, 95% CI 9.24 to 14.37; T4: F 13.38, 95% CI 10.11 to 16.66).

Clinician-reported HoNOSCA predicted scores (figure 2, online supplemental table 7S) showed a significant decline over time, with a gender effect dependent on diagnosis. Males were generally more impaired, particularly in anxiety/somatic/trauma (T1: M 15.07, 95% CI 12.69 to 17.44 vs F 13.46, 95% CI 11.93 to 14.99) and ED/OCD (T1: M 12.16, 95% CI 8.94 to 15.38 vs F 11.52, 95% CI 9.79 to 13.25), while females were more affected in PD/CD/SUD (T1: F 17.42, 95% CI 14.89 to 19.96 vs M 13.63, 95% CI 11.18 to 16.07). Parent-reported SLOF predicted scores (figure 2, online supplemental table 8S) consistently showed better functioning in females across nearly all groups, especially in ADHD (T1: F 189.49, 95% CI 183.62 to 195.36 vs M 185.10, 95% CI 180.57 to 189.62), ASD (T1: F 176.94, 95% CI 170.11 to 183.77 vs M 171.16, 95% CI 165.13 to 177.20), depression/bipolar (T1: F 192.50, 95% CI 189.37 to 195.62 vs M 186.73, 95% CI 180.46 to 193.00) and SSD (T1: F 185.50, 95% CI 176.06 to 194.95 vs M 174.86, 95% CI 160.27 to 189.45). Both sexes showed slight improvements over time, with a very slight increase of the gap at T4. Overall, males exhibited higher symptom severity and impairment, while females showed better adaptive functioning across diagnostic categories.

Finally, the variance partitioning (see online supplemental table 10S) indicated that the majority of the variability was attributable to the subject-level differences, with a smaller proportion attributable to centre-level variation across all outcome measures.

The sensitivity analysis performed using data imputation via multivariate imputation by chained equations showed consistent estimates of the different effects without relevant variations. To explore the potential impact of the informative missingness, supplementary pattern-mixture models including missingness pattern and pattern-by-gender interaction terms were performed. LRTs showed no significant pattern-by-gender interaction across outcomes, suggesting that the main gender contrasts were broadly consistent across missingness patterns (see online supplemental table 15S). A significant main effect of missingness pattern emerged for ASEBA EP and SLOF scores, while borderline evidence was observed for HoNOSCA. In particular, differences were more evident for patterns characterised by missing assessments at later follow-up visits (eg, missing third and/or fourth visit). Conversely, no substantial pattern effect was observed for ASEBA IP.

Estimates from the pattern-mixture multilevel models are reported in online supplemental table 16S. Online supplemental table 16S presents diagnosis-specific estimates across follow-up time points (T1–T4) for ASEBA IP, ASEBA EP, HoNOSCA and SLOF outcomes; the table is found in the online supplemental material. Overall, although outcome levels varied according to missingness patterns, the absence of significant pattern-by-gender interactions suggests that the main findings regarding gender differences were not materially driven by differential missingness mechanisms.

FIGURE 1.

