Journal of Clinical Psychology

WILEY

RESEARCH ARTICLE OPEN ACCESS

Bridging Perspectives: Clinician–Adolescent Agreement on Psychopathological Severity in the European MILESTONE Cohort

Federica Marcolini[1] | Marta Magno[2] | Silvia Leone[2] | Donato Martella[2,3] | Anna Caterina Leucci[1] | Anna Rita Atti[1] | Samuele Cortese[4,5,6] | Diana De Ronchi[1] | Gwen Dieleman[7] | Tomislav Franic[8] | Suzanne Gerritsen[7] | Athanasios Maras[9] | Fiona McNicholas[10,11,12] | Diane Purper-Ouakil[13,14] | Paramala Santosh[15,16,17] | Ulrike M. E. Schulze[18] | Cathy Street[19] | Swaran Singh[19] | Sabine Tremmery[20] | Helena Tuomainen[19] | Larissa S. van Bodegom[7] | Dieter Wolke[19,21] | Giovanni de Girolamo[2] | The MILESTONE Consortium

[1]Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy | [2]Unit of Epidemiological Psychiatry and Digital Mental Health, IRCCS Istituto Fatebenefratelli, Brescia, Italy | [3]Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milano, Italy | [4]Faculty of Medicine, University of Southampton, Southampton, UK | [5]University of Bari “Aldo Moro”, Bari, Italy | [6]New York University Child Study Center, New York, USA | [7]Erasmus Medical Center, Rotterdam, The Netherlands | [8]Department of Psychiatry, Medical School University of Split, Split, Croatia | [9]ARQ National Psychotrauma Centre, Diemen, The Netherlands | [10]Lucena Clinic, Saint John of God, Dublin, Ireland | [11]University College Dublin, Dublin, Ireland | [12]Children's Health Ireland at Crumlin, Dublin, Ireland | [13]CHU Montpellier, Saint Eloi and Colombière Hospitals, Montpellier, France | [14]INSERM U 1018, Centre for Epidemiology and Population Health (CESP), Villejuif, France | [15]King's College London, London, UK | [16]South London and Maudsley NHS Foundation Trust, London, UK | [17]HealthTracker Ltd, Kent, UK | [18]University of Ulm, Ulm, Germany | [19]Warwick Medical School, University of Warwick, Coventry, UK | [20]Department of Neurosciences, KU Leuven, Leuven, Belgium | [21]Department of Psychology, University of Warwick, Coventry, UK

Correspondence: Giovanni de Girolamo (gdegirolamo@fatebenefratelli.eu)

Received: 10 January 2025 | **Revised:** 8 July 2025 | **Accepted:** 21 November 2025

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

Keywords: CAMHS to AMHS transition | clinician–adolescent agreement | HoNOSCA | longitudinal assessment | psychopathological severity

ABSTRACT

Objectives: Adolescents transitioning from child and adolescent mental health services (CAMHS) to adult mental health services (AMHS) may face challenges in accurately identifying and reporting their mental health symptoms, often leading to discrepancies between clinician and patient evaluations. Using data from the MILESTONE project, this study aims to assess clinician-adolescent concordance over 24 months and identify domains of psychopathology with the highest disparities.

Methods: Participants were assessed at baseline, 9, 15, and 24 months using the Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA) scale and were categorized in four diagnostic groups. Hierarchical cluster analysis identified symptom-based subgroups of patients based on clinician and patient-rated HoNOSCA scores. Concordance was evaluated through multilevel linear regression models, while Bland–Altman plots examined agreement between scores across time points.

Results: Two clusters of patients were identified: one characterized by lower severity and greater prevalence, the other by higher complexity and fewer individuals. Clinician–patient concordance increased over time, rising from 77% to 83% by

---

Federica Marcolini and Marta Magno are joint first authors.

the second time point and stabilizing. Concordance varied across diagnostic categories, with anxiety showing the highest agreement and ADHD the lowest.

Conclusions: Improved communication, psychoeducation, and tailored interventions may facilitate greater patient–clinician alignment, thereby supporting more favorable outcomes during this critical developmental period.

Trial Registration: ISRCTN83240263; NCT03013595

1 | Introduction

Adolescence, defined as a developmental stage from ages 10–19, is commonly viewed as a period free of significant health issues (Jaworska and MacQueen 2015). Nonetheless, ~20% of adolescents are affected by mental health challenges, with depression and anxiety being the most prevalent (World Health Assembly 2015; World Health Organization 2003; World Health Organization Regional Office for Europe 2016). Adolescents may face difficulties in identifying and articulating mental health symptoms due to the dynamic nature of their emotional and cognitive development (Beckman et al. 2023; Westberg et al. 2022). In clinical contexts, accurately aligning clinician-administered evaluations with adolescents' self-reported assessments is essential, as there is frequently a discrepancy between patient self-assessments and clinical evaluations (Berlin et al. 2021; Hershenberg et al. 2020). These differences are especially pronounced in younger populations, where cognitive maturity and emotional regulation are still developing, impacting the accuracy of self-assessments (Lin et al. 2024). Recognizing these nuances is critical for treatment outcomes, as awareness of health status among those with mental health conditions plays a central role. Such disparities in clinician and self-assessed reports often arise due to variations in self-awareness, symptom interpretation, and the prioritization of symptoms, shaped by subjective experiences versus clinical criteria (Conijn et al. 2018; Dunlop et al. 2011; Kuitunen-Paul et al. 2023). In 2005, De Los Reyes and Kazdin addressed the inconsistencies often observed in reports of childhood psychopathology from different informants, such as parents, teachers, and children themselves. These discrepancies present challenges for clinicians and researchers when assessing and diagnosing children's mental health conditions (De Los Reyes and Kazdin 2005).

Assessing how young people subjectively experience their condition becomes especially critical during this transitional phase (Guyer 2020), such as the shift from pediatric psychiatric services (CAMHS) to adult mental health services (AMHS). During this period, adolescents are expected to take on a more active and autonomous role in their care, which underscores the importance of aligning clinician assessments with their self-reported experiences to ensure continuity and enhance treatment effectiveness (Livanou et al. 2021; Medforth and Boyle 2023). This shift may have been emotionally burdensome, requiring additional support from family members, peers, and healthcare providers (Tuomainen et al. 2018). However, research has highlighted a persistent gap in the agreement on the severity of psychopathology between young people and clinicians, with self-reported tools often being underutilized or undervalued in the clinical setting (Paul et al. 2015). The limited use of self-report measures in mental health evaluations has been attributed to concerns about the reliability and validity of these assessments, particularly when compared to clinician-administered tools (Kroenke et al. 2001). Moreover, despite the relevance of this topic (Abrams et al. 2018; Dunlop et al. 2011; Hauschildt et al. 2019; Ung et al. 2014) longitudinal studies specifically examining clinician–adolescent agreement during this transition phase are still scarce, likely due to the inherent challenges in conducting studies of this scale within transitional care settings. Tracking clinician–adolescent agreement over time helps capture how perceptions of symptom severity evolve during the transition from CAMHS to AMHS. Multi-informant assessments enhance clinical understanding by providing complementary perspectives, while discrepancies may reflect differences in insight, engagement, or relational dynamics (De Los Reyes and Kazdin 2005; Kraemer et al. 2003). Agreement can shift as adolescents' self-awareness increases and clinicians refine their understanding through ongoing therapeutic contact, making longitudinal assessment essential for guiding targeted interventions and ensuring continuity of care. The MILESTONE project (2014–2019) was a longitudinal, cohort study with a cluster-randomized controlled trial designed to advance the understanding and efficacy of transitional care from CAMHS to AMHS across various European healthcare systems (Singh et al. 2017).

Using data from the MILESTONE project, this prospective cohort study aims to assess the concordance between clinician and adolescent ratings on the Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA) scale (Gowers et al. 1999) over a 24-month period, with evaluations at four distinct time points (TP) (at baseline, 9, 15, and 24 months of follow-up). We decided to assess the degree of agreement between severity of psychopathology over 2 years because it allowed us to evaluate the stability and consistency of symptom severity over time. The study also seeks to identify specific domains of psychopathology with the highest clinician–patient discrepancies, enhancing the understanding of mental health assessments across diverse disorder categories and exploring how these alignments may impact treatment outcomes during this pivotal developmental stage. Clustering techniques have gained traction in clinical psychology as data-driven tools to identify transdiagnostic subgroups of patients based on symptom presentation. Unlike categorical diagnoses, which may obscure individual variability, cluster analyses allow for the emergence of empirical profiles that reflect clinical complexity and functional impairment more directly. This approach has proven particularly relevant in youth mental health, where comorbidity and developmental changes challenge rigid diagnostic frameworks. In the context of informant agreement, generating symptom-based clusters separately for clinicians and patients allows for a more nuanced understanding of whether both perspectives converge not only in terms of severity but in classifying individuals within similar clinical profiles. This methodology aligns with current person-centered frameworks and addresses the need to operationalize agreement beyond total score comparisons.

