Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Markerless video analysis is feasible during a soccer program for autistic children in the SPARK pilot study

Giovanni Assenza

g.assenza@policlinicocampus.it

Università Campus Bio-Medico

Mario Merone

Università Campus Bio-Medico

Lorenzo Ricci

Università Campus Bio-Medico

Stefano Barboni

AS Ostiamare

Claudia Ferraro

Università Campus Bio-Medico

Davide Fornari

Azienda Sanitaria Locale Viterbo

Gian Luca Idini

Barbara Manzi

Azienda Sanitaria Locale Roma 2

Margareth Martino

ScopriAmo l'Autismo

Michele Ribolsi

Link Campus University

Vincenzo Di Lazzaro

Università Campus Bio-Medico

Emiliano Schena

Università Campus Bio-Medico

Leandro Pecchia

Università Campus Bio-Medico

Article

Keywords: autism, markerless video analysis, pose estimation, soccer, social responsiveness, parenting stress

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11388326/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Objective measurement of behavior during real-world interventions in autistic children remains challenging. This prospective, single-arm pilot study assessed the feasibility of markerless video analysis during an eight-week structured soccer program. Ten boys aged 7.5–10 years with level 1 autism completed eight weekly 60-minute sessions. Recordings from sessions 1 and 8 were processed using two-dimensional pose estimation to extract seven candidate motor and socio-motor metrics. Paired recordings were analyzable in all participants. HeadUp ratio, Posture Stability, Movement Fluidity, Active Time, Proximity, and Orientation increased in every child, whereas Restlessness decreased (exact two-sided Wilcoxon $p = 0.00195$ for each metric; all significant after false-discovery-rate correction). Among paired caregiver questionnaires, mean social responsiveness raw scores decreased from 97.22 to 76.78 and parenting stress scores from 100.11 to 85.56 ($p = 0.0117$ for each). Digital and questionnaire changes were directionally congruent, without formal association testing. Markerless analysis provided feasible, low-burden longitudinal measurement during naturalistic activities in autistic children. The uncontrolled design and absence of formal measurement validation preclude conclusions about intervention efficacy, reliability, or construct validity. Further validation is required before these metrics can serve as established clinical outcomes.

Introduction

Autism is a heterogeneous neurodevelopmental condition characterized by persistent differences in social communication and interaction, together with restricted or repetitive behaviors, interests, or activities [1]. Motor differences are also common and may involve balance, coordination, praxis, postural control, object-control skills, and interpersonal synchrony [2, 3]. These features are clinically relevant because motor competence contributes to participation in play, physical activity, and peer-based recreational contexts, and meta-analytic evidence suggests a modest but consistent relationship between gross-motor and social skills in autistic children [3].

Children and adolescents with Autism are less likely than their non-autistic peers to achieve recommended levels of moderate-to-vigorous physical activity [4]. Participation may be limited by motor difficulties, sensory sensitivities, behavioral dysregulation, social exclusion, reduced access to adapted programs, and limited availability of appropriately trained instructors [5, 6]. Physical-activity and motor interventions have therefore been increasingly investigated as adjunctive components of multidisciplinary support. Reviews suggest benefits for motor performance and possible effects on social communication, behavioral regulation, and executive functioning, although the evidence remains heterogeneous and is limited by small samples, variable protocols, and frequent reliance on unblinded caregiver-reported outcomes [7–13]. Physical activity should therefore be considered primarily as a health- and participation-promoting intervention rather than an established treatment for core autism features.

Organized group sports may be particularly relevant because motor learning occurs within repeated interpersonal contingencies. Soccer is widely available, relatively inexpensive, and readily adaptable to different motor and communication profiles. Ball control, passing, shooting, goalkeeping, and role rotation naturally require spatial orientation, anticipation, motor planning, visual monitoring, turn-taking, and interpersonal timing. Previous studies suggest that community football programs for autistic children are feasible and may be associated with improvements in motor, behavioral, communication, and social outcomes [14–17]. However, these studies do not resolve how behavioral change should be measured objectively during the activity itself.

Most intervention studies rely on caregiver questionnaires, standardized assessments performed outside the intervention, or manual behavioral coding. Caregiver reports provide clinically meaningful information across everyday contexts but may be influenced by expectancy, recall, and lack of assessor blinding. Standardized motor tests offer controlled measurement but may not capture behavior in naturalistic group settings, while manual video coding is ecologically relevant but labor-intensive, observer-dependent, and difficult to scale.

Several technologies can provide more objective movement assessment, but each involves trade-offs between accuracy, burden, portability, and ecological validity. Laboratory optical motion-capture systems provide highly accurate kinematics but require dedicated infrastructure, calibration, trained personnel, and reflective markers [18]. Wearable inertial sensors are more portable but still require body-mounted hardware that may alter spontaneous behavior or be poorly tolerated in children with sensory sensitivities. Depth or RGB-D cameras can provide three-dimensional information without markers, but capture volume, hardware requirements, and sensitivity to environmental conditions can limit use in larger or outdoor spaces [19]. By contrast, markerless pose estimation from standard RGB video combines low participant burden, inexpensive and portable acquisition, scalability, and the possibility of retrospective re-analysis with improved algorithms. Its limitations include reduced precision for detailed three-dimensional kinematics, out-of-plane movements, and occluded body segments [18–20]. For longitudinal monitoring of observable behavior in real-world settings rather than laboratory-grade joint kinematics, this balance favored standard RGB video coupled with markerless pose-estimation analysis.

AI-enabled markerless computer vision can identify body keypoints directly from ordinary video and transform their trajectories into quantitative descriptors of posture, movement variability, active participation, interpersonal distance, body orientation, and repetitive movement. Computer-vision approaches have already been explored in autism research [21, 22], and markerless motion capture is increasingly used across neurological and pediatric rehabilitation [20]. Although sensitivity to occlusion, camera viewpoint, and population-specific validation remain important limitations [18–20], the absence of markers or wearable sensors makes this approach particularly attractive for repeated assessment during naturalistic activities in autistic children.

The Soccer Program for Autism: Reinforcing Kinship and Behavior (SPARK) was developed as an eight-week structured soccer program for autistic children requiring level 1 support, integrating technical

soccer skills with balance, coordination, spatial orientation, turn-taking, role alternation, and adaptation to another participant's movement. Its principal innovation was not the soccer intervention itself, but the integration of markerless computer vision into a naturalistic group-based program. Seven candidate digital measures were derived from standardized video recordings—HeadUp ratio, Posture Stability, Movement Fluidity, Active Time, Proximity, Orientation, and Restlessness—to characterize longitudinal changes in observable motor and socio-motor behavior rather than to diagnose autism or infer higher-order social cognition.