PREDICTED VALUES STRATIFIED BY SEX OF ASEBA IP AND ASEBA EP

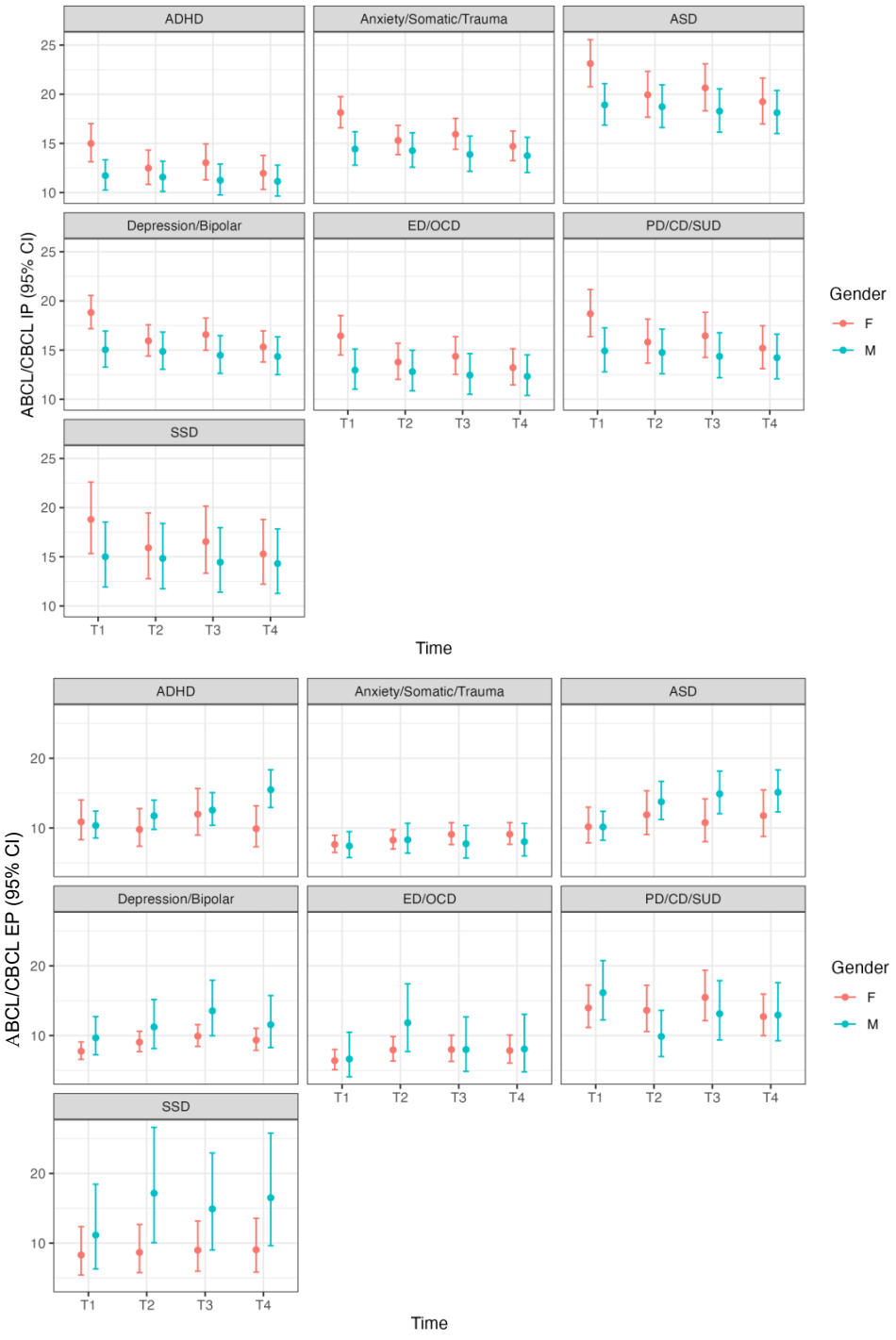

Figure 1 Predicted values stratified by sex for Achenbach System of Empirically Based Assessment (ASEBA) IP and ASEBA EP. Mean ASEBA IP and EP scores are plotted by sex and diagnostic category. Female participants showed higher IP scores across most diagnostic groups, particularly in ADHD, ASD and PD/CD/SUD, with minimal hange over time. In contrast, males exhibited higher EP scores, notably in depression/BD and SSD, with small sex differences in ADHD, ASD and ED/OCD. Symptom trajectories showed slight increases over time in some groups but remained relatively stable overall. ABCL, Adult Behavior Checklist; ADHD, attention deficit hyperactivity disorder; ASD, autism spectrum disorder; BD, bipolar disorder; CBCL, Child Behavior Checklist; CD, conduct disorder; ED, eating disorder; EP, externalising problems; IP, internalising problems; OCD, obsessive-compulsive disorder; PD, personality disorder; SSD, schizophrenia spectrum disorder; SUD, substance use disorder.

