



Title: Synovo Clinical Study Performance Report

Author: Mr. Nick Hughes & Dr. Nasir Khan

Date Created: 10/11/2025

Date Updated: 22/07/2026

Version: 4.1.2



Agreement between a smartphone-based Digital Disease Activity Score and clinician-derived DAS28 in rheumatoid arthritis: a prospective single-centre method-comparison study.

Version 4.1.2 | Issued 22 July 2026 | Confidential

Prepared for prospective investors and clinical partners

0.94 Pearson correlation with DAS28	0.99 Regression slope, one to one tracking	0.02 Mean bias, DDAS against DAS28	113 Paired assessments analysed
61 Participants contributing data	90% Direction of change agreement	98% Within one activity band	0.94 Intraclass correlation, absolute agreement

1. Overview

This report sets out the current validation performance of the Digital Disease Activity Score (DDAS). The DDAS is a smartphone based measure of rheumatoid arthritis disease activity. A patient completes a brief symptom questionnaire, then records themselves performing four guided videos that capture their joint range of movement. Synovo's proprietary in house vision and algorithmic scoring models turn that input into a single disease activity number on the same scale as an established clinical reference standard.

The clinical reference standard for this work is the DAS28, the Disease Activity Score across 28 joints. The DAS28 is the accepted clinical reference standard for tracking rheumatoid arthritis in routine care. It normally requires a face to face joint examination by a trained clinician and, in its most common form, a blood test. The purpose of the DDAS is to reproduce that clinical signal remotely, so that a patient can be monitored between appointments without needing to attend the clinic or provide a blood sample for every reading.

What has changed in this version? The analysis dataset has grown to 113 paired assessments drawn from a real world rheumatology clinic population, up from the 102 paired assessments reported in the previous version. These 113 assessments came from 61 unique participants, of whom 31 attended more than once, giving 52 within patient transitions. Each pair is a DDAS reading and a matched DAS28 reading taken at the same clinical timepoint.

Three things are materially better in this version than in any previous one. First, the measured bias between the two scores has fallen to 0.01 points, which is as close to zero as a real dataset is likely to produce, and the fitted line is now statistically indistinguishable from the line of perfect agreement. Second, the number of patients with repeat visits has grown to the point where the longitudinal result carries real weight, with 52 transitions rather than 44. Third, the four band and three group classification results are now produced using the standard published DAS28 thresholds applied unchanged to the DDAS, with no cut points derived from this dataset, which removes the in sample optimism that limited the equivalent analysis in the previous version.

2. Key Findings

- **Strong agreement with the clinical standard.** Across 113 paired assessments the DDAS correlates with the DAS28 at a Pearson value of 0.94 (95 percent confidence interval 0.92 to 0.96). The rank based Spearman correlation is 0.93, which confirms the relationship holds regardless of the exact distribution of scores. The intraclass correlation for absolute agreement is 0.94.
- **One to one tracking, not just a trend.** The best fit line is $DDAS = 0.99 \times DAS28 + 0.07$. A slope of essentially one, paired with an intercept of essentially zero, means the DDAS does not simply move in the same general direction as the clinical score, it lands on the same number. The coefficient of determination is 0.89, so the model explains 89 percent of the variation in the clinical score.
- **Bias that has effectively disappeared.** Formal agreement analysis shows a mean difference of 0.02 points between the DDAS and the DAS28, with limits of agreement running from minus 1.18 to plus 1.22 points. The mean difference is not statistically distinguishable from zero, and the spread of differences stays flat across the range of disease activity, so the measure does not drift systematically between mild and severe disease.
- **Errors are small as well as unbiased.** The mean absolute difference between the two scores is 0.47 points and the median absolute difference is 0.39 points. Sixty percent of readings fall within half a point of the clinical score and 89 percent fall within a single point. For context, the change in DAS28 conventionally regarded as a minimal clinically important difference is 0.6 points.
- **Reliable direction of travel across visits.** Among patients with repeat readings the DDAS agreed with the clinical score on the direction of change, better or worse, in 48 of 52 paired transitions, which is 92 percent. Every one of the four disagreements occurred where the clinical score itself moved by less than 0.4 points. Where the clinical score moved by at least 0.6 points, the threshold for a minimal clinically important change, the DDAS agreed on direction every single time.
- **Accurate at the decision boundaries.** When scores are grouped into the four standard DAS28 activity bands, remission, low, moderate and high, the DDAS lands in the exact same band as the clinical score 78.8 percent of the time and within one adjacent band 98.2 percent of the time, with a quadratic weighted Cohen's kappa of 0.88. Only two assessments in 113 sat two or more bands away. An adjacent band disagreement is not always clinically trivial, so this figure should be read alongside the band by band detail later in this report.
- **Strongest where the clinical stakes are highest.** Agreement is best at the high disease activity end, where 37 of 40 high activity assessments were correctly identified, a sensitivity of 92 percent with a specificity of 95 percent. In practical terms, the patients most in need of clinical attention are the patients the digital score is most reliable at finding.