The primary aim of this pilot study was to determine whether markerless video analysis could be feasibly implemented to quantify motor and socio-motor behavior in autistic children during a naturalistic soccer-based intervention. The secondary aim was to examine whether the direction and pattern of change in video-derived metrics were congruent with caregiver-reported changes in social responsiveness and family-related stress.

Methods

Study design

SPARK was a prospective, longitudinal, single-center, single-arm feasibility and proof-of-concept study (Fig. 1). The project was academically promoted by Università Campus Bio-Medico di Roma and conducted as an independent study. The intervention consisted of eight weekly 60-minute soccer sessions. Clinical assessments were obtained before the first session (T0) and after completion of the eighth session (T1). Standardized exercises performed during sessions 1 and 8 were video-recorded for computer-vision analysis. The study was designed primarily to determine whether markerless video analysis could be feasibly and consistently implemented in autistic children during a naturalistic group-sport intervention and whether the direction and pattern of change in video-derived metrics were congruent with changes in caregiver-reported outcomes. It was not designed to establish intervention efficacy The soccer program served as the naturalistic setting in which the markerless measurement framework was tested, rather than constituting the primary experimental treatment whose efficacy the study was designed to establish..

Setting and recruitment

Participants were recruited in collaboration with the ScopriAmo l'Autismo not-for-profit association. Training sessions were conducted at the facilities of Accademia Gialloazzurri ASD under the supervision of qualified soccer coaches and therapists experienced in supporting children with developmental and behavioral disorders. Study conception, clinical assessment, data management, and statistical analysis were coordinated by investigators at Università Campus Bio-Medico di Roma.

Participants

Ten children were enrolled. All participants were male and had a median age of 9 years (IQR 8–9; range 7.5–10). Consistent with current inclusive-language recommendations, identity-first language (e.g., “autistic children”) is used throughout, while “autism spectrum disorder” is retained when referring specifically to the clinical diagnosis. Eligibility criteria were: age between 6 and 10 years; clinical diagnosis of autism spectrum disorder requiring level 1 support according to DSM-5-TR criteria [23]; participation in an ongoing standardized rehabilitation program (i.e., ABA intervention); ability to understand and follow basic instructions required for adapted soccer activities; and written informed consent provided by a parent or legal guardian. Exclusion criteria included motor impairment preventing safe participation; lower-limb contracture or spasticity substantially limiting locomotion; medical conditions or pharmacological treatments substantially impairing exercise tolerance; cognitive or language difficulties preventing participation despite support; severe behavioral dysregulation preventing safe engagement; and other contraindications to recreational physical exercise.

Soccer intervention

Participants attended one 60-minute session per week for eight consecutive weeks. Sessions were delivered by coaches and therapists in a structured but flexible manner intended to support motivation, spontaneous participation, enjoyment, and shared engagement. Activities were generally performed in a reduced field area to limit environmental complexity and facilitate supervision. Core intervention components included ball control, passing, and shooting; slalom and obstacle negotiation; balance and jumping; spatial orientation; visual monitoring; waiting and turn-taking; role alternation; coach-mediated ball exchange; peer-mediated complementary tasks; and cooperation toward shared goals.

Peer-mediated exercise

Two children began simultaneously from separate starting positions and completed complementary motor sequences. One child crossed obstacles, performed a slalom, and shot at goal, while the other completed a separate pathway and moved into the goalkeeper position. Participants were encouraged to monitor the other child’s position and adapt their own timing accordingly.

Coach-mediated exercise

The child negotiated obstacles, identified the colored position of a coach, verbally named the color, exchanged the ball with that coach, and subsequently shot at goal. Participants then rotated roles with the goalkeeper. Variants increased motor complexity by requiring ball control with the feet or negotiation of obstacles of different heights.

Social responsiveness

Social functioning was assessed using the caregiver-report Social Responsiveness Scale, Second Edition (SRS-2) [24]. The questionnaire consists of 65 items and yields a total raw score and five domain scores: Social Awareness, Social Cognition, Social Communication, Social Motivation, and Restricted Interests and Repetitive Behavior, reported in the study database as Mannerisms. Higher values indicate greater social difficulty. The total score was considered the principal exploratory clinical outcome. Domain analyses were considered supportive and were adjusted for multiple comparisons. For exploratory interpretation of clinical relevance, we used the provisional minimal clinically important difference (MCID) proposed by Tan et al. [25]. In that study, an anchor-based analysis using Clinical Global Impression of Improvement ratings from previous autism trials estimated an SRS-2 raw-score MCID of 5.61 points; changes exceeding this value may be considered potentially clinically meaningful, although this threshold has not been independently validated across intervention contexts. Nine parents completed the planned clinical assessments at T0 and T1, resulting in nine complete paired SRS-2 records available for the current analysis.

Parenting stress

Parenting stress was assessed using the 36-item Parenting Stress Index–Short Form (PSI-SF), derived from the Parenting Stress Index – Fourth Edition (PSI-4) [26]. Each item is rated on a five-point Likert scale, yielding a Total Stress score ranging from 36 to 180, with higher scores indicating greater parenting stress. The PSI-SF also comprises three 12-item domains—Parental Distress, Parent–Child Dysfunctional Interaction, and Difficult Child—which should be analyzed separately if complete item-level data are available. In PSI-SF literature, scores at or above the 85th percentile are generally considered elevated or borderline clinically significant, while some studies use a Total Stress raw score > 90 as a practical clinical threshold; because percentile conversion depends on the normative reference, the > 90 raw-score threshold was used here only for descriptive exploratory classification [26, 27]. The Italian PSI-4 reference used in the study was Guarino et al. [26]. Nine caregivers completed the questionnaire at both time points (T0 and T1).