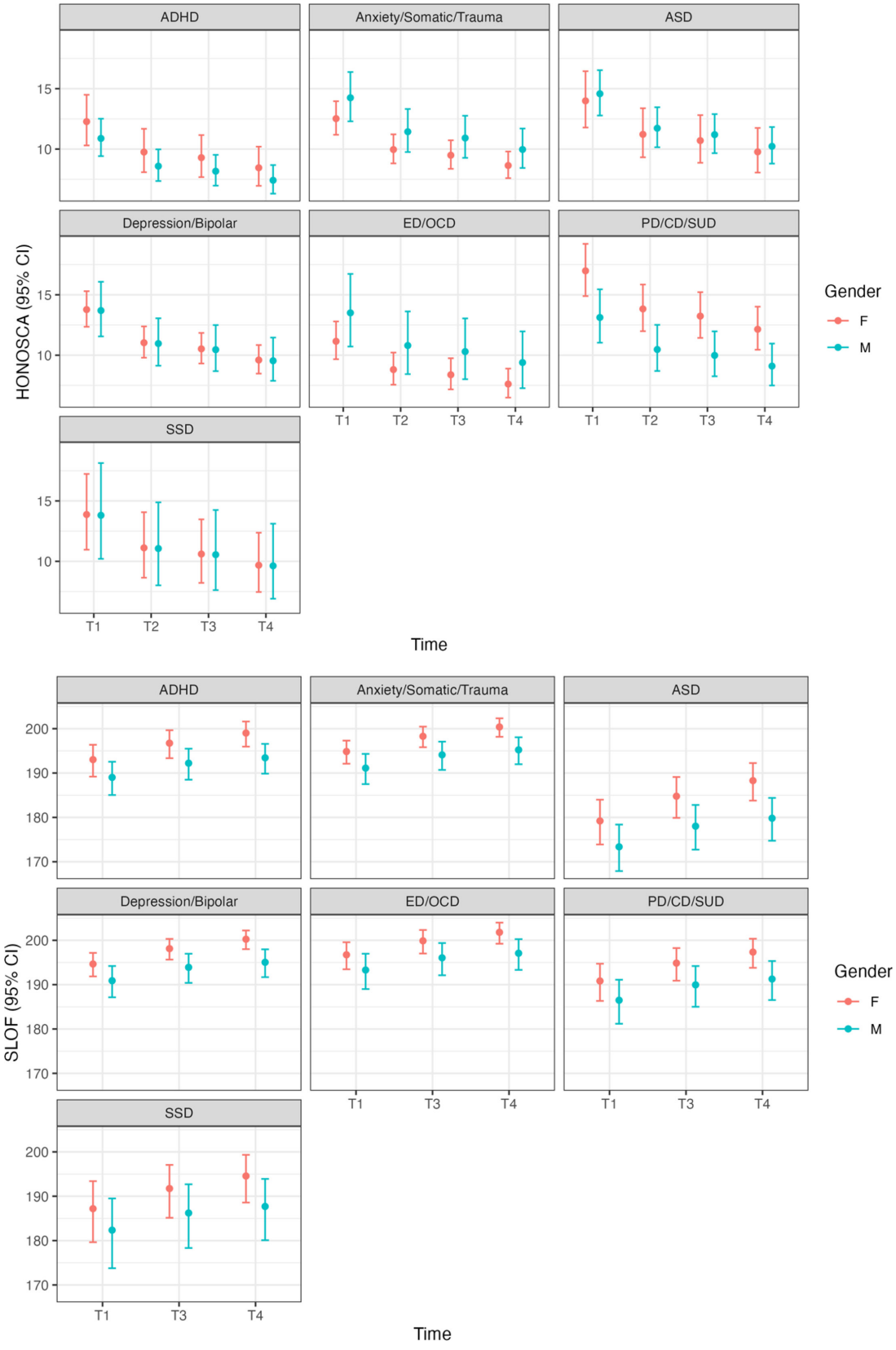

Figure 2 Predicted values stratified by sex for HoNOSCA and SLOF. Mean HoNOSCA and SLOF scores are shown by sex and diagnostic category. HoNOSCA scores, indicating symptom severity and impairment, were generally higher in males, especially in anxiety/somatic/trauma, ED/OCD and SSD, while females showed higher impairment in PD/CD/SUD. SLOF scores, reflecting adaptive functioning, were consistently higher in females across most diagnostic groups, particularly in ADHD, ASD, depression/BD and SSD. Both HoNOSCA and SLOF scores showed slight improvement over time, though sex differences remained stable. ADHD, attention deficit hyperactivity disorder; ASD, autism spectrum disorder; BD, bipolar disorder; CD, conduct disorder; ED, eating disorder; HoNOSCA, Health of the Nation Outcome Scales for Children and Adolescents; OCD, obsessive-compulsive disorder; PD, personality disorder; SLOF, Specific Level of Functioning Scale; SSD, schizophrenia spectrum disorder; SUD, substance use disorder.

DISCUSSION

This study explored sex-specific differences in psychiatric symptoms and clinical outcomes during the transition from adolescence to adulthood in a large European sample. Overall, the analyses identified some differences in symptom profiles, general mental health and psychosocial functioning between males and females; however, most sex×diagnosis×time interactions were non-significant and effect sizes were modest. Therefore, the findings should be interpreted as reflecting general tendencies rather than robust sex-specific developmental trajectories. At baseline, females tended to report higher levels of internalising symptoms, whereas males more often showed externalising symptoms and, in some diagnostic groups, greater overall severity. An exception was observed in females with PD/CD/SUD, who displayed relatively higher severity. Females also generally showed better psychosocial functioning. Longitudinally, no significant sex differences emerged in internalising symptom trajectories, while externalising symptoms showed only small increases, slightly more evident in males with ADHD. Across follow-up, both sexes showed modest improvements, and sex differences remained largely stable.

Our sex-stratified findings should be interpreted within the broader MILESTONE trial, which documented substantial cross-country heterogeneity in transition rates to AMHS and a non-uniform response to managed transition across diagnostic groups. This context is informative in two respects. First, in service systems where continuity of care is fragile, young people with more severe or externalising presentations—disproportionately male in our cohort—may be particularly vulnerable to disengagement, even when formal transition protocols are in place. Second, the most pronounced sex-related divergences in our analyses emerged precisely in the diagnostic categories (ADHD, ASD, internalising disorders) for which the parent trial reported differential intervention effects. Together, these observations suggest that one-size-fits-all transition models may inadequately address the differential clinical needs of male and female young people, and that sex should be considered alongside diagnosis when stratifying transition pathways.