2 | Materials and Methods

2.1 | Study Design and Participants

The MILESTONE study was designed to investigate service utilization, mental health, and additional outcomes during a 2-year monitoring period among a group of 1004 young individuals aged 17–19 who have reached the upper age limit of their CAMHS across 8 European countries (Belgium, Croatia, France, Germany, Ireland, Italy, the Netherlands, and the United Kingdom). Previous studies have described their study design, inclusion and exclusion criteria, and recruitment procedures in detail (for a thorough review of the research protocol, please refer to Singh et al. 2017).

2.2 | Measures, Procedures, and Variables

Here, we focus on analyzing data from the participants who took part in all four assessments ($N = 614$) over a 24-month follow-up period: 3 months before transitioning from CAMHS to AHMS (T1), 9 months after T1 (T2), 15 months after T1 (T3), and 24 months after T1 (T4). At T1, all patients received CAMHS treatment. Following informed consent, participants and parents took part in a baseline assessment (T1) in CAMHS ~6 months before reaching the upper age limit. Trained researchers collected sociodemographic information and evaluated care needs through interviews held at CAMHS, in participants' homes, or via telephone, depending on their preferences. The sociodemographic data were obtained from a variety of sources, including clinicians, parents, medical records, and online questionnaires accessed through HealthTracker (https://www.healthtracker.co.uk).

The assessment battery included several structured rating scales; however, in this study, we focused our analysis specifically on the HoNOSCA (range 0–52), one of the most widely utilized clinician-rated instruments in adolescent mental health (Mathiassen et al. 2012). The HoNOSCA has demonstrated reliability and sensitivity to change in clinical settings, providing a comprehensive overview of mental health difficulties across various psychiatric conditions (Pirkis et al. 2005). The HoNOSCA consists of 13 items assessing severity of functioning in dimensions of behavioral problems (e.g., hyperactivity, aggression), emotional and somatic symptoms (e.g., anxiety, psychosis), social functioning impairments (e.g., peer conflict, school refusal) in children and adolescents with psychiatric disorders. It has demonstrated solid psychometric properties, including good inter-rater reliability ($\text{ICC} > 0.70$), sensitivity to clinical change over time, and strong construct and concurrent validity. Differences in scores between inpatient and outpatient groups further support its discriminant validity. A standardized glossary and scoring guide support consistent use in both clinical and research contexts (Gowers et al. 1999). In addition to the clinician-rated version of HoNOSCA, a self-rated variant was used, enabling young people to directly report their experiences and symptoms, thus incorporating both perspectives into the diagnostic and treatment process (Gowers et al. 2002). This tool was administered through semi-structured interviews with the young person, and, where feasible, with a parent or caregiver, as well as the relevant clinician. If the clinician was not available, information was gathered from a review of medical records.

To evaluate the agreement between clinicians and patients on the severity of psychopathology, we categorized our sample into four diagnostic groups based on core psychopathological characteristics, using their primary diagnosis: (i) anxiety disorders; (ii) depressive disorders; (iii) attention deficit and hyperactivity disorder (ADHD); and (iv) obsessive–compulsive disorders (OCD) alongside eating disorders (ED).

2.3 | Statistical Analysis

Descriptive statistics included frequency tables for categorical variables and mean values with corresponding standard deviations [SD] for continuous variables. We selected patients from a total sample of 474 who had available data for at least three-time points. To handle missing clinical scores (HoNOSCA), we employed multiple imputation at all four-time points to identify temporal trends prior to analysis (van Buuren and Groothuis-Oudshoorn 2011).

Using all the items in HoNOSCA, this study proposes a multidimensional analysis approach based on unsupervised learning methods. Hierarchical classification analyses (HCA) were performed for all available time points, separately for clinician and patient evaluations, using all HoNOSCA items as classification variables. HCA is an unsupervised learning technique that iteratively groups observations into clusters based on similarity across multiple dimensions—in this case, HoNOSCA item scores. We employed Euclidean distance as the proximity measure and complete linkage as the agglomeration method, which considers the greatest distance between elements across clusters. This approach enabled us to identify specific patient groups characterized by high intra-cluster homogeneity. Clusters were determined by inspecting the resulting dendrograms and selecting a number of groupings that balanced within-group homogeneity with conceptual interpretability. These clusters represent potential targets for delineating patient profiles based on HoNOSCA items.

The results can be interpreted at two levels: (i) a temporal level and (ii) a categorical level. We selected HCA over alternative clustering methods such as k-means or latent profile analysis (LPA) for several reasons. HCA does not require a priori specification of the number of clusters and is more robust to small and unequal cluster sizes, which aligns with our sample's diagnostic heterogeneity and the natural variation in symptom severity. Unlike LPA, HCA is distribution-free, making it better suited to the ordinal nature of the HoNOSCA data. This clustering approach was chosen to identify latent symptom configurations independently derived from clinician and patient assessments. By comparing group assignments across raters, we were able to examine alignment (or misalignment) in symptom patterns from a person-centered perspective. This method enables evaluation of concordance not only at the level of aggregated scores but also at the profile level, providing a nuanced understanding of how different informants perceive the same case. A reading on the temporal level allowed us to assess whether the patient's assignment to a group was stable or not at different time-points, and which target was characterized by greater concordance/discordance in the evaluation over time. Thus, this reading is only a comparison over time, not between clinician and patient evaluation categories.

The categorical level reading, on the other hand, allowed a comparison of clinicians and patient evaluations. For each time point, the assessment examined whether the patient's evaluation aligned with the clinician's evaluation within the same target group.

To examine whether clinician–patient agreement in cluster assignment changed significantly over time, we applied Cochran's Q test to repeated binary measures of agreement (agreement vs. disagreement) across the four assessment points. When the test was significant, post hoc pairwise comparisons between time points were conducted using McNemar's tests with Bonferroni correction. These analyses were also repeated within diagnostic subgroups to explore whether trends in agreement varied by clinical category. A supplementary analysis was carried out to check the robustness of the results. HoNOSCA total ratings by patients and clinicians were compared using the Bland-Altman method. The Bland–Altman plot visually represents this comparison, showing the mean of the measurements on the x-axis and the difference between the two sets of scores on the y-axis. Horizontal lines indicate the mean bias and limits of agreement, highlighting any systematic differences.

To examine the relationship between the mean of the measurements and the differences, linear regression was used according to the nature of the data. The linear model assumptions were assessed for normality of residuals using a Q–Q plot and the Shapiro–Wilk test, and no violations were detected. Additionally, a random intercept for each country was included in the model to account for potential variability across different countries. The slope of the regression line shows how the difference changes as the mean increases, and the intercept provides insight into the average bias. The p value associated with the slope reveals the statistical significance of the relationship, with a high p value suggesting good agreement between the two measurement methods.

All analyses were conducted using R software version 4.3.2 with a significance threshold of 0.05.

TABLE 1 | Descriptive statistics for sociodemographic characteristics of young people in the MILESTONE Cohort assessed at baseline ($n = 474$).

Variable	*N* (%) or mean (SD)
Gender	
Female	312 (65.8%)
Male	172 (34.2%)
Age	17.5 (0.6)
Ethnicity	
Caucasian	359 (75.7%)
Other ethnic groups	40 (8.4%)
Missing	75 (15.9%)
Country	
Belgium	58 (12.2%)
Croatia	17 (3.6%)
France	69 (14.6%)
Germany	50 (10.5%)
Ireland	23 (4.8%)
Italy	92 (19.4%)
The Netherlands	77 (16.2%)
UK London	31 (6.5%)
UK West Mids	57 (12.2%)
Living situation	
With biological parents	272 (57.4%)
With one biological parent	144 (30.4%)
Adoptive/foster parent(s) or other living arrangements	41 (8.6%)
Missing	17 (3.6%)
Current education	
Secondary/vocational	19 (4.0%)
Higher (under/postgraduate)	6 (1.3%)
No current school attendance	343 (72.4%)
Other	18 (3.8%)
Missing	88 (18.5%)
Diagnosis	
Anxiety	106 (22.4%)
Depressive	154 (32.5%)
ADHD	131 (27.6%)
OCD/ED	83 (17.5%)

Note: Percentages are based on the total sample.
Abbreviations: ADHD = attention-deficit/hyperactivity disorder, OCD/ED = obsessive–compulsive disorder or eating disorders.