Video acquisition

Video recordings were acquired using two fixed DJI Osmo Action 4 cameras (DJI, Shenzhen, China), operating at 4K resolution (3840 × 2160 pixels) and 30 frames/s. The cameras were positioned around the training area to provide complementary views of the standardized soccer exercises and to reduce the impact of transient occlusions. The same camera configuration and field layout were maintained during sessions 1 and 8 to maximize longitudinal comparability. The two-camera configuration provided complementary viewpoints; however, the present analysis was based on two-dimensional pose estimation and did not involve three-dimensional reconstruction or multi-view triangulation. Video recordings were processed offline using YOLOv8m-pose (Ultralytics) for two-dimensional human pose estimation. For each detected participant, the algorithm estimated the coordinates and confidence scores of the principal anatomical landmarks, including the head, shoulders, elbows, wrists, hips, knees,

and ankles. A confidence threshold of 0.50 was applied to individual keypoints; keypoints below this threshold were treated as missing. Participant identity across consecutive frames was maintained through temporal association based on bounding-box overlap and spatial continuity of the detected body centroid. Short gaps in keypoint trajectories of up to five consecutive frames were reconstructed using linear interpolation, whereas longer gaps were treated as missing and excluded from the corresponding metric calculation. To reduce high-frequency frame-to-frame jitter in the estimated trajectories, a five-frame moving-average filter was applied before feature extraction. The same pose-estimation model, confidence threshold, tracking procedure, missing-data handling, smoothing parameters, and feature-extraction pipeline were applied to all participants and to both assessment sessions. The objective of the computer-vision analysis was not to diagnose autism or quantify autism characteristics, but to provide objective and reproducible measures describing motor organization, active participation, and socio-motor behaviour during the intervention itself.

Video-derived metrics

Seven digital metrics were extracted from the trajectories of body keypoints and interpersonal spatial relationships obtained through the markerless pose-estimation pipeline. These metrics were defined as study-specific candidate digital measures for the present proof-of-concept analysis and should not be considered standardized clinical measures or validated digital biomarkers. Consequently, established normative values, clinical reference ranges, and validated cut-off values are not currently available for these measures. HeadUp ratio was calculated as the proportion of valid frames in which the participant maintained the head in a horizontal or upward position, based on the relative configuration of the head and shoulder keypoints. Posture Stability was derived from the temporal variability of the shoulder–hip configuration, with lower variability corresponding to greater postural stability. Movement Fluidity was calculated from the temporal smoothness and continuity of the principal body-keypoint trajectories across consecutive frames.

Active Time was defined as the proportion of valid frames in which the normalized frame-to-frame displacement of the body centroid exceeded 0.02 of the participant's estimated body height, providing an estimate of active engagement during the exercise.

Proximity was defined as the proportion of valid frames in which the distance between the participant and the interaction partner was below 1.5 normalized body lengths. Orientation was defined as the proportion of valid frames in which the participant's trunk was directed toward the interaction partner within an angular tolerance of ± 45°.

Restlessness was derived from repetitive low-amplitude frame-to-frame displacement of the head and wrist keypoints, providing a quantitative descriptor of motor agitation.

All derived measures were transformed to a 0–1 range using min-max normalization. The same normalization procedure and parameter definitions were applied consistently across all participants and both assessment sessions. Higher values indicated better performance for HeadUp ratio, Posture

Stability, Movement Fluidity, Active Time, Proximity, and Orientation, whereas lower values indicated better performance for Restlessness. The choice of observable constructs was informed by prior computer-vision studies of head and postural movement in autistic children [28], interpersonal distance and facing direction in social interaction [29], and movement duration and velocity during caregiver-child interaction [30]. In this study, HeadUp ratio and Posture Stability describe head and trunk configuration; Movement Fluidity, Active Time, and Restlessness describe trajectory smoothness, movement presence, and repetitive low-amplitude motion; and Proximity and Orientation describe spatial relations to an interaction partner. These publications support the relevance of analogous observable features, but do not validate the present seven formulas or their soccer-specific thresholds. Head position was not interpreted as eye gaze, and repetitive low-amplitude movement was not classified as stereotypy.

Statistical analysis

The distribution of continuous variables was assessed using the Shapiro-Wilk test. Variables showing no significant departure from normality were summarized as mean ± standard deviation (SD), whereas non-normally distributed variables were reported as median and interquartile range (IQR). Paired changes were described using the corresponding distribution-appropriate summary measures. Because of the small sample, T0-T1 comparisons were performed using exact two-sided Wilcoxon signed-rank tests. The seven video-derived metrics constituted the primary digital outcome family. The SRS-2 total score and PSI-SF total score were treated as secondary exploratory clinical outcomes. The five SRS-2 domains were treated as a separate outcome family and corrected using the Benjamini-Hochberg false-discovery-rate (FDR) procedure. The PSI-SF total score was analyzed as a separate exploratory outcome. The seven video-derived metrics were adjusted using the Benjamini-Hochberg procedure. In keeping with the pilot aims, congruence between digital and clinical findings was assessed descriptively by comparing the direction and overall pattern of T0-T1 change; no formal correlation or association analysis between video-derived and questionnaire changes was performed. As an exploratory analysis of individual PSI-SF change, the Reliable Change Index (RCI) was calculated according to Jacobson and Truax [31], using the baseline study SD and the published PSI-SF Total Stress test-retest reliability coefficient of 0.84 [27]. The standard error of the difference was calculated as $SD_{baseline} \times \sqrt{[2(1 - r)]}$, and $|RCI| \geq 1.96$ was considered evidence that change exceeded expected measurement error. Because one participant had different caregiver respondents at T0 and T1, the robustness of this classification was also considered after exclusion of that pair. Missing values were not imputed.

Ethics

The study was approved by the Lazio Area 2 Ethic Committee at 17th/10/2024, approval number 254.24 CET2 cbm. The study was conducted in accordance with the Declaration of Helsinki and applicable institutional and national regulations. Written informed consent was obtained from a parent or legal guardian before study procedures. The consent process covered clinical assessments and video recording and analysis. Data were pseudonymized.

Use of generative artificial intelligence

ChatGPT (OpenAI) was used to assist with editorial restructuring and language editing during preparation of this manuscript. Authors reviewed the manuscript, maintained responsibility for the final content, and disclosed any additional AI use.]

Results

Feasibility and data completeness

Ten boys (median age, 9 years; IQR, 8–9; range, 7.5–10) entered and completed the eight-session soccer program. All eight weekly sessions were recorded in full for each of the 10 children, yielding complete video coverage of 80 child-session observations. For the prespecified longitudinal comparison, recordings from sessions 1 and 8 were analyzable in all 10 children (10/10 paired datasets), and all seven video-derived metrics could be extracted for every participant. Together, these results support the operational feasibility of video acquisition and markerless processing under the conditions of this pilot study.

Complete paired SRS-2 and PSI-SF data were available for nine participants. No adverse events were reported in the available study documentation.