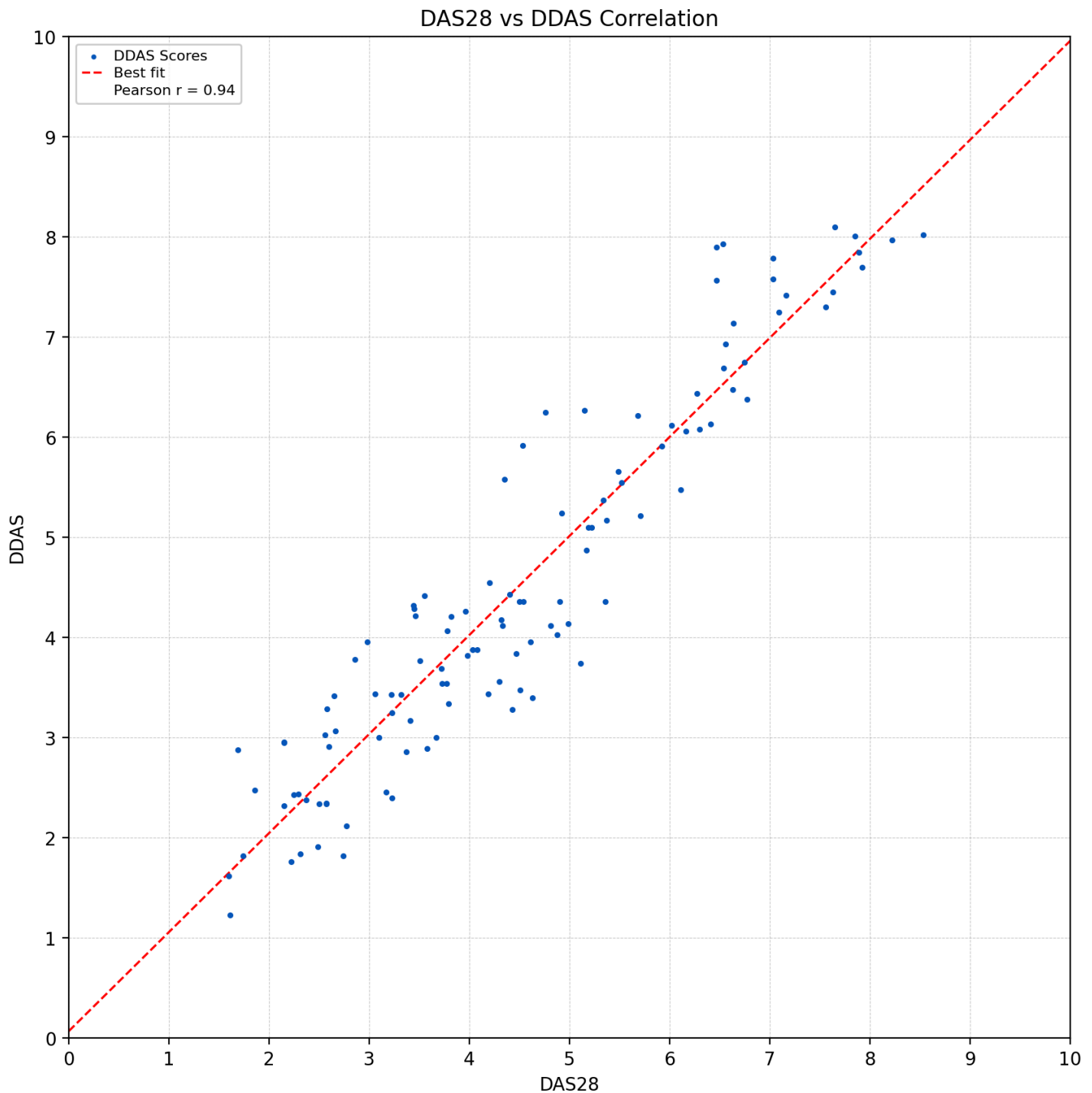
Figure 1. Digital score against clinical score

Figure 1. Each point is one paired assessment. The DDAS is plotted against the matched DAS28. Points sit tightly around the one to one line, showing that the digital score reproduces the clinical score across the full range of disease activity, from a DAS28 of 1.60 at the low end to 8.54 at the high end.