3 | Results

3.1 | Sociodemographic and Clinical Characteristics

Table 1 and Table S1 report the sociodemographic and clinical characteristics of the adolescents included in this study. The sample comprised predominantly female participants (65.8%), with a mean age of 17.5 years (SD = 0.6). The majority were living with both biological parents (57.4%) and were not currently attending school (72.4%). Clinically, based on clinician ratings on the Clinical Global Impression—Severity scale, most participants were considered mildly or moderately ill (60.1%). The mean quality of life score, as measured by the WHOQOL-

TABLE 2 | HoNOSCA total ratings stratified by diagnosis across four time points.

Diagnosis	Evaluation	Time Point 1	Time Point 2	Time Point 3	Time Point 4
Anxiety	Clinician	11.7 (7.7)	9.1 (7.2)	9.8 (6.9)	8.2 (7.0)
	Patient	12.0 (8.9)	9.0 (7.5)	9.3 (7.3)	9.0 (7.0)
Depressive	Clinician	14.8 (7.7)	11.6 (7.4)	10.0 (6.5)	9.9 (7.2)
	Patient	14.9 (8.5)	13.0 (7.9)	12.1 (8.2)	11.0 (8.1)
ADHD	Clinician	11.7 (6.8)	9.9 (6.3)	8.7 (5.1)	9.1 (6.1)
	Patient	9.6 (7.4)	9.0 (7.1)	9.3 (6.5)	9.2 (7.1)
OCD/ED	Clinician	10.9 (6.4)	9.4 (6.8)	9.7 (7.3)	8.5 (6.1)
	Patient	12.0 (8.8)	11.1 (7.8)	11.1 (8.6)	9.4 (7.2)

Note: Percentages are based on the total sample.
Abbreviations: ADHD = attention-deficit/hyperactivity disorder, OCD/ED = obsessive–compulsive disorder or eating disorders.

TABLE 3 | Number of patients in each cluster and at each time point.

Time Point	Cluster	Patient's evaluation	Clinician's evaluation
Time Point 1	Cluster (i)	334	462
	Cluster (ii)	131	11
Time Point 2	Cluster (i)	396	461
	Cluster (ii)	470	4
Time Point 3	Cluster (i)	435	441
	Cluster (ii)	39	33
Time Point 4	Cluster (i)	552	611
	Cluster (ii)	63	14

Note: Results from HCA procedures. Clusters (i) and (ii) refer to groupings derived from cluster analysis based on symptom severity ratings. "Patient's evaluation" indicates self-reported HoNOSCA ratings; "Clinician's evaluation" refers to clinician-rated HoNOSCA scores.

BREF, was 79.4 (SD = 11.7), while the mean score for everyday functional skills (SLOF total) was 189.9 (SD = 18.6). Participants reported an average of 1.9 (SD = 1.6) stressful life events (range 0–13). Regarding service use, the majority had been in contact with CAMHS for more than 5 years (25.1%), while 70.9% had used services for up to 5 years.

Table 2 summarizes the HoNOSCA total mean scores from clinicians and patients across four time points for various diagnostic groups. Clinician ratings indicate a general decrease in symptom severity over time, especially in the depressive and anxiety groups. Patient ratings also show improvement but with more variability, particularly in the depressive group, where patients report a slower decline. ADHD ratings by patients remain stable, while clinician ratings initially improve and then slightly increase at the final time point.

3.2 | Classification Analysis

Two clusters of patients emerged from the HCA using both the clinician and patient evaluation variables. Statistical comparisons of total HoNOSCA scores between Clusters 1 and 2 were reported in Table S2 and were conducted using Mann–Whitney *U* tests at each time point. The first Cluster (1) is characterized by a lower mean of the individual scores than the second Cluster (2), which is the one with the highest mean values. Patient evaluations showed that, across all time points, individuals in Cluster (ii) reported significantly higher levels of difficulty than those in Cluster (i). These differences were particularly evident in areas such as emotional symptoms, relationships with peers and family, independence, skills, and self-harm. Emotional symptoms were notably higher in Cluster (ii) throughout (e.g., T1: 2.84 vs. 1.19). Although there were slight variations in the *p* values over time, the pattern of greater severity in Cluster (ii) remained stable. Clinician evaluations also revealed consistently higher scores for Cluster (ii) in almost all areas. Key differences were observed in emotional and somatic symptoms, family relationships, and attendance, with large effect sizes and highly significant *p* values (mostly $p < 0.001$). Notably, independence and skills also clearly distinguished the clusters. Group size (Table 3) is inversely proportional to complexity, with the most patients concentrated in Group (1) and the least in Group (2) (Figure 1).

The concordance between both the clinicians' evaluation and the patient's evaluation rose from around 77% at the first time point to 83% at Time Point 2 and remains stable at this percentage at the other two time points (Figure 2a). Looking at the temporal concordance for clinician and patient separately (Figure 2b), a greater stability over time can be seen in the clinician's assessment, which goes from 97% concordance between the first and second time points to 92% in the transition from the third to the fourth time point. The patient ratings are less stable over time, with the 77% stability observed from the first to the second time point rising to 82% from the penultimate