Exploratory clinical outcome: social responsiveness

Complete paired SRS-2 data were available for nine participants. All T0 and T1 SRS-2 total and domain score distributions showed no significant departure from normality on the Shapiro-Wilk test (all $p > 0.05$). The SRS-2 total raw score decreased from 97.22 ± 15.03 at T0 to 76.78 ± 15.86 at T1, with a mean paired change of − 20.44 ± 21.28 points (exact Wilcoxon signed-rank $p = 0.0117$). Seven of nine participants showed a reduction in SRS-2 total score exceeding the provisional 5.61-point MCID, all in the direction of improvement; no participant showed a worsening exceeding this threshold. Individual trajectories are shown in Fig. 2. Social Cognition decreased from 18.78 ± 2.49 to 15.00 ± 2.35 (exact $p = 0.0156$; FDR $q = 0.0391$), and Social Motivation from 14.44 ± 4.80 to 10.00 ± 3.24 (exact $p = 0.0156$; FDR $q = 0.0391$). Social Communication decreased from 33.67 ± 6.86 to 27.78 ± 5.09; because the paired change was not normally distributed (Shapiro-Wilk $p = 0.030$), the change is reported as median − 4 points (IQR − 8 to 0). The nominal pre-post difference (exact $p = 0.0313$) did not remain significant after FDR correction ($q = 0.0521$). Social Awareness changed from 10.33 ± 2.45 to 8.89 ± 3.41 (exact $p = 0.125$; $q = 0.125$), and Mannerisms from 20.00 ± 4.47 to 15.11 ± 6.25 (exact $p = 0.0508$; $q = 0.0635$) (Table 1).

Table 1
Pre-post changes in SRS-2 total and domain raw scores in participants included in the current paired analysis.

SRS-2 domain	T0, mean ± SD	T1, mean ± SD	Mean change	Exact p	FDR q
Social Awareness	10.3 ± 2.4	8.9 ± 3.4	−1.4	0.125	0.125
Social Cognition	18.8 ± 2.5	15.0 ± 2.3	−3.8	0.0156	0.0391
Social Communication	33.7 ± 6.9	27.8 ± 5.1	−5.9	0.0313	0.0521
Social Motivation	14.4 ± 4.8	10.0 ± 3.2	−4.4	0.0156	0.0391
Mannerisms	20.0 ± 4.5	15.1 ± 6.3	−4.9	0.0508	0.0635
Total score	97.2 ± 15.0	76.8 ± 15.9	−20.4	0.0117	—

Table 1 **legend. Values are mean ± SD for nine paired participants; change is T1 − T0. P values are from exact two-sided Wilcoxon signed-rank tests. FDR q values apply to the five domain comparisons; the total score was a separate exploratory outcome. Social Communication change is additionally summarized in the text as median and IQR because its paired-change distribution was non-normal. SRS-2, Social Responsiveness Scale, Second Edition; SD, standard deviation; FDR, false discovery rate.**

Figure 2 **legend. Each colored line represents one participant, with the corresponding colored marker shown in the adjacent table. The table reports individual T0 and T1 raw scores and the change calculated as Δ = T1 − T0; negative Δ values indicate improvement and positive values indicate worsening. Delta values shown in bold identify participants whose absolute change exceeded the provisional 5.61-point minimal clinically important difference (MCID) proposed by Tan et al.** [25]. **This threshold is presented as an exploratory reference and should not be interpreted as a universally validated SRS-2 MCID across intervention settings.**

Exploratory clinical outcome: family-related stress

Complete paired PSI-SF data were available for nine caregivers. The T0, T1, and paired-change distributions showed no significant departure from normality on the Shapiro-Wilk test (all $p > 0.05$). The mean score decreased from 100.11 ± 21.55 at T0 to 85.56 ± 17.99 at T1, with a mean paired change of −14.56 ± 11.13 points. Eight of nine caregivers showed a reduction in score, whereas one showed a slight increase. The exact Wilcoxon signed-rank test indicated a significant paired change ($W = 1.5$, exact $p = 0.0117$). Using the commonly cited PSI-SF Total Stress raw-score threshold > 90 as an exploratory marker of clinically elevated parenting stress, 6/9 caregivers (66.7%) were above threshold at T0 compared with 3/9 (33.3%) at T1. This raw-score classification is descriptive because formal PSI-SF clinical interpretation is norm-referenced and typically based on percentile scores. In the exploratory RCI analysis, calculated according to the Jacobson–Truax method [31], the baseline SD of 21.55 and a published PSI-SF Total Stress test-retest reliability of 0.84 [27] yielded a standard error of the difference of 12.19 points and a reliable-change threshold of approximately 23.89 raw-score points. Two of nine caregivers showed reliable improvement (RCI ≤ − 1.96), none showed reliable worsening, and the result

remained two of eight after excluding the participant whose caregiver respondent differed between T0 and T1 (Fig. 3).

Figure 3 **legend. Each colored line represents one caregiver-participant pair; the adjacent table reports T0 and T1 raw scores, Δ = T1 − T0, and the corresponding RCI. Negative Δ and RCI values indicate improvement. Bold values identify reliable change (|RCI| ≥ 1.96). The RCI was calculated using the study baseline SD (21.55) and the published PSI-SF Total Stress test-retest reliability coefficient (r = 0.84), corresponding to an approximate raw-score threshold of 23.89 points. One participant had different caregiver respondents at T0 and T1; excluding this pair did not alter the number classified as showing reliable improvement.**

Primary digital outcomes: markerless video analysis

Computer-vision-derived metrics were available for all 10 participants, demonstrating complete paired digital data acquisition at the two prespecified time points. All participants showed changes in the prespecified direction for each of the seven digital metrics, with increases in HeadUp ratio, Posture Stability, Movement Fluidity, Active Time, Proximity, and Orientation, together with a reduction in Restlessness (Table 2). This uniform pattern indicates that the markerless framework was sufficiently sensitive to capture longitudinal within-participant change across multiple dimensions of observable motor and socio-motor behavior.

The largest proportional changes were observed for HeadUp ratio, Posture Stability, Orientation, and Proximity. These findings are consistent with greater head-up behavior, more stable postural organization, and increased spatial orientation toward coaches or peers during the standardized activities. Movement Fluidity and Active Time also increased consistently, while Restlessness decreased in every participant. Because these metrics are candidate digital measures rather than validated clinical biomarkers, they were interpreted as objective descriptors of behavior during the task rather than direct measures of attention, social cognition, or clinical improvement.