ASEBA scores in males and females

Consistent with previous literature, females reported higher internalising symptoms than males, particularly in ADHD, ASD and PD/CD/SUD. These findings align with evidence of greater vulnerability to internalising symptoms among adolescent females,[5 7 8] although effect sizes were small and remained stable over time. Elevated internalising symptoms in females with PD/CD/SUD may reflect emotional distress associated with these conditions,[14 16] while similar patterns in neurodevelopmental disorders may relate to biological and psychosocial factors.[12 26] For example, females with ASD may experience greater social pressures and engage more frequently in camouflaging strategies, which have been associated with increased stress and affective symptoms.[13 27] However, as no sex×time interactions emerged, these interpretations remain speculative.

Males generally showed higher levels of externalising symptoms across several diagnostic groups, consistent with prior research linking male sex to disruptive behaviours in neurodevelopmental and mood disorders.[11 14 18] However, longitudinal changes were small and statistically significant only in limited subgroups.

HoNOSCA scores in males and females

Sex differences in clinician-rated psychopathological severity were observed in some diagnostic categories but were inconsistent overall. Males tended to show higher severity in SSD, anxiety-related disorders and ED/OCD, whereas females showed higher severity in PD/CD/SUD. These patterns broadly reflect previous findings,[9 14–16] but longitudinal effects were weak and mostly non-significant. Greater severity in females with PD/CD/SUD may be related to the complexity and psychosocial burden of these conditions,[14 16] while biological and social factors, including gender roles and hormonal fluctuations, may also contribute to women's vulnerability to addiction.[1 6]

The higher severity in males with anxiety-related disorders contrasts with their lower prevalence in epidemiological studies.[5 7] This discrepancy may reflect differences in referral pathways, clinical presentation or reporting styles rather than true differences in underlying psychopathology.[28] Overall, sex differences in severity were limited.

SLOF scores in males and females

Females showed higher SLOF scores, indicating better psychosocial functioning, consistent with previous studies.[8 29] However, these differences were relatively stable over time and did not indicate divergent developmental trajectories. The finding of better functioning in females with ADHD may partly reflect referral bias, with diagnosed females representing a higher functioning subgroup.[11] Caregiver ratings may also be influenced by gender expectations, with daughters perceived as functioning better than sons with similar symptoms.[14 28]

Better functioning in ASD females may reflect camouflaging rather than lower impairment.[13] Camouflaging, including imitation of social behaviours and verbal strategies, is not autism specific nor necessarily protective, but may contribute—together with gender-related diagnostic expectations—to delayed or missed recognition in females.[13] In SSD, poorer functioning in males aligns with earlier onset and severity.[15] Overall, sex differences were modest and stable, with no consistent sex×diagnosis×time effects. Consequently, interpretations regarding underlying mechanisms should be considered exploratory and mainly informed by previous literature rather than directly supported by the present findings.

Mental health trajectories over the 2-year follow-up

A general (non-significant) improvement in symptom severity and psychosocial functioning was observed over the 2-year follow-up; however, these changes did not reach statistical significance and effect sizes were modest. No substantial sex differences emerged in internalising symptoms from T1 to T4, while a slight worsening of externalising symptoms was observed in males with ADHD.

These findings should be interpreted cautiously. Interpretation is limited by the lack of detailed treatment data. Previous studies suggest treatment may reduce sex differences and produce similar improvement trajectories across sexes.[1 2 5 6] Functional improvement may occur even with persistent symptoms, reflecting developmental adaptation.[28] These interpretations extend beyond the present data and should be considered hypothesis generating rather than definitive.

Strengths and limitations

This exploratory study has several limitations. First, the inclusion of multiple predictors and outcomes increased the risk of type I error, while ethical constraints limited documentation of screening and recruitment procedures. In addition, missing SLOF data at T2 reduced completeness for some analyses.

Second, the absence of detailed information on participants' service pathways during follow-up limits interpretation of longitudinal findings. Without knowing whether young people transitioned to AMHS, remained in CAMHS or disengaged from care, it is difficult to distinguish developmental changes from service-related effects or selection processes.[19]

Third, diagnoses were based only on baseline routine clinical assessments from CAMHS records or treating clinicians. Given the diagnostic instability often observed during adolescence, the lack of diagnostic updates over the 2-year follow-up may have introduced misclassification bias in diagnosis-specific and sex-specific analyses. The relatively low representation of girls in some diagnostic groups may also affect the generalisability of sex-specific findings.[2 8] In disorders such as ADHD, under-recognition of female presentations may contribute to this imbalance, as standard rating scales may not adequately capture girls' symptom profiles.[11 29] These considerations highlight the importance of comprehensive interview-based assessments and caution against excluding ADHD solely on the basis of emotional or internalising symptoms.