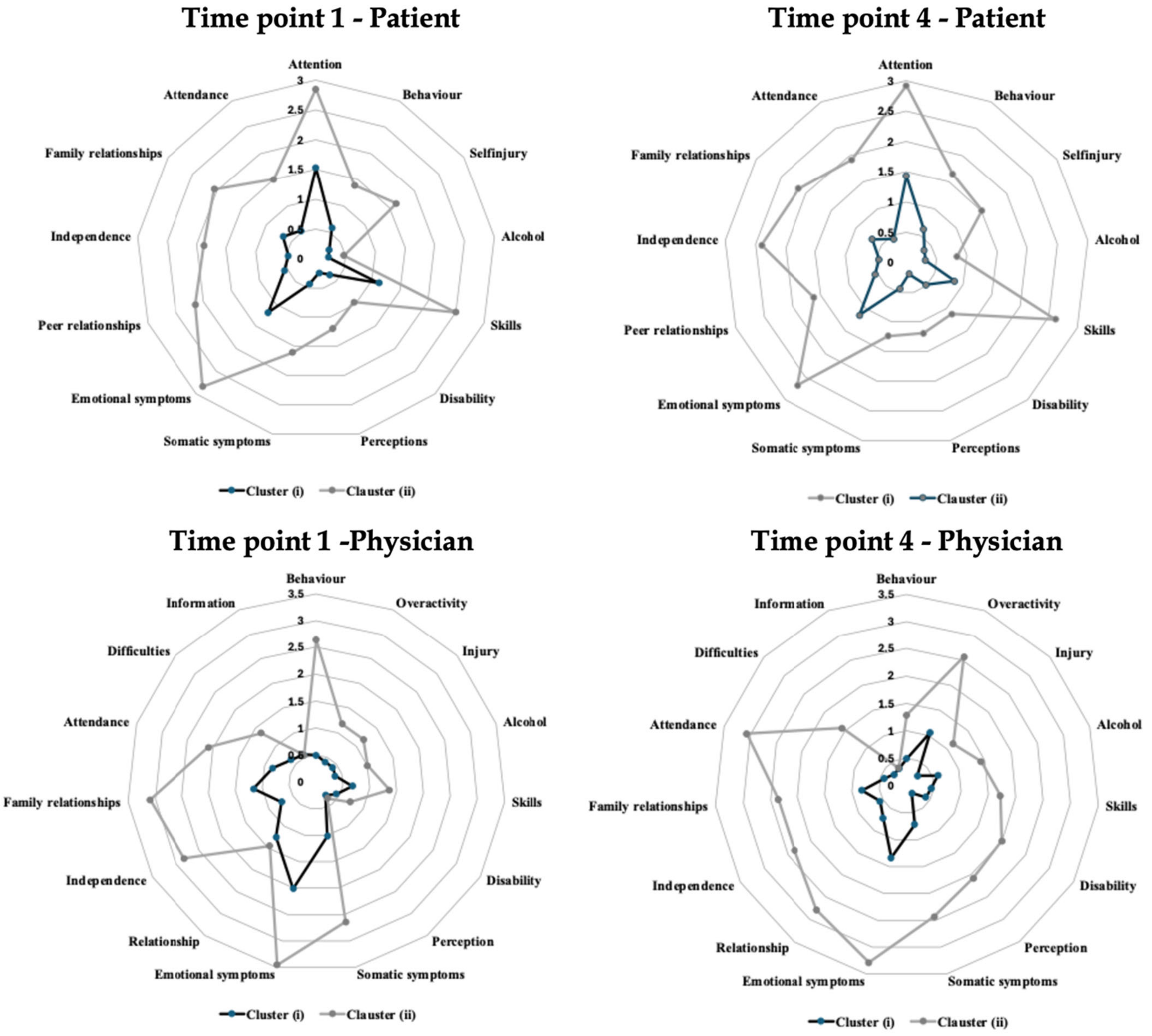

FIGURE 1 | Average item scores of patients and clinicians' evaluation in the two clusters at Time Points 1 and 4. Descriptive for complexity group extracted from HCA procedures. Patients evaluations: Time Point 1: *N* patients in Cluster (i) = 334; *N* patients in Cluster (ii) = 131; Time Point 4: *N* patients in Cluster (i) = 462; *N* patients in Cluster (ii) = 11; Clinicians evaluations: Time Point 1: *N* patients in Cluster (i) = 562; *N* patients in Cluster (ii) = 53; Time Point 4: *N* patients in Cluster (i) = 611; *N* patients in Cluster (ii) = 14.

to the last time point. Analysis stratified by disorder shows an increasing trend in the inter-rater concordance for the four diagnoses considered (Figure 2c).

The lowest concordance percentage at Time Point 1 is 64% and refers to patients diagnosed with ADHD, while the highest value is recorded for patients with depression (82%). For all diagnoses, there is an increase in the degree of concordance at subsequent time points.

3.3 | Robustness Analysis

Stratified analysis by diagnosis shows that for anxiety, the agreement varies across time points. At T1 ($p = 0.067$) patients generally report lower scores than clinicians. However, at T2 ($p = 0.694$), T3 ($p = 0.597$), and T4 ($p = 0.624$), there is strong agreement, indicating minimal bias. Overall, the agreement improves at the later time points (Table 4).

For patients with depression, T1 ($p = 0.033$) and T3 ($p = 0.002$) show significant discordance, with patients reporting notably lower scores than clinicians. T2 ($p = 0.288$) shows good agreement, and T4 ($p = 0.068$) is nearly significant. Overall, there is more discordance at T1 and T3, with improved agreement at T2 and T4.

In the ADHD group, T1 ($p = 0.295$) shows weak agreement, but there is no significant improvement in concordance from T2 ($p = 0.027$) to T4 ($p = 0.002$).

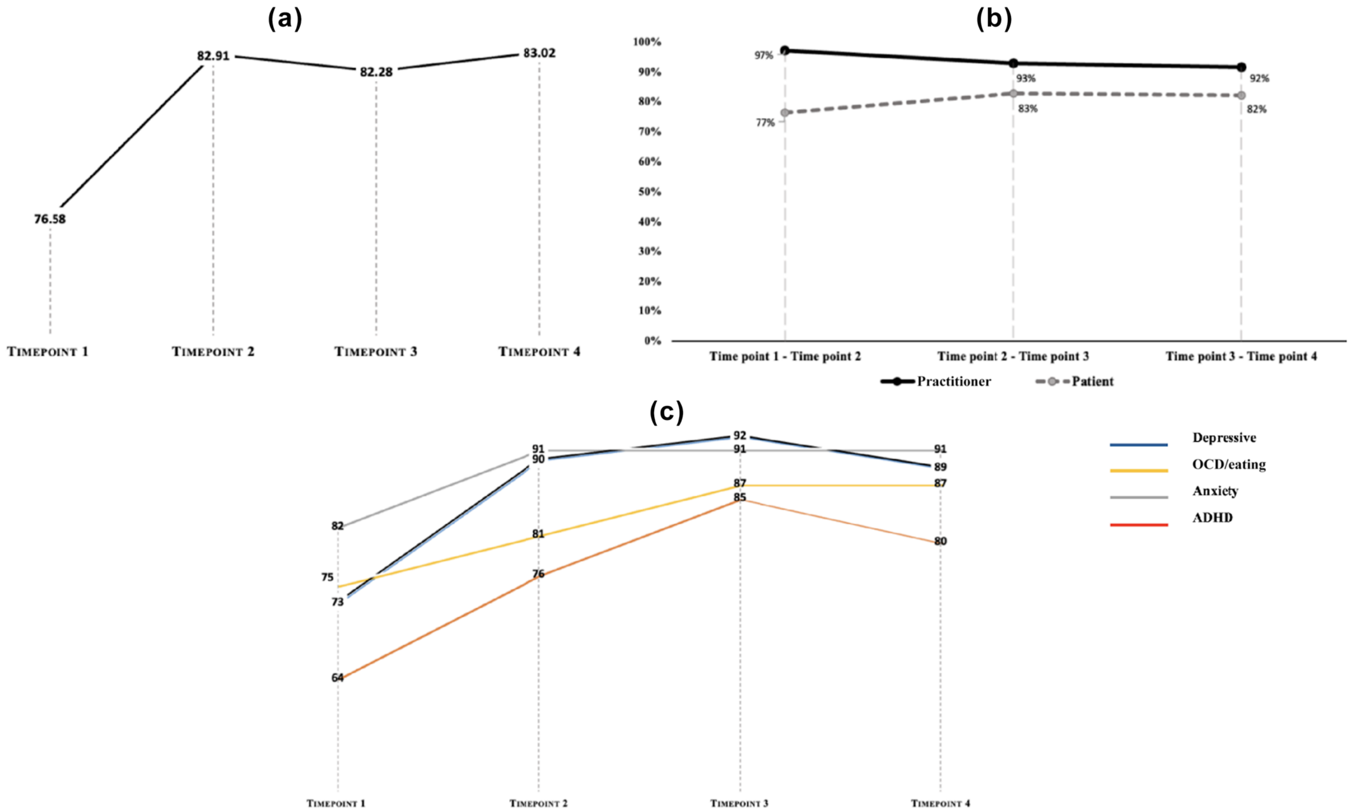

FIGURE 2 | Percentage of concordance at the four time points: (a) between clinician and patient evaluations; (b) across time points; (c) stratified by diagnosis. Concordance defined as assignment to the same severity cluster in different HCA procedures: (a) comparison of the cluster assigned by HCA based on clinician evaluation and HCA based on patient evaluation; (b) comparison of the clusters assigned at the four time points, separately for clinician and patient; (c) as (a), stratified by diagnosis. The overall Cochran's *Q* test was not significant across all time points, but a significant increase in agreement was observed between T1 and T2 only for patient evaluation (a). In stratified analyses (c) significant increases were also observed between T1 and T2 for depressive and anxiety disorders, and between T1 and T3 for OCD/eating and ADHD groups based on patient ratings.

TABLE 4 | Estimates from the linear regression model.

Diagnosis	Time point	Intercept	Slope	*p**
Anxiety	Time Point 1	−2.21	0.17	0.067
	Time Point 2	−0.11	0.04	0.694
	Time Point 3	−0.89	0.05	0.597
	Time Point 4	1.42	−0.05	0.624
Depressive	Time Point 1	−1.96	0.20	**0.033**
	Time Point 2	0.30	0.09	0.288
	Time Point 3	−1.26	0.31	**0.002**
	Time Point 4	0.16	0.14	0.068
ADHD	Time Point 1	−3.72	0.13	0.295
	Time Point 2	−3.96	0.24	**0.027**
	Time Point 3	−2.98	0.34	**0.001**
	Time Point 4	−3.94	0.32	**0.002**
OCD/ED	Time Point 1	−5.13	0.51	**< 0.001**
	Time Point 2	−0.87	0.26	**0.014**
	Time Point 3	−0.44	0.21	**0.042**
	Time Point 4	−0.39	0.22	0.063

Note: Bland–Altman regression results assessing clinician–patient agreement across mental health conditions and time points.
*The Bold values are significatives.