All paired comparisons reached statistical significance using exact Wilcoxon signed-rank tests (exact p = 0.00195) and remained significant after false-discovery-rate correction. The identical exact p-values across the seven video-derived metrics arise from the uniform direction of change observed across all 10 participants. When all 10 non-zero paired differences have the same sign, the minimum attainable two-sided exact Wilcoxon signed-rank p-value is 2/2^10 = 0.001953125. Thus, the identical p-values reflect the same extreme signed-rank configuration across the seven metrics rather than a computational or reporting error. The uniform directional pattern shows that the proposed markerless framework consistently detected longitudinal within-task change across all seven prespecified metrics. This finding should not be interpreted as evidence of measurement responsiveness, reliability, or construct validity, none of which was formally established in this pilot study.

Table 2
Pre-post changes in computer-vision-derived metrics between sessions 1 and 8.

Video metric	Session 1	Session 8	Mean change	Exact p
HeadUp ratio	0.435 ± 0.048	0.523 ± 0.051	+ 0.088	0.00195
Posture Stability	0.477 ± 0.040	0.562 ± 0.043	+ 0.085	0.00195
Movement Fluidity	0.426 ± 0.039	0.505 ± 0.039	+ 0.079	0.00195
Active Time	0.595 ± 0.032	0.653 ± 0.029	+ 0.058	0.00195
Proximity	0.305 ± 0.030	0.380 ± 0.031	+ 0.075	0.00195
Orientation	0.363 ± 0.039	0.443 ± 0.037	+ 0.080	0.00195
Restlessness	0.265 ± 0.036	0.202 ± 0.032	−0.063	0.00195

Table 2 **legend. Values are mean ± SD for 10 paired participants. Metrics were normalized to a 0–1 range; change is session 8 − session 1. P values are from exact two-sided Wilcoxon signed-rank tests. All seven comparisons remained significant after Benjamini–Hochberg FDR correction (q = 0.00195 for each). SD, standard deviation; FDR, false discovery rate.**

Discussion

The central question of this pilot study was not whether soccer is an effective treatment for autism, but whether markerless video analysis can be implemented in autistic children to obtain quantitative longitudinal measures during a naturalistic motor and social intervention. Three findings address this question. First, paired recordings were analyzable in all 10 participants and the same markerless pipeline yielded all prespecified digital metrics without wearable or body-mounted instrumentation. Second, all seven video-derived metrics showed consistent longitudinal within-task change in the prespecified direction between the first and eighth sessions. Third, the direction and overall pattern of these digital changes were descriptively congruent with secondary exploratory caregiver-reported outcomes. The clinical findings therefore provide supportive context, but do not establish responsiveness, construct validity, a participant-level association, or intervention efficacy.

Previous soccer-based studies in autism have focused mainly on feasibility and changes in motor, behavioral, communication, or social outcomes [14–17]. SPARK addresses a different methodological gap: objective measurement of observable motor and socio-motor behavior within the intervention itself. Its contribution is therefore primarily technological and methodological rather than evidence of soccer efficacy.

As a secondary exploratory outcome, the SRS-2 total raw score decreased by approximately 20 points, with seven of nine participants exceeding the provisional 5.61-point MCID proposed by Tan et al. [25] (Fig. 2). Social Cognition and Social Motivation remained significant after multiple-comparison correction. Because the MCID was derived in a different therapeutic context and has not been

independently validated across intervention settings, these findings should be viewed as supportive clinical context rather than evidence of treatment efficacy.

The most distinctive result was the feasibility and internal coherence of the markerless video framework. HeadUp, Proximity, and Orientation describe observable head position and spatial relation to interaction partners but should not be interpreted as gaze, joint attention, or communicative intent. Posture Stability, Movement Fluidity, Active Time, and Restlessness describe motor organization and task engagement. All seven metrics showed consistent longitudinal change across all 10 participants, demonstrating that the framework detected within-task change under the study conditions. This does not establish responsiveness or clinical validity, because practice, habituation, familiarity with the exercises, and task geometry may all have contributed.

The PSI-SF showed the same overall direction: mean Total Stress decreased and the proportion above the descriptive > 90 raw-score threshold fell from 6/9 to 3/9. The exploratory RCI was more conservative, identifying reliable improvement in two caregivers and no reliable worsening (Fig. 3). These results support the overall direction of change but neither validate the digital measures nor demonstrate soccer efficacy.

The main strengths were the prospective design, standardized exercises at both video time points, complete paired video data, and acquisition in an ecologically valid community setting. Importantly, the markerless approach used ordinary RGB video and required no wearable sensors, reflective markers, or body-mounted devices, reducing participant burden and preserving naturalistic behavior.

The study has several important limitations. First, the sample was small and included only autistic boys requiring level 1 support, limiting generalizability. Second, the absence of a control condition prevents separation of intervention-related change from maturation, concurrent rehabilitation, repeated exposure, familiarity with coaches, or simple task practice. Third, caregivers were aware of participation, introducing expectancy and common-method bias into the clinical outcomes. Fourth, the post-intervention SRS-2 referred to behavior over a six-month period, whereas the intervention lasted eight weeks. Fifth, complete paired questionnaire data were available for nine participants. Sixth, formal reliability of the video-derived metrics was not established, and the measures have not yet been externally validated against blinded human coding, independent standardized motor assessments, or a reference motion-capture system. Seventh, the present sample was too small for robust participant-level association analyses; therefore, congruence between digital and clinical changes was assessed descriptively rather than statistically. Finally, the highly concordant video changes may partly reflect familiarization with the standardized exercises rather than generalized behavioral improvement.

A measurement check could be undertaken retrospectively on the existing recordings, if the source videos remain available and their use is covered by consent. A subset of clips from both sessions could be selected by a documented rule and coded independently by blinded raters for visible head position, active movement, interpersonal proximity, and trunk orientation. Agreement between raters and between human coding and the corresponding automated metrics, along with sensitivity to keypoint-confidence

thresholds, interpolation, camera view, and task segment, could then be reported as explicitly post hoc analyses. These analyses were not performed in the present study. Even if favorable, they would establish agreement under these recording conditions rather than test-retest reliability, construct validity, or intervention efficacy.