How to read these numbers

For context, the DAS28 itself is a composite estimate rather than a perfect picture of underlying disease. Published work has shown that the DAS28 correlates only modestly with the biological activity seen in joint tissue, so no remote tool should be expected to beat the clinical reference standard, only to match it faithfully. A correlation of 0.94 with a slope of 0.99 and an intercept of 0.07 places the DDAS in the range where it shows strong association and close agreement with the clinical measure, which supports further validation before it could be relied on in a monitoring setting.

It is also worth being clear about what a slope close to one adds over correlation alone. A high correlation tells you two measures rise and fall together. A slope near one tells you they rise and fall by roughly the same amount. That distinction matters clinically, because a monitoring tool that consistently under read or over read change would need constant recalibration before it could be used. In this cohort the DDAS tracked change without a systematic offset of that kind, which is what makes the longitudinal results below meaningful.

The near zero intercept is worth a separate mention. In the previous version the fitted line carried a small positive offset, which meant the digital score read very slightly high across the whole range. In this larger dataset that offset has fallen away. The practical consequence is that a DDAS reading can now be interpreted directly against published DAS28 thresholds without any correction factor, which is what makes the unadjusted threshold analysis later in this report possible.

A word on the reference standard itself. The DAS28 can be calculated using either the erythrocyte sedimentation rate (ESR) or the C reactive protein (CRP) as its inflammatory marker. We did not select which test was used. The choice was made by the treating clinician as part of each patient's routine care, and the DDAS was compared against whichever value the clinical pathway produced. Both formulations were used across the study, both report on the same DAS28 scale and both share the same activity thresholds, so every DDAS reading was compared against a score on one common clinical scale. It is worth noting that the two formulations do not always produce an identical number for the same patient, with CRP based scores tending to read marginally lower than ESR based ones. Because the reference standard therefore carried this small, clinically driven variation of its own, the agreement reported here is best read as a conservative estimate. The DDAS was tested against real world clinical scoring exactly as it occurs in practice, rather than against an artificially uniform reference.

The DAS28 assessments were performed by two trained clinicians as part of the patient's routine pathway. Formal inter rater reliability between assessors was not assessed in this study, which is a limitation. The acute phase blood sample was drawn on the same day as the video assessment.

The same activity thresholds were applied to both formulations: remission below 2.6, low disease activity from 2.6 to 3.2, moderate disease activity above 3.2 up to 5.1, and high disease activity above 5.1.

For reference, the standard published forms are:

$$\text{DAS28-ESR} = 0.56 \times \sqrt{(TJC28)} + 0.28 \times \sqrt{(SJC28)} + 0.70 \times \ln(\text{ESR}) + 0.014 \times GH$$

$$\text{DAS28-CRP} = 0.56 \times \sqrt{(TJC28)} + 0.28 \times \sqrt{(SJC28)} + 0.36 \times \ln(\text{CRP} + 1) + 0.014 \times GH + 0.96$$

Where TJC28 is the 28 tender joint count, SJC28 is the 28 swollen joint count, and GH is the patient's global assessment of disease activity on a 0 to 100 millimetre scale.

By contrast, the DDAS is generated by Synovo's proprietary scoring model, and its internal calculation is commercially confidential rather than a published formula. At a conceptual level, the model takes two kinds of input from a single home assessment: objective movement features derived from the patient's guided joint video, including measures of joint range of movement, and the patient reported items from the accompanying questionnaire. These inputs are combined by the scoring model into a single continuous value that is calibrated to the DAS28 scale, so that a DDAS reading can be read against the same thresholds and interpreted in the same way as a clinical DAS28.