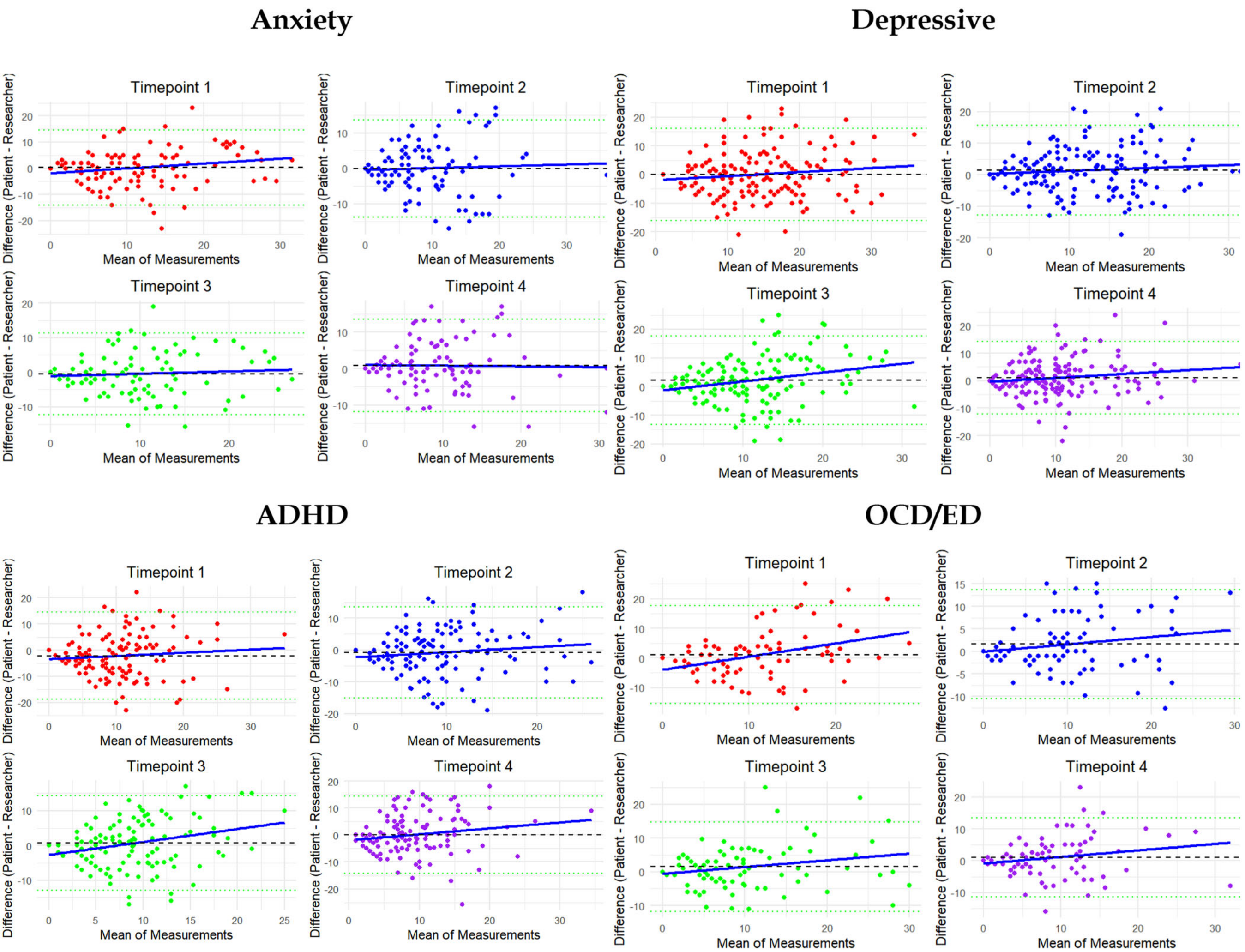

FIGURE 3 | Bland–Altman plot for evaluating differences between patients and clinicians' ratings. Bland–Altman plots for assessing agreement between clinicians and patients across four time points.

For OCD or ED, significant discordance is observed at T1 ($p < 0.001$) and T2 ($p = 0.014$), with patients reporting much lower scores than clinicians. The difference is smaller but still significant at T3 ($p = 0.042$). The scores converge by T4 ($p = 0.063$), showing no significant difference. Overall, there is a clear improvement in agreement over time (Figure 3).

4 | Discussion

The analysis of patient evaluations using HCA revealed a distinctive distribution pattern, elucidating the interplay between clinical complexity and patient representation. Two primary clusters were identified: the first, encompassing the majority of patients, was characterized by lower average scores, indicative of less severe or more straightforward cases. In contrast, the second cluster, although numerically smaller, exhibited higher scores and greater clinical complexity, suggestive of severe symptomatology or multifactorial presentations. This inverse relationship between the size of the group and the complexity of the cases aligns with common trends in healthcare systems, where the majority of patients present with less demanding needs, while a smaller subset, though less frequent, requires significantly greater resources and more nuanced approaches (Starfield et al. 2005). This pattern is particularly pronounced in mental health services, where a large portion of patients benefit from standard treatments such as psychoeducation, low-intensity cognitive-behavioral therapies, or medication management, aimed at addressing common conditions like mild-to-moderate anxiety and depression (Kazdin and Blase 2011). However, a smaller but critical subgroup presents with complex and severe conditions, necessitating more tailored and resource-intensive care, including multidisciplinary approaches, long-term therapy, or inpatient care (McHugh and Barlow 2010; Özge et al. 2023). This dual-tiered demand creates a challenge for mental health systems, as resources are often disproportionately consumed by the high-needs population, leaving fewer for the majority with less severe issues (Patel et al. 2018). Moreover, these high-need cases require nuanced therapeutic alliances, innovative treatments, and greater clinician expertise to address both the severity and heterogeneity of their presentations (Fonagy and Bateman 2006).

The progressive increase in concordance between clinician and patient assessments, from 77% at Time Point 1 to 83% at Time Point 2 and stabilizing thereafter, highlights a dynamic improvement in mutual understanding or adherence to treatment over time. This trend may be due to several factors, including improved communication between clinicians and patients, increased patient familiarity with the therapeutic process, or refinement of diagnostic tools. Interestingly, the long-term dynamics of concordance show different patterns between clinicians and patients. Clinicians' ratings showed greater stability, with consistently high concordance (dropping only slightly from 97% to 92%), reflecting their reliance on standardized protocols and professional expertise in assessing disorder progression. In contrast, patient ratings showed more variability, improving from 77% to 82% at the final time point. This variability likely reflects the subjective nature of patient perceptions, which can be influenced by external factors such as life events, treatment side effects or social support, as well as internal factors such as changes in self-awareness and symptom perception. Addressing these disparities requires integrating stepped care models, which stratify services based on patients' needs, ensuring efficient resource allocation while maintaining high-quality care across the spectrum (Bower and Gilbody 2005).

Patients experiencing anxiety disorders show greater concordance regarding the severity of their psychopathology throughout the course of treatment compared to individuals with other diagnostic groups. In another study, Mendoza et al. (2024) explored the relationship between various health indices, stressful life events, and anxiety symptoms, showing that both patients and clinicians acknowledged the influence of physical and mental health on anxiety severity. Patients with poorer mental health and those who experienced more stressful life events exhibited higher anxiety levels, which clinicians also identified during assessments. This shared understanding of the factors influencing anxiety severity contributes to agreement between patients and clinicians regarding the severity of psychopathology (Mendoza et al. 2024).

In contrast, patients suffering from depression exhibit a different trend in agreement with clinicians' evaluations. The degree of concordance between clinician-rated and patient-rated scales of depressive severity has been demonstrated to vary considerably in a number of studies (Domken et al. 1994; Rush et al. 1987). Research examining the correlation between these two types of assessments has produced inconsistent results regarding the level of concordance. A study conducted by Yamada et al. (2023) compared clinician-rated and self-reported measures of depression severity, showing that discrepancies between the two assessments are not uncommon. However, strong correlations were observed in certain cases, particularly when patients were actively engaged in their treatment (Yamada et al. 2023). Increased patient involvement in the treatment process fosters better alignment in the perception of depressive severity, as both the clinician and the patient work toward a shared understanding of the condition (IsHak et al. 2014).

On the other hand, patients suffering from ED, OCD, and ADHD demonstrate a less favorable alignment with clinicians' assessments. Indeed, patients suffering from ED often describe their experiences of psychopathology in ways that differ from clinicians' assessments (Fassino et al. 2009; Wallier et al. 2009). Research has demonstrated that individuals diagnosed with anorexia nervosa may not perceive the severity of their symptoms in a manner that aligns with the clinical assessment of the disorder's severity (Mac Donald et al. 2022). Similarly, patients with OCD may underestimate the intensity of their symptoms, frequently due to limited insight into the impact of their condition (Storch et al. 2015). This lack of self-awareness can contribute to a significant gap between the severity of symptoms as perceived by the patient and as assessed by the clinician, with the latter often recognizing functional impairments that the patient may minimize or fail to acknowledge (Cervin et al. 2024).