The next step is formal validation of the markerless framework, including test-retest reliability, robustness to camera placement and occlusion, agreement with blinded human coding, and comparison with independent motor or motion-capture measures. Beyond validation, the technological value of this approach lies in low-burden acquisition from standard RGB video, scalability across sites, and the possibility of retrospective re-analysis as pose-estimation algorithms improve. Because acquisition does not depend on sport-specific hardware, the same framework could potentially be adapted to rehabilitation, school-based activity, adapted sport, and other real-world settings in which repeated ecological measurement is desirable. Larger controlled studies should preregister a limited set of digital endpoints, include repeated baselines to quantify habituation, and test whether participant-level digital change agrees with validated clinical and functional outcomes.

This prospective pilot study supports the feasibility of using markerless computer vision to quantify motor and socio-motor behavior in autistic children during a naturalistic group-sport intervention. Paired recordings were analyzable in all participants, and the same framework generated interpretable longitudinal metrics without wearable sensors, reflective markers, or body-mounted devices.

The direction and overall pattern of change in the video-derived measures were descriptively congruent with improvements in the secondary exploratory caregiver-reported outcomes. This agreement is hypothesis-generating only: it does not establish a statistical association, measurement responsiveness, construct validity, or intervention efficacy.

The principal implication therefore extends beyond soccer itself: markerless video analysis may provide a low-burden, scalable approach for objective monitoring of motor and socio-motor behavior during real-world rehabilitative or sport-based interventions in autism. Formal validation of reliability, construct validity, responsiveness, and clinical interpretability is required before these measures can be considered established digital outcome measures.

Data availability

De-identified participant-level data supporting the results are available on demand.

Code availability

The statistical analysis code is available on demand

Declarations

Additional information

Competing interests

All authors declare no competing interests

Funding

All intervention procedures were provided free of charge by the staff of ScopriAmo l'Autismo and Accademia Gialloazzurri. Researchers from Università Campus Bio-Medico di Roma (UCBM) provided scientific support, while UCBM covered the publication and insurance costs.

Author Contribution

G.A. conceived and designed the study, coordinated the project, supervised the clinical and scientific activities, contributed to data interpretation, and drafted the manuscript. M.M. contributed to the methodological design of the markerless video-analysis framework, software development, data processing, and interpretation of the digital outcomes. L.R. contributed to study methodology, data analysis and interpretation, and critical revision of the manuscript. S.B., D.F., G.L.I., B.M., and M.M. contributed to participant recruitment, organization and delivery of the soccer intervention, and data acquisition. C.F. contributed to data management, methodological support, and analysis of the digital outcomes. M.R. contributed to the clinical interpretation of the findings and critically revised the manuscript. V.D.L. contributed to study supervision, interpretation of the results, and critical revision of the manuscript. E.S. contributed to methodological supervision of the video-acquisition and measurement framework, interpretation of the technical findings, and critical revision of the manuscript. L.P. contributed to study methodology, statistical and technological supervision, interpretation of the results, and critical revision of the manuscript. All authors reviewed and approved the final version of the manuscript and agree to be accountable for the work.

Acknowledgement

The authors thank the participating children and families, ScopriAmo l'Autismo, Accademia Gialloazzurri ASD, and the coaches and therapists involved in the program.

Data Availability

Data availabilityDe-identified participant-level data supporting the results are available on demand.Code availabilityThe statistical analysis code is available on demand.

References

1. Lord, C. et al. Autism spectrum disorder. *Nat. Rev. Dis. Primers*. **6**, 5. https://doi.org/10.1038/s41572-019-0138-4 (2020).
2. Zampella, C. J., Wang, L. A. L., Haley, M., Hutchinson, A. G. & de Marchena, A. Motor skill differences in autism spectrum disorder: a clinically focused review. *Curr. Psychiatry Rep.* **23**, 64. https://doi.org/10.1007/s11920-021-01280-6 (2021).
3. Wang, L. A. L., Petrulla, V., Zampella, C. J., Waller, R. & Schultz, R. T. Gross motor impairment and its relation to social skills in autism spectrum disorder: a systematic review and two meta-analyses. *Psychol. Bull.* **148**, 273–300. https://doi.org/10.1037/bul0000358 (2022).
4. Liang, X. et al. Age-related differences in accelerometer-assessed physical activity and sleep parameters among children and adolescents with and without autism spectrum disorder: a meta-analysis. *JAMA Netw. Open.* **6**, e2336129. https://doi.org/10.1001/jamanetworkopen.2023.36129 (2023).
5. Must, A., Phillips, S., Curtin, C. & Bandini, L. G. Barriers to physical activity in children with autism spectrum disorders: relationship to physical activity and screen time. *J. Phys. Act. Health*. **12**, 529–534. https://doi.org/10.1123/jpah.2013-0271 (2015).
6. Arkesteyn, A. et al. Physical activity correlates in children and adolescents with autism spectrum disorder: a systematic review. *Disabil. Rehabil*. **44**, 6539–6550. https://doi.org/10.1080/09638288.2021.1970251 (2022).
7. Ruggeri, A., Dancel, A., Johnson, R. & Sargent, B. The effect of motor and physical activity intervention on motor outcomes of children with autism spectrum disorder: a systematic review. *Autism* **24**, 544–568. https://doi.org/10.1177/1362361319885215 (2020).
8. Healy, S., Nacario, A., Braithwaite, R. E. & Hopper, C. The effect of physical activity interventions on youth with autism spectrum disorder: a meta-analysis. *Autism Res.* **11**, 818–833. https://doi.org/10.1002/aur.1955 (2018).
9. Chan, J. S. Y., Deng, K. & Yan, J. H. The effectiveness of physical activity interventions on communication and social functioning in autistic children and adolescents: a meta-analysis of controlled trials. *Autism* **25**, 874–886. https://doi.org/10.1177/1362361320977645 (2021).
10. Howells, K. et al. Efficacy of group-based organised physical activity participation for social outcomes in children with autism spectrum disorder: a systematic review and meta-analysis. *J. Autism Dev. Disord*. **49**, 3290–3308. https://doi.org/10.1007/s10803-019-04050-9 (2019).
11. Jia, S. et al. The effect of physical exercise on disordered social communication in individuals with autism spectrum disorder: a systematic review and meta-analysis of randomized controlled trials. *Front. Pediatr.* **11**, 1193648. https://doi.org/10.3389/fped.2023.1193648 (2023).
12. Qi, K. et al. Effect of physical activity on social communication impairments in children with autism spectrum disorder: a meta-analysis. *Heliyon* 10, e39053 (2024). https://doi.org/10.1016/j.heliyon.2024.e39053