Fourth, interpretation of sex differences in psychosocial functioning should consider possible measurement-related biases. As noted by Slof-Op 't Landt *et al*,[23] sex differences in sum scores may be difficult to interpret when measurement invariance is not established. This issue is particularly relevant for clinician-rated measures such as the SLOF and HoNOSCA.[22 24] Since invariance across sex was not formally tested, observed differences may partly reflect measurement properties rather than true functional differences. Furthermore, the use of multiple informants introduced additional complexity. In this study, CBCL and ASEBA scores were self-reported, SLOF ratings were caregiver reported and overall severity was clinician rated using the HoNOSCA. Informant discrepancies are common in adolescent mental health research and may influence interpretation of sex differences.[29] Previous work from the MILESTONE cohort showed only moderate agreement between clinicians and adolescents, particularly for less observable symptoms.[16] Parent–child agreement is generally low to moderate, and reporting discrepancies may vary according to the child's sex, with parents tending to under-report emotional symptoms in girls and conduct problems in boys.[1 16 20 29]

Despite these limitations, the study has important strengths. The large multicentric sample across eight European countries enhances statistical power and generalisability. The longitudinal design with four assessment points, the use of standardised instruments, the application of MICE to address missing data and the inclusion of adolescent, caregiver and clinician-reported measures provided complementary perspectives on symptom trajectories, severity and psychosocial functioning across diagnostic categories.

CONCLUSIONS AND CLINICAL IMPLICATIONS

Sex differences in adolescent mental health should be considered in transitional care. Evidence indicates differences in prevalence, symptom expression and developmental trajectories.[1 19 29] Failure to account for sex-specific patterns may contribute to delayed diagnosis and suboptimal care, particularly in under-recognised groups such as girls with ADHD/ASD or boys with ED.[11 13 15]

Further longitudinal research is needed to clarify the clinical relevance, stability and mechanisms underlying these differences.[6 8]

Author affiliations
[1]Unit of Epidemiological Psychiatry and Digital Mental Health, IRCCS Fatebenefratelli, Brescia, Italy
[2]Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy
[3]Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy
[4]Department of Statistics and Quantitative Methods, University of Milan-Bicocca, Milan, Italy
[5]Developmental EPI (Evidence synthesis, Prediction, Implementation) Lab, Centre for Innovation in Mental Health, School of Psychology, Faculty of Enviromental and Life Sciences, University of Southampton, Southampton, UK
[6]Clinical and Experimental Sciences (CNS and Psychiatry), Faculty of Medicine, University of Southampton, Southampton, UK
[7]Hampshire and Isle of Wight Healthcare, NHS Foundation Trust, Southampton, UK
[8]Hassenfeld Children's Hospital at NYU Langone, New York University Child Study Center, New York, NYC, USA
[9]DiMePRe-J-Department of Precision and Rigenerative Medicine-Joinc Area, University of Bari "Aldo Moro", Bari, Italy
[10]Department of Child and Adolescent Psychiatry and Psychology, Erasmus Medical Center, Rotterdam, Netherlands
[11]Department of Psychiatry, Clinical Hospital Center Split, Split, Croatia
[12]ARQ National Psychotrauma Centre, Diemen, Netherlands
[13]St John of God, Lucena Rathgar Dublin 6, Ireland
[14]University College Dublin, Belfield, Dublin 4, Ireland
[15]Children's Health Ireland (CHI) Crumlin, Dublin 12, Ireland
[16]CHU Monpellier, Child and Adolescent Psychiatry Unit (MPEA1), University of Monpellier, Monpellier, France
[17]INSERM CESP U1018, Villejuif, France
[18]Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, UK
[19]Centre for Interventional Pediatric Psychopharmacology and Rare Diseases (CIPPRD) & CIPP Rett Centre, National and Specialist Child and Adolescent Mental Health Services, Michael Rutter Centre, Maudsley Hospital, London, UK
[20]HealthTracker Ltd, Gillingham, UK
[21]University of Ulm, Ulm, Germany
[22]School of Nursing & Midwifery, University of Worcester, Worcester, UK
[23]Paediatrics, Warwick Medical School, Coventry, UK
[24]Department of Neurosciences, KU Leuven, Leuven, Belgium
[25]Department of Psychology, University of Warwick, Coventry, UK
[26]Warwick Medical School, Coventry, UK
[27]Child and Adolescent Psychiatry Unit, Bambino Gesu Pediatric Hospital, IRCCS, Rome, Italy
[28]Child Neuropsychiatry, Catholic University, Rome, Italy
[29]IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

Acknowledgements We thank and appreciate all the study participants, their families and carers, as well as the clinicians for their contribution. We are also grateful to the wider MILESTONE project consortium for their valuable collaboration during the study.