A qualitative review conducted by Marshall et al. 2019 found that subjective self-reports from patients often do not align with objective assessments made by clinicians. Patients may underreport or misinterpret their symptoms due to cognitive biases or emotional factors, resulting in a mismatch between their self-perception and the clinical evaluation of their condition. This discrepancy is particularly evident in instances where patients experience challenges in accurately recognizing or articulating their ADHD symptoms (Adamis et al. 2024; Marshall et al. 2019; Swanson et al. 2017).

4.1 | Implications for Clinical Practice

These findings carry important practical implications for clinical practice and therapeutic relationships in clinical settings. The improvement in clinician–patient concordance over time may represent an indirect indicator of therapeutic success and may reflect an improved quality of the therapeutic alliance (Aafjes-van Doorn et al. 2024): indeed, in Bordin's three-dimensional model of therapeutic alliance it is stated that a positive alliance is achieved through agreement on treatment goals and tasks (Bordin 1979). The therapeutic alliance is widely recognized as one of the most robust and consistent predictors of treatment outcome across different disorders, treatment settings and therapeutic approaches (Aafjes-van Doorn et al. 2024; Horvath and Symonds 1991). Enhancing this alignment through interventions that promote open and transparent communication could expedite this process, fostering mutual understanding and trust (Asan et al. 2021).

The observation of lower initial concordance in conditions such as depression highlights the need for tailored treatment and communication strategies for these populations. Programs that include psychoeducation or interventions aimed at enhancing patient awareness and engagement could prove especially effective in addressing these challenges (Breznoscakova et al. 2024; Mhango et al. 2023). Additionally, the role of symptom complexity in influencing agreement patterns suggests a need for more personalized assessment approaches. Clinical profiling through multi-informant ratings conducted longitudinally across the transition from adolescence to emerging adulthood provides a nuanced understanding of symptoms and their associated impairments, enabling personalized care. Moreover, careful documentation of information sources and attention to discrepancies between informants offer valuable insights for clinicians. Such discrepancies are not merely points of divergence but may serve as critical data to enrich diagnostic formulations and to inform treatment planning, fostering a more tailored and collaborative approach to care.

Practitioners should consider how demographic, social, and therapeutic factors may influence clinician–patient concordance during evaluations. Tailoring interventions—whether psychotherapy, pharmacotherapy, or a combination—could enhance alignment and foster a shared understanding of symptom severity; moreover, investigating the role of the intervention type and its impact on improving concordance could provide valuable insights. Additionally, it would be worthwhile to examine whether variations in concordance between clinicians and patients are associated with long-term therapeutic outcomes, potentially shedding light on the broader implications of alignment in symptom evaluation for treatment efficacy and patient well-being. While our analysis focused on descriptive tracking of cluster membership and agreement over time, formal longitudinal modeling could provide additional insights into the dynamics of symptom trajectories and cluster stability. Future studies with larger and more balanced samples may enable such modeling, thereby strengthening causal inferences about changes in clinician–adolescent concordance.

4.2 | Limitations and Strengths of the Study

A notable limitation of the present study is our inability to account for the natural course of symptom progression or remission over time and its potential impact on the level of agreement between clinicians and participants. In the absence of a control group, it is not possible to disentangle the specific effects of the intervention from potential spontaneous fluctuations in symptomatology, regression to the mean, or other time-related influences. This limits our ability to draw causal inferences about the observed outcomes. Moreover, we cannot determine how symptom trajectories may have evolved independently of the intervention, nor whether such changes may have contributed to the observed level of concordance. It is plausible that natural developmental improvements occurred in some diagnostic categories, or that the treatment itself had a positive impact on participants' well-being, although these possibilities lie beyond the scope of the present analysis. What can be observed, however, is a clear reduction in symptom severity reported by both clinicians and participants at T4, suggesting a likely convergence of perceptions at that specific time point. To our knowledge, this is the first study to investigate clinician–patient agreement rates during the critical developmental phase of transition from CAMHS to AMHS. This transitional period is characterized not only by increased vulnerability but also by diagnostic complexity, particularly for conditions such as personality disorders, which are often difficult or inappropriate to diagnose in adolescents and young adults. Another major strength of the study is its large and diverse sample, drawn from multiple European countries, which enhances the generalizability of the findings across different healthcare systems and sociocultural contexts. Moreover, the 24-month longitudinal design enables a more nuanced understanding of how clinician–patient agreement on psychopathological severity evolves over time, especially around the transition point. The inclusion of four distinct assessment points further strengthens the study by providing repeated measures that offer insight into symptom trajectories and shifts in agreement over an extended period.

5 | Conclusions

In conclusion, this study highlights the dynamic interplay between clinician and patient evaluations, which are shaped by time, diagnosis, and clinical complexity. By addressing initial discordance and leveraging the stability observed in longitudinal data, clinicians can optimize therapeutic strategies to align with patient needs and improve outcomes.

Author Contributions

Federica Marcolini, Marta Magno, Silvia Leone, and Giovanni de Girolamo conceived and designed the study. Federica Marcolini, Marta Magno, and Silvia Leone consulted literature and wrote the paper. Donato Martella and Anna Caterina Leucci performed the statistical data analysis. Federica Marcolini, Marta Magno, and Silvia Leone did the psychosocial interpretation of the results. Dieter Wolke, Paramala Santosh, Diane Purper-Ouakil, Cathy Street, Diana De Ronchi, Anna Rita Atti, and Samuele Cortese did the revision. Giovanni de Girolamo is the last author. All authors read and approved the manuscript.

Acknowledgments

The MILESTONE project was funded by the European Union's Seventh Framework Programme for research, technological development, and demonstration (Grant Agreement: 602442). This work was also supported by the Italian Ministry of Health (Ricerca Corrente). The Ministry of Health has no role in the analyses and interpretation of study findings. We extend our thanks and appreciation to all the study participants, their families and carers, and clinicians for their contribution. The authors are also grateful to the wider MILESTONE project consortium for their valuable collaboration during the all study: Jason Madan (University of Warwick, Coventry, UK); Moli Paul (University of Warwick, Coventry, UK); Mathilde M. Overbeek (Yulius Academy, Dordrecht, Netherlands); Nikolina Davidović, Helena Jerkovic (University of Split, Croatia); Virginie Maurice, Frédérick Russet (Centre Hospitalier Universitaire de Montpellier, Montpellier, France); Anne Sartor (Josefinum Klinik, Augsburg, Germany); Rebecca Appleton (University College London, London, UK); Federico Fiori (HealthTracker, Kent, UK); Gaëlle Hendrickx (KU Leuven, Leuven, Belgium); Kate Lievesley, Mathilde Mastroianni, Jatinder Singh (King's College London, London, UK); Aesa Parenti (CHU Lille, Lille, France); Melanie Saam (University of Ulm, Ulm, Germany); Anne Sartor (Josefinum Klinik, Augsburg, Germany); Giulia Signorini (IRCCS Fatebenefratelli, Brescia, Italy); Priya Tah (University of Oxford, Oxford, UK); Giovanni Allibrio (ASST Bergamo Ovest, Treviglio, Italy); Angelo Bertani (ASST Santi Paolo e Carlo, Milan, Italy); Antonella Squarcia (AUSL of Parma, Parma, Italy); Patrizia Conti (ASST of Como, Como, Italy); Francesco Margari, Adriana Pastore (University of Bari, Italy); Ottaviano Martinelli (ASST of Lecco, Lecco, Italy); Renata Nacinovich (IRCCS San Gerardo Dei Tintori, Monza, Italy); Paolo Scocco (Padua, Italy); Francesco Rinaldi (ASST Valcamonica, Esine, Italy); Paolo Stagi (Azienda USL Toscana Centro, Lucca, Italy); Stefano Vicari (IRCCS Bambino Gesù Children's Hospital, IRCCS, Rome, Italy). The MILESTONE project was funded by the European Union's Seventh Framework Programme for research, technological development, and demonstration (Grant Agreement: 602442). This work was also supported by the Italian Ministry of Health (Ricerca Corrente).