13. Koh, S. H. Analyzing the influence of physical exercise interventions on social skills in children with autism spectrum disorder: insights from meta-analysis. *Front. Psychol.* **15**, 1399902. https://doi.org/10.3389/fpsyg.2024.1399902 (2024).
14. Howells, K. et al. Can participation in a community organized football program improve social, behavioural functioning and communication in children with autism spectrum disorder? A pilot study. *J. Autism Dev. Disord*. **50**, 3714–3727. https://doi.org/10.1007/s10803-020-04423-5 (2020).
15. Howells, K. et al. Can a community-based football program benefit motor ability in children with autism spectrum disorder? A pilot evaluation considering the role of social impairments. *J. Autism Dev. Disord*. **52**, 402–413. https://doi.org/10.1007/s10803-021-04933-w (2022).
16. López-Díaz, J. M. & Felgueras Custodio, N. Garrote Camarena, I. Football as an alternative to work on the development of social skills in children with autism spectrum disorder with level 1. *Behav. Sci.* **11**, 159. https://doi.org/10.3390/bs11110159 (2021).
17. Chambers, C. & Radley, K. C. Training soccer skills to adolescents with autism spectrum disorder via peer-mediated behavioral skills training. *Behav. Anal. Pract.* **13**, 454–461. https://doi.org/10.1007/s40617-019-00381-2 (2020).
18. Wade, L., Needham, L., McGuigan, P. & Bilzon, J. Applications and limitations of current markerless motion capture methods for clinical gait biomechanics. *PeerJ* **10**, e12995. https://doi.org/10.7717/peerj.12995 (2022).
19. Scott, B., Seyres, M., Philp, F., Chadwick, E. K. & Blana, D. Healthcare applications of single camera markerless motion capture: a scoping review. *PeerJ* **10**, e13517. https://doi.org/10.7717/peerj.13517 (2022).
20. Lam, W. W. T., Tang, Y. M. & Fong, K. N. K. A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation. *J. Neuroeng. Rehabil*. **20**, 57. https://doi.org/10.1186/s12984-023-01186-9 (2023).
21. Kaur, M., Srinivasan, S. M. & Bhat, A. N. Comparing motor performance, praxis, coordination, and interpersonal synchrony between children with and without autism spectrum disorder (ASD). *Res. Dev. Disabil.* **72**, 79–95. https://doi.org/10.1016/j.ridd.2017.10.025 (2018).
22. Sabatos-DeVito, M. et al. Methodological considerations in the use of Noldus EthoVision XT video tracking of children with autism in multi-site studies. *Biol. Psychol.* **146**, 107712. https://doi.org/10.1016/j.biopsycho.2019.05.012 (2019).
23. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. *text. revision (American Psychiatric Association Publishing*. https://doi.org/10.1176/appi.books.9780890425787 (2022). 5th edn.
24. Constantino, J. N. & Gruber, C. P. *Social Responsiveness Scale, Second Edition (SRS-2): Manual* (Western Psychological Services, 2012).
25. Tan, H. et al. Accelerated continuous theta burst stimulation targeting left primary motor cortex for children with autism spectrum disorder: multicentre randomised sham controlled trial. *BMJ* **393**, e086295. https://doi.org/10.1136/bmj-2025-086295 (2026).

26. Guarino, A., Laghi, F., Serantoni, G., Di Blasio, P. & Camisasca, E. *Parenting Stress Index – Fourth Edition (PSI-4)*Giunti O.S., (2016).
27. Abidin, R. R. *Parenting Stress Index, Fourth Edition Professional Manual* (Psychological Assessment Resources, 2012).
28. Dawson, G. et al. Atypical postural control can be detected via computer vision analysis in toddlers with autism spectrum disorder. *Sci. Rep.* **8**, 17008. https://doi.org/10.1038/s41598-018-35215-8 (2018).
29. Budman, I. et al. Quantifying the social symptoms of autism using motion capture. *Sci. Rep.* **9**, 7712. https://doi.org/10.1038/s41598-019-44180-9 (2019).
30. Bey, A. L. et al. Automated video tracking of autistic children's movement during caregiver-child interaction: an exploratory study. *J. Autism Dev. Disord*. **54**, 3706–3718. https://doi.org/10.1007/s10803-023-06107-2 (2024).
31. Jacobson, N. S. & Truax, P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J. Consult Clin. Psychol.* **59**, 12–19. https://doi.org/10.1037/0022-006X.59.1.12 (1991).