Collaborators MILESTONE consortium members: Jason Madan (Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, Coventry, UK); Moli Paul (Warwick Medical School, University of Warwick, Coventry, United Kingdom; Coventry and Warwickshire Partnership NHS Trust, Coventry, United Kingdom); Mathilde M Overbeek (Yulius Mental Health Organization, Yulius Academy, Dordrecht, 3300 BA, Netherlands; Department of Child and Adolescent Psychiatry and Psychology, Erasmus Medical Center, Rotterdam, 3000 CB, Netherlands; ARQ National Psychotrauma Centre, Diemen, Netherlands); Nikolina Davidović (University Hospital Split, Split, Croatia; School of Medicine, University of Split, Croatia); Virginie Maurice (Centre Hospitalier Universitaire de Montpellier, Saint Eloi Hospital, Montpellier, France); Frédérick Russet (Centre Hospitalier Universitaire de Montpellier, Saint Eloi Hospital, Montpellier, France); Anne Sartor (Josefinum Augsburg, Klinik für Kinder- und Jugendpsychiatrie und -Psychotherapie, Augsburg, Germany); Rebecca Appleton (NIHR Mental Health Policy Research Unit, Division of Psychiatry, University College London, London, UK); Federico Fiori (Department of Child & Adolescent Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, Kings College London, London, UK; Centre for Interventional Paediatric Psychopharmacology and Rare Diseases, South London and Maudsley NHS Foundation Trust, London, UK, HealthTracker, Kent, UK); Gaëlle Hendrickx (Department of Neurosciences, Centre for Clinical Psychiatry, KU Leuven, Leuven, Belgium); Laura Iozzino (IRCCS San Giovanni di Dio Fatebenefratelli, Brescia, Italy); Kate Lievesley (Department of Child & Adolescent Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, Kings College London, London, UK); Mathilde Mastroianni (Department of Child & Adolescent Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, Kings College London, London, UK, Centre for Interventional Paediatric Psychopharmacology and Rare Diseases, South London and Maudsley NHS Foundation Trust, London, UK); Virginie Maurice (Centre Hospitalier Universitaire

de Montpellier, Saint Eloi Hospital, Montpellier, France); Aesa Parenti (Child and Adolescent Psychiatry Department, Fontan Hospital, CHU Lille, Lille, France); Frédérick Russet (Centre Hospitalier Universitaire de Montpellier, Saint Eloi Hospital, Montpellier, France); Melanie Saam (Department of Child and Adolescent Psychiatry and Psychotherapy, University of Ulm, Ulm, Germany); Anne Sartor (Josefinum Augsburg, Klinik für Kinder und Jugendpsychiatrie und -Psychotherapie, Augsburg, Germany); Giulia Signorini (IRCCS San Giovanni di Dio Fatebenefratelli, Brescia, Italy); Jatinder Singh (Department of Child & Adolescent Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, Kings College London, London, UK, Centre for Interventional Paediatric Psychopharmacology and Rare Diseases, South London and Maudsley NHS Foundation Trust, London, UK); Priya Tah (Rees Centre, Department of Education, University of Oxford, Oxford, UK); Helena Jerkovic (University Hospital Split, Split, Croatia; School of Medicine, University of Split, Croatia), Giovanni Allibrio (ASST of Bergamo Ovest, Treviglio), Angelo Bertani (ASST Santi Paolo e Carlo, Milan), Sabrina Ferrari, Clin Psych (AUSL of Parma, Parma, Italy), Patrizia Conti (ASST of Como, Como), Francesco Margari (University of Bari, Bari), Ottaviano Martinelli (ASST of Lecco, Lecco), Renata Nacinovich (Fondazione IRCCS San Gerardo dei Tintori, Monza), Paolo Scocco (formerly ASL of Padua, Padua), Francesco Rinaldi (ASST Valcamonica, Esine), Paolo Stagi (Azienda USL Toscana Centro, Lucca), Stefano Vicari (Child Neuropsychiatric Unit, Department of Neuroscience, IRCCS Bambino Gesù Children's Hospital, IRCCS, Rome, Italy; Department of Life Sciences and Public Health, Catholic University, Rome).

Contributors MDA, SL, MM, FM: conceptualisation, validation, writing—original draft, writing—review and editing, visualisation. ACL, SC, MC, DM: methodology, software, validation, formal analysis, data curation. ARA, LI, SC, GD, TF, SG, AM, FMN, DPO, PS, UMES, CS, SS, ST, HT, LSB, DW, SV: validation, writing—review and editing, supervision. GG: conceptualisation, methodology, validation, investigation, resources, writing—review and editing, visualisation, supervision, project administration, funding acquisition. All authors read and approved the manuscript. GG is the guarantor of the study.

Funding The MILESTONE project was funded by the European Union's Seventh Framework Programme for research, technological development and demonstration (grant agreement number: 602442). This work was also supported by the Italian Ministry of Health (Ricerca Corrente), Ministero della Salute.