Ethics Statement

The study was approved by the UK National Research Ethics Service Committee West Midlands, South Birmingham (15/WM/0052) and the Ethical Committee of the Italian Coordinating Site IRCCS Fatebenefratelli on 4.12.2014 (permission 70/2014). All procedures contributing to this study comply with the Helsinki Declaration of 1975, as revised in 2024.

Consent

Informed consent has been appropriately obtained from patients and their parents.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

Aafjes-van Doorn, K., D. S. Spina, S. J. Horne, and V. Békés. 2024. "The Association Between Quality of Therapeutic Alliance and Treatment Outcomes in Teletherapy: A Systematic Review and Meta-Analysis." *Clinical Psychology Review* 110: 102430. https://doi.org/10.1016/j.cpr.2024.102430.

Abrams, J., S. V. Faraone, K. Y. Woodworth, T. J. Spencer, I. Biederman, and J. Biederman. 2018. "Are Adult ADHD Patients Good Informants of Their Symptoms? A Qualitative Literature Review of Concordance Between Clinician and Self-Report ADHD Symptoms." *Journal of Nervous & Mental Disease* 206, no. 9: 739–743.

Adamis, D., J. Singh, I. Coada, M. Wrigley, B. Gavin, and F. McNicholas. 2024. "Measuring Clinical Outcomes in Adult ADHD Clinics: Psychometrics of a New Scale, the Adult ADHD Clinical Outcome Scale." *BJPsych Open* 10, no. 6: e180. https://doi.org/10.1192/bjo.2024.739.

Asan, O., Z. Yu, and B. H. Crotty. 2021. "How Clinician-Patient Communication Affects Trust in Health Information Sources: Temporal Trends From a National Cross-Sectional Survey." *PLoS One* 16, no. 2: e0247583. https://doi.org/10.1371/journal.pone.0247583.

Beckman, L., S. Hassler, and L. Hellström. 2023. "Children and Youth's Perceptions of Mental Health-A Scoping Review of Qualitative Studies." *BMC Psychiatry* 23, no. 1: 669. https://doi.org/10.1186/s12888-023-05169-x.

Berlin, J., A. Tärnhäll, B. Hofvander, and M. Wallinius. 2021. "Self-Report Versus Clinician-Ratings in the Assessment of Aggression in Violent Offenders." *Criminal Behaviour and Mental Health* 31, no. 3: 198–210. https://doi.org/10.1002/cbm.2201.

Bordin, E. S. 1979. "The Generalizability of the Psychoanalytic Concept of the Working Alliance." *Psychotherapy: Theory, Research & Practice* 16, no. 3: 252–260. https://doi.org/10.1037/h0085885.

Bower, P., and S. Gilbody. 2005. "Stepped Care in Psychological Therapies: Access, Effectiveness and Efficiency. Narrative Literature Review." *British Journal of Psychiatry* 186: 11–17. https://doi.org/10.1192/bjp.186.1.11.

Breznoscakova, D., M. Pallayova, L. Izakova, and M. Kralova. 2024. "In-Person Psychoeducational Intervention to Reduce Rehospitalizations and Improve the Clinical Course of Major Depressive Disorder: A Non-Randomized Pilot Study." *Frontiers in Psychiatry* 15: 1429913. https://doi.org/10.3389/fpsyt.2024.1429913.

Cervin, M., D. R. M. A. Højgaard, S. Jensen, et al. 2024. "A General Factor of Psychopathology Predicts Treatment and Long-Term Outcomes in Children and Adolescents With Obsessive-Compulsive Disorder." *Journal of the American Academy of Child & Adolescent Psychiatry* 64: 810–821. https://doi.org/10.1016/j.jaac.2024.06.003.

Conijn, J. M., W. H. M. Emons, B. F. Page, et al. 2018. "Response Inconsistency of Patient-Reported Symptoms as a Predictor of Discrepancy Between Patient and Clinician-Reported Depression Severity." *Assessment* 25, no. 7: 917–928. https://doi.org/10.1177/1073191116666949.

De Los Reyes, A., and A. E. Kazdin. 2005. "Informant Discrepancies in the Assessment of Childhood Psychopathology: A Critical Review, Theoretical Framework, and Recommendations for Further Study." *Psychological Bulletin* 131, no. 4: 483–509. https://doi.org/10.1037/0033-2909.131.4.483.

Domken, M., J. Scott, and P. Kelly. 1994. "What Factors Predict Discrepancies Between Self and Observer Ratings of Depression?." *Journal of Affective Disorders* 31, no. 4: 253–259. https://doi.org/10.1016/0165-0327(94)90101-5.

Dunlop, B. W., T. Li, S. G. Kornstein, et al. 2011. "Concordance Between Clinician and Patient Ratings as Predictors of Response, Remission, and Recurrence in Major Depressive Disorder." *Journal of Psychiatric Research* 45, no. 1: 96–103. https://doi.org/10.1016/j.jpsychires.2010.04.032.

Fassino, S., A. Pierò, E. Tomba, and G. Abbate-Daga. 2009. "Factors Associated With Dropout From Treatment for Eating Disorders: A Comprehensive Literature Review." *BMC Psychiatry* 9: 67. https://doi.org/10.1186/1471-244X-9-67.

Fonagy, P., and A. W. Bateman. 2006. "Mechanisms of Change in Mentalization-Based Treatment of BPD." *Journal of Clinical Psychology* 62, no. 4: 411–430. https://doi.org/10.1002/jclp.20241.

Gowers, S. G., R. C. Harrington, A. Whitton, et al. 1999. "Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA). Glossary for HoNOSCA Score Sheet." *British Journal of Psychiatry* 174: 428–431. https://doi.org/10.1192/bjp.174.5.428.

Gowers, S., W. Levine, S. Bailey-Rogers, A. Shore, and E. Burhouse. 2002. "Use of a Routine, Self-Report Outcome Measure (HoNOSCA–SR) in Two Adolescent Mental Health Services." *British Journal of Psychiatry* 180: 266–269. https://doi.org/10.1192/bjp.180.3.266.

Guyer, A. E. 2020. "Adolescent Psychopathology: The Role of Brain-Based Diatheses, Sensitivities, and Susceptibilities." *Child Development Perspectives* 14, no. 2: 104–109. https://doi.org/10.1111/cdep.12365.

Hauschildt, M., R. Dar, J. Schröder, and S. Moritz. 2019. "Congruence and Discrepancy Between Self-Rated and Clinician-Rated Symptom Severity on the Yale–Brown Obsessive-Compulsive Scale (Y-BOCS) Before and After a Low-Intensity Intervention." *Psychiatry Research* 273: 595–602.

Hershenberg, R., W. M. McDonald, A. Crowell, et al. 2020. "Concordance Between Clinician-Rated and Patient Reported Outcome Measures of Depressive Symptoms in Treatment Resistant Depression." *Journal of Affective Disorders* 266: 22–29. https://doi.org/10.1016/j.jad.2020.01.108.

Horvath, A. O., and B. D. Symonds. 1991. "Relation Between Working Alliance and Outcome in Psychotherapy: A Meta-Analysis." *Journal of Counseling Psychology* 38, no. 2: 139–149. https://doi.org/10.1037/0022-0167.38.2.139.

IsHak, W. W., J. Mirocha, S. Pi, et al. 2014. "Patient-Reported Outcomes Before and After Treatment of Major Depressive Disorder." *Dialogues in Clinical Neuroscience* 16, no. 2: 171–183. https://doi.org/10.31887/DCNS.2014.16.2/rcohen.

Jaworska, N., and G. MacQueen. 2015. "Adolescence as a Unique Developmental Period." *Journal of Psychiatry & Neuroscience: JPN* 40, no. 5: 291–293. https://doi.org/10.1503/jpn.150268.

Kazdin, A. E., and S. L. Blase. 2011. "Rebooting Psychotherapy Research and Practice to Reduce the Burden of Mental Illness." *Perspectives on Psychological Science* 6, no. 1: 21–37. https://doi.org/10.1177/1745691610393527.

Kraemer, H. C., G. A. Morgan, N. L. Leech, J. A. Gliner, J. J. Vaske, and R. J. Harmon. 2003. "Measures of Clinical Significance." *Journal of the American Academy of Child and Adolescent Psychiatry* 42, no. 12: 1524–1529.