Figures

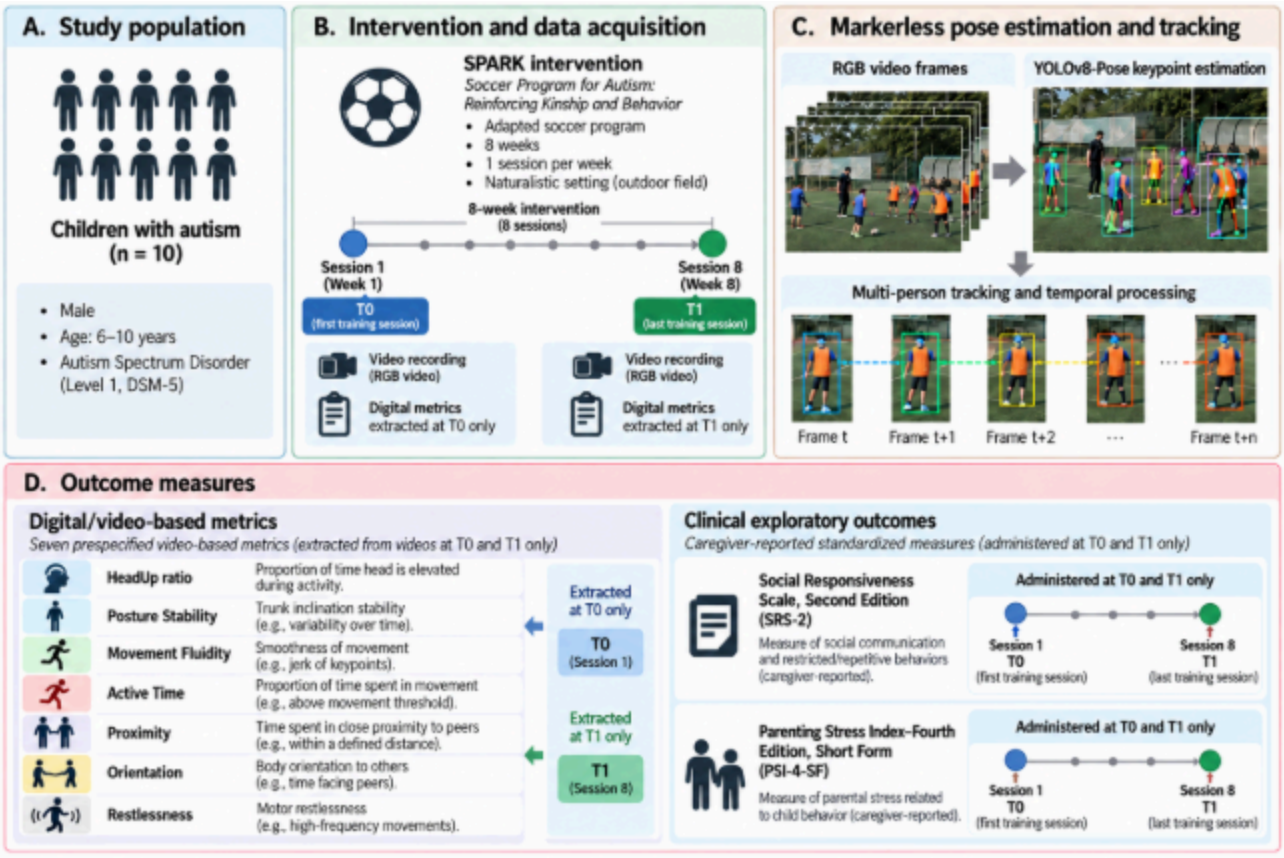
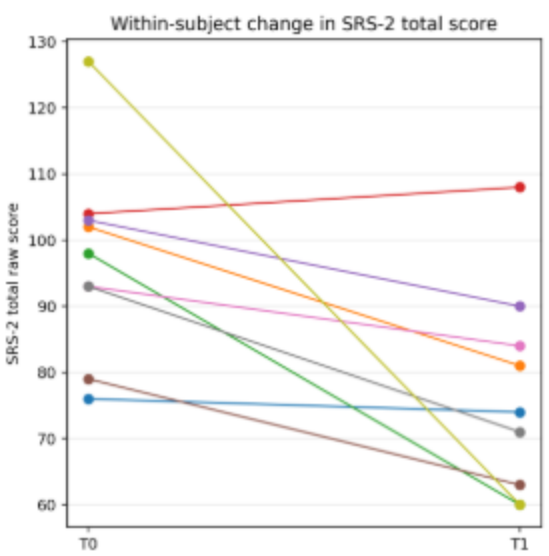


Figure 1. Overview of the SPARK study design, markerless video-analysis pipeline, and outcome measures.

Figure 1

Overview of the SPARK study design and multimodal assessment workflow. Ten autistic boys aged 7.5–10 years requiring level 1 support completed an eight-week structured soccer program. Standard RGB video acquired during the intervention was processed using markerless pose estimation to derive seven prespecified digital metrics. Caregiver-reported secondary exploratory outcomes were assessed with the Social Responsiveness Scale, Second Edition (SRS-2), and the Parenting Stress Index–Short Form (PSI-SF) The human figures shown in this schematic are AI-generated illustrative elements and do not depict study participants or images extracted from the study dataset..



Individual SRS-2 changes

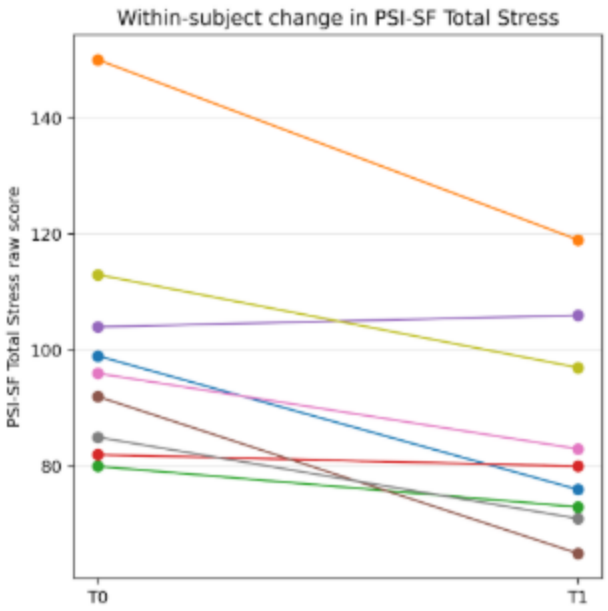
Participant	T0	T1	Δ (T1−T0)
	76	74	-2
	102	81	**-21**
	98	60	**-38**
	104	108	+4
	103	90	**-13**
	79	63	**-16**
	93	84	**-9**
	93	71	**-22**
	127	60	**-67**

Bold values indicate |Δ| > 5.61 raw points.

Figure 2

Individual SRS-2 total-score trajectories from T0 to T1.

Figure 2 legend. Each colored line represents one participant, with the corresponding colored marker shown in the adjacent table. The table reports individual T0 and T1 raw scores and the change calculated as Δ = T1 − T0; negative Δ values indicate improvement and positive values indicate worsening. Delta values shown in bold identify participants whose absolute change exceeded the provisional 5.61-point minimal clinically important difference (MCID) proposed by Tan et al. [25]. This threshold is presented as an exploratory reference and should not be interpreted as a universally validated SRS-2 MCID across intervention settings.



Individual PSI-SF changes

Participant	T0	T1	Δ (T1−T0)	RCI
●	99	76	-23	-1.89
●	150	119	**-31**	**-2.54**
●	80	73	-7	-0.57
●	82	80	-2	-0.16
●	104	106	+2	0.16
●	92	65	**-27**	**-2.21**
●	96	83	-13	-1.07
●	85	71	-14	-1.15
●	113	97	-16	-1.31

Bold = reliable change (|RCI| ≥ 1.96; raw-change threshold ≈ 23.9 points).

Figure 3

Individual PSI-SF Total Stress trajectories from T0 to T1 with exploratory Reliable Change Index (RCI) classification.

Figure 3 legend. Each colored line represents one caregiver-participant pair; the adjacent table reports T0 and T1 raw scores, Δ = T1 − T0, and the corresponding RCI. Negative Δ and RCI values indicate improvement. Bold values identify reliable change (|RCI| ≥ 1.96). The RCI was calculated using the study baseline SD (21.55) and the published PSI-SF Total Stress test-retest reliability coefficient (r = 0.84), corresponding to an approximate raw-score threshold of 23.89 points. One participant had different caregiver respondents at T0 and T1; excluding this pair did not alter the number classified as showing reliable improvement.