Disclaimer The views expressed in this publication are those of the authors and not necessarily those of the NIHR, NHS or the UK Department of Health and Social Care.

Competing interests PS is on the Advisory Board of Acadia Pharmaceuticals. He was previously a Principal Investigator (PI) for the Sarizotan clinical trial (protocol number: Sarizotan/001/II/2015; ClinicalTrials.gov identifier: NCT02790034) and the Anavex Life Sciences Corp trial (protocol number: ANAVEX2-73-RS-002). He is currently the PI for the Anavex Life Sciences Corp trial (protocol number: ANAVEX2-73-RS-003) for Rett syndrome (RTT). He is a coinventor of HealthTracker and serves as CEO and shareholder of HealthTracker Ltd. SC is an NIHR Research Professor and is funded by the NIHR for this research project (NIHR303122). SC is also supported by NIHR grants NIHR203684, NIHR203035, NIHR130077, NIHR128472, RP-PG-0618-20003, as well as grant 101095568-HORIZONHLTH-2022-DISEASE-07-03 from the European Research Executive Agency. SC has received reimbursement for travel and accommodation expenses from the Association for Child and Adolescent Central Health (ACAMH) in relation to lectures delivered for ACAMH, the Canadian AADHD Alliance Resource, the British Association of Psychopharmacology, and Healthcare Convention and CCM Group educational activities on ADHD. SC has also received honoraria from Medice.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the local ethics committee on 4 December 2014 (opinion 70/2014). This study was conducted in accordance with ethical standards, and informed consent was obtained from all participants prior to their involvement. For participants under the age of 18, parental consent and child assent were secured. The authors affirm that all procedures involved in this work adhere to the ethical standards set by the appropriate national and institutional committees on human experimentation, in accordance with the Helsinki Declaration of 1975, as amended in 2008. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID iDs
Miriam D'Addazio https://orcid.org/0000-0001-9361-4607
Silvia Leone https://orcid.org/0009-0000-8804-4849
Anna Caterina Leucci https://orcid.org/0000-0003-3562-6524
Marta Magno https://orcid.org/0000-0002-0400-774X
Anna Rita Atti https://orcid.org/0000-0002-6188-2647
Stefano Calza https://orcid.org/0000-0003-4996-7995
Martina Carnevale https://orcid.org/0009-0002-4855-0776
Elisa Caselani https://orcid.org/0000-0002-3042-0319
Laura Iozzino https://orcid.org/0000-0002-3633-8316
Federica Marcolini https://orcid.org/0000-0001-5875-0822
Donato Martella https://orcid.org/0009-0004-4398-0747
Samuele Cortese https://orcid.org/0000-0001-5877-8075
Gwendolyn Dieleman https://orcid.org/0000-0001-9700-5606
Tomislav Franić https://orcid.org/0000-0002-5240-7166
Athanasios Maras https://orcid.org/0000-0001-9361-1414
Fiona McNicholas https://orcid.org/0000-0001-9428-6908
Diane Purper-Ouakil https://orcid.org/0000-0003-0454-1579
Paramala Santosh https://orcid.org/0000-0003-4830-5893
Ulrike M E Schulze https://orcid.org/0000-0002-6350-5140
Cathy Street https://orcid.org/0000-0002-9949-0032
Swaran Preet Singh https://orcid.org/0000-0003-3454-2089
Sabine Tremmery https://orcid.org/0009-0008-5799-5965
Helena Tuomainen https://orcid.org/0000-0003-1636-8187
Larissa van Bodegom https://orcid.org/0009-0002-3020-0751
Dieter Wolke https://orcid.org/0000-0003-0304-268X
Stefano Vicari https://orcid.org/0000-0002-5395-2262
Giovanni de Girolamo https://orcid.org/0000-0002-1611-8324