Kroenke, K., R. L. Spitzer, and J. B. W. Williams. 2001. "The PHQ-9: Validity of a Brief Depression Severity Measure." *Journal of General*

Internal Medicine 16, no. 9: 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Kuitunen-Paul, S., A. Eichler, M. Wiedmann, L. A. Basedow, V. Roessner, and Y. Golub. 2023. "Comparing Self-Report and Parental Report of Psychopathologies in Adolescents With Substance Use Disorders." *European Child & Adolescent Psychiatry* 32, no. 2: 331–342. https://doi.org/10.1007/s00787-021-01865-9.

Lin, S. C., C. Kehoe, E. Pozzi, D. Liontos, and S. Whittle. 2024. "Research Review: Child Emotion Regulation Mediates the Association Between Family Factors and Internalizing Symptoms in Children and Adolescents - A Meta-Analysis." *Journal of Child Psychology and Psychiatry* 65, no. 3: 260–274. https://doi.org/10.1111/jcpp.13894.

Livanou, M., S. D'Souza, R. Lane, B. La Plante, and S. P. Singh. 2021. "Challenges and Facilitators During Transitions From Adolescent Medium Secure Units to Adult Services in England: Interviews With Mental Healthcare Professionals." *Administration and Policy in Mental Health and Mental Health Services Research* 48, no. 6: 1089–1104. https://doi.org/10.1007/s10488-021-01115-9.

Mac Donald, B., C. M. Bulik, L. V. Petersen, and L. Clausen. 2022. "Influence of Eating Disorder Psychopathology and General Psychopathology on the Risk of Involuntary Treatment in Anorexia Nervosa." *Eating and Weight Disorders* 27, no. 8: 3157–3172. https://doi.org/10.1007/s40519-022-01446-y.

Marshall, P., J. Hoelzle, and M. Nikolas. 2019. "Diagnosing Attention-Deficit/Hyperactivity Disorder (ADHD) in Young Adults: A Qualitative Review of the Utility of Assessment Measures and Recommendations for Improving the Diagnostic Process." *Clinical Neuropsychologist* 35, no. 1: 165–198. https://doi.org/10.1080/13854046.2019.1696409.

Mathiassen, B., P. H. Brøndbo, K. Waterloo, et al. 2012. "IQ as a Moderator of Outcome in Severity of Children's Mental Health Status After Treatment in Outpatient Clinics." *Child and Adolescent Psychiatry and Mental Health* 6, no. 1: 22. https://doi.org/10.1186/1753-2000-6-22.

McHugh, R. K., and D. H. Barlow. 2010. "The Dissemination and Implementation of Evidence-Based Psychological Treatments. A Review of Current Efforts." *American Psychologist* 65, no. 2: 73–84. https://doi.org/10.1037/a0018121.

Medforth, N., and C. Boyle. 2023. "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* 46, no. 3: 180–200. https://doi.org/10.1080/24694193.2023.2245473.

Mendoza, L. S. M., M. Trujillo-Güiza, D. A. Forero, and S. Baez. 2024. "Health, Psychosocial and Cognitive Factors Associated With Anxiety Symptoms." *Current Psychology* 43, no. 26: 22376–22388. https://doi.org/10.1007/s12144-024-05998-3.

Mhango, W., L. Crowter, D. Michelson, and D. Gaysina. 2023. "Psychoeducation as an Active Ingredient for Interventions for Perinatal Depression and Anxiety in Youth: A Mixed-Method Systematic Literature Review and Lived Experience Synthesis." *BJPsych Open* 10, no. 1: e10. https://doi.org/10.1192/bjo.2023.614.

Özge, A., F. M. Domaç, N. Tekin, et al. 2023. "One Patient, Three Providers: A Multidisciplinary Approach to Managing Common Neuropsychiatric Cases." *Journal of Clinical Medicine* 12, no. 17: 5754. https://doi.org/10.3390/jcm12175754.

Patel, V., S. Saxena, C. Lund, et al. 2018. "The Lancet Commission on Global Mental Health and Sustainable Development." *Lancet* 392, no. 10157: 1553–1598. https://doi.org/10.1016/S0140-6736(18)31612-X.

Paul, M., C. Street, N. Wheeler, and S. P. Singh. 2015. "Transition to Adult Services for Young People With Mental Health Needs: A Systematic Review." *Clinical Child Psychology and Psychiatry* 20, no. 3: 436–457. https://doi.org/10.1177/1359104514526603.

Pirkis, J. E., P. M. Burgess, P. K. Kirk, S. Dodson, T. J. Coombs, and M. K. Williamson. 2005. "A Review of the Psychometric Properties of the Health of the Nation Outcome Scales (HoNOS) Family of Measures." *Health and Quality of Life Outcomes* 3: 76. https://doi.org/10.1186/1477-7525-3-76.

Rush, A. J., W. Hiser, and D. E. Giles. 1987. "A Comparison of Self-Reported Versus Clinician-Related Symptoms in Depression." *Journal of Clinical Psychiatry* 48, no. 6: 246–248.

Singh, S. P., H. Tuomainen, and G. Girolamo, MILESTONE Consortium, et al. 2017. "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* 7, no. 10: 016055. https://doi.org/10.1136/bmjopen-2017-016055.

Starfield, B., L. Shi, and J. Macinko. 2005. "Contribution of Primary Care to Health Systems and Health." *Milbank Quarterly* 83, no. 3: 457–502. https://doi.org/10.1111/j.1468-0009.2005.00409.x.

Storch, E. A., A. S. De Nadai, M. Conceição do Rosário, et al. 2015. "Defining Clinical Severity in Adults With Obsessive-Compulsive Disorder." *Comprehensive Psychiatry* 63: 30–35. https://doi.org/10.1016/j.comppsych.2015.08.007.

Swanson, J. M., L. E. Arnold, and B. S. G. Molina, MTA Cooperative Group, et al. 2017. "Young Adult Outcomes in the Follow-Up of the Multimodal Treatment Study of Attention-Deficit/Hyperactivity Disorder: Symptom Persistence, Source Discrepancy, and Height Suppression." *Journal of Child Psychology and Psychiatry* 58, no. 6: 663–678. https://doi.org/10.1111/jcpp.12684.

Tuomainen, H., U. Schulze, and J. Warwick, MILESTONE consortium, et al. 2018. "Managing the Link and Strengthening Transition From Child to Adult Mental Health Care in Europe (MILESTONE): Background, Rationale and Methodology." *BMC Psychiatry* 18, no. 1: 167. https://doi.org/10.1186/s12888-018-1758-z.

Ung, D., E. B. Arnold, A. S. De Nadai, et al. 2014. "Inter-Rater Reliability of the Anxiety Disorders Interview Schedule for DSM-IV in High-Functioning Youth With Autism Spectrum Disorder." *Journal of Developmental and Physical Disabilities* 26: 53–65.

van Buuren, S., and K. Groothuis-Oudshoorn. 2011. "Mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software* 45, no. 3: 1–67. https://doi.org/10.18637/jss.v045.i03.

Wallier, J., S. Vibert, S. Berthoz, C. Huas, T. Hubert, and N. Godart. 2009. "Dropout From Inpatient Treatment for Anorexia Nervosa: Critical Review of the Literature." *International Journal of Eating Disorders* 42, no. 7: 636–647. https://doi.org/10.1002/eat.20609.

Westberg, K. H., M. Nyholm, J. M. Nygren, and P. Svedberg. 2022. "Mental Health Problems Among Young People-A Scoping Review of Help-Seeking." *International Journal of Environmental Research and Public Health* 19, no. 3: 1430. https://doi.org/10.3390/ijerph19031430.

World Health Assembly. 2015. *Adolescent Health: Report by the Secretariat*. World Health Organization.

World Health Organization. 2003. Caring for Children and Adolescents With Mental Disorders: Setting WHO Directions. World Health Organization.

World Health Organization Regional Office for Europe. 2016. *Integrated Care Models: An Overview: Working Document*. World Health Organization. Regional Office for Europe. WHO IRIS iris.who.int/handle/10665/375502; http://iris.who.int/.

Yamada, R., T. Fujii, K. Hattori, et al. 2023. "Discrepancy Between Clinician-Rated and Self-Reported Depression Severity Is Associated With Adverse Childhood Experience, Autistic-Like Traits, and Coping Styles in Mood Disorders." *Clinical Psychopharmacology and Neuroscience* 21, no. 2: 296–303. https://doi.org/10.9758/cpn.2023.21.2.296.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.
Table S1: Type of service use (%) by children and adolescents over time.
Table S2: Item average in cluster at each time-point for patient and medical evaluation.