REFERENCES

1 Christiansen DM, McCarthy MM, Seeman MV. Editorial: Understanding the influences of sex and gender differences in mental disorders. *Front Psychiatry* 2022;13:984195.
2 Murray AL, Ushakova A, Speyer L, *et al*. Sex/gender differences in individual and joint trajectories of common mental health symptoms in early to middle adolescence. *JCPP Adv* 2022;2:e12057.
3 Livanou MI, Bull M, Lane R, *et al*. Transition outcomes for young people discharged from adolescent medium secure services in England: A qualitative study exploring adolescents' and carers' experiences. *Clin Child Psychol Psychiatry* 2021;26:1227–42.
4 Medforth N, Boyle C. Challenges, Complexity, and Developments in Transition Services for Young People with Disabilities, Mental Health, and Long-Term Conditions: An Integrative Review. *Comprehensive Child and Adolescent Nursing* 2023;46:180–200.
5 Serio B, Kohler R, Ye F, *et al*. A multidimensional approach to understanding the emergence of sex differences in internalizing symptoms in adolescence. *Dev Cogn Neurosci* 2022;58:101182.
6 Singh SP, Tuomainen H, Girolamo G de, *et al*. Protocol for a cohort study of adolescent mental health service users with a nested cluster randomised controlled trial to assess the clinical and cost-effectiveness of managed transition in improving transitions from child to adult mental health services (the MILESTONE study). *BMJ Open* 2017;7:e016055.
7 Alloy LB, Hamilton JL, Hamlat EJ, *et al*. Pubertal Development, Emotion Regulatory Styles, and the Emergence of Sex Differences in Internalizing Disorders and Symptoms in Adolescence. *Clin Psychol Sci* 2016;4:867–81.
8 McKinney A, O'Brien S, Maybin JA, *et al*. Camouflaging in neurodivergent and neurotypical girls at the transition to adolescence and its relationship to mental health: A participatory methods research study. *JCPP Adv* 2024;4:e12294.
9 Silén Y, Keski-Rahkonen A. Worldwide prevalence of DSM-5 eating disorders among young people. *Curr Opin Psychiatry* 2022;35:362–71.
10 Zona K, Milan S. Gender differences in the longitudinal impact of exposure to violence on mental health in urban youth. *J Youth Adolesc* 2011;40:1674–90.
11 Fraticelli S, Caratelli G, De Berardis D, *et al*. Gender differences in attention deficit hyperactivity disorder: an update of the current evidence. *Riv Psichiatr* 2022;57:159–64.
12 Di Vara S, Guerrera S, Menghini D, *et al*. Characterizing individual differences in children and adolescents with autism spectrum disorder: a descriptive study. *Front Psychol* 2024;15:1323787.
13 Fombonne E. Camouflage and autism. *J Child Psychol Psychiatry* 2020;61:735–8.

14 Boson K, Anderberg M, Melander Hagborg J, *et al*. Adolescents with substance use problems in outpatient treatment: a one-year prospective follow-up study focusing on mental health and gender differences. *Subst Abuse Treat Prev Policy* 2022;17:53.
15 Sommer IE, Tiihonen J, van Mourik A, *et al*. The clinical course of schizophrenia in women and men-a nation-wide cohort study. *NPJ Schizophr* 2020;6:12.
16 Marcolini F, Magno M, Leone S, *et al*. Bridging Perspectives: Clinician-Adolescent Agreement on Psychopathological Severity in the European MILESTONE Cohort. *J Clin Psychol* 2026;82:317–29.
17 Plass-Christl A, Otto C, Klasen F, *et al*. Trajectories of mental health problems in children of parents with mental health problems: results of the BELLA study. *Eur Child Adolesc Psychiatry* 2018;27:867–76.
18 Fernandez Castelao C, Kröner-Herwig B. Developmental trajectories and predictors of externalizing behavior: a comparison of girls and boys. *J Youth Adolesc* 2014;43:775–89.
19 Singh K, Wendt FR. Effects of sex and gender on the etiologies and presentation of select internalizing psychopathologies. *Transl Psychiatry* 2024;14:73:73:.
20 Achenbach TM, Rescorla LA. *Manual for the ASEBA School-Age: Forms &Amp; Profiles: An Integrated System of Multi-Informant Assessment: Child Behavior Checklist for Ages 6-18, Teacher's Report For, Youth Self-Report*. ASEBA, 2001.
21 Achenbach TM. *Manual for the ASEBA School-Age Forms & Profiles*. University of Vermont, 2001.
22 Gowers SG, Harrington RC, Whitton A, *et al*. Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA). Glossary for HoNOSCA score sheet. *Br J Psychiatry* 1999;174:428–31.
23 Slof-Op 't Landt MCT, Dolan CV, Rebollo-Mesa I, *et al*. Sex differences in sum scores may be hard to interpret: the importance of measurement invariance. *Assessment* 2009;16:415–23.
24 Mucci A, Rucci P, Rocca P, *et al*. The Specific Level of Functioning Scale: construct validity, internal consistency and factor structure in a large Italian sample of people with schizophrenia living in the community. *Schizophr Res* 2014;159:144–50.
25 Kubinec R. Ordered Beta Regression: A Parsimonious, Well-Fitting Model for Continuous Data with Lower and Upper Bounds. *Polit Anal* 2023;31:519–36.
26 Tutz G, Hennevogl W. Random effects in ordinal regression models. *Computational Statistics & Data Analysis* 1996;22:537–57.
27 Quigley JA, Logsdon MK, Turner CA, *et al*. Sex differences in vulnerability to addiction. *Neuropharmacology* 2021;187.
28 Duax JM, Youngstrom EA, Calabrese JR, *et al*. Sex differences in pediatric bipolar disorder. *J Clin Psychiatry* 2007;68:1565–73.
29 Williamson D, Johnston C. Gender differences in adults with attention-deficit/hyperactivity disorder: A narrative review. *Clin Psychol Rev* 2015;40:15–27.