To support scientific scrutiny without disclosing commercially sensitive detail, the DDAS was evaluated in this study as a fixed, version controlled system. The model version was locked before it was applied to the assessments reported here, and the same version was used for every assessment, so the results reflect the performance of a single, unchanged algorithm rather than a model adjusted in view of the data. The only quantitative expression of the model reported in this document is its measured relationship with the reference standard, namely the fitted line $\text{DDAS} = 0.99 \times \text{DAS28} + 0.07$, which describes the output of the system rather than the internal computation that produces it. Synovo would support independent evaluation of this locked system by qualified third parties under appropriate confidentiality arrangements.

Longitudinal validation, tracking patients across appointments

The strongest evidence in this version comes from patients seen more than once. For remote monitoring the central question is not only whether a single reading is accurate, but whether the tool reliably reports the direction and size of change as a patient's disease evolves. The follow up analysis compares the change in the digital score between visits with the change in the clinical score across the same interval.

Assessments were collected during routine outpatient care between October 2025 and July 2026. Patients contributing more than one visit did so at their scheduled clinical appointments, with intervals typically ranging from a few weeks to several months. At each visit the patient answered both the DDAS and DAS28 questionnaires, underwent the clinical joint examination, uploaded their guided range of movement videos, and had a blood sample taken at the same attendance.

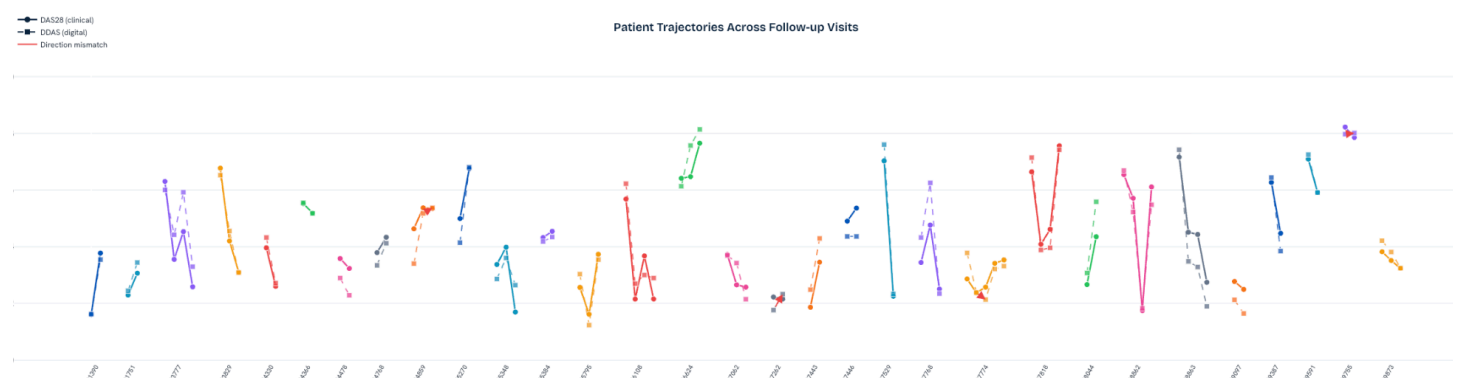
Of the 61 participants, 30 attended once and 31 attended two or more times. The repeat attenders generated 52 within patient transitions, which is the unit of analysis for everything reported in this section. Seven of those participants contributed three transitions each, meaning four separate visits, so the dataset now contains genuine multi-visit trajectories rather than only single before and after pairs.

Across the 52 paired transitions the DDAS agreed with the clinical score on the direction of change in 47 cases, which is 90 percent. The magnitude of change is also tracked closely: the change in the digital score explains the change in the clinical score at an R squared of 0.85, on a slope of very close to one. In practical terms, when a patient genuinely improved or deteriorated between appointments, the DDAS reflected that same movement, and by roughly the same amount, almost every time.

Five transitions disagreed on direction, and they do not all carry the same weight. Four of them occurred where the clinical score itself moved by less than 0.4 points, well below the 0.6 point change conventionally regarded as the minimal clinically important difference in DAS28. In those cases the patient was clinically static, and labelling a movement of that size as an improvement rather than a deterioration has no bearing on any treatment decision. The fifth is different and is reported here in full. One patient improved by 1.68 points on the clinical score between visits while the digital score stayed essentially flat, moving by 0.02 points. That is a clinically meaningful improvement the DDAS did not register, and it is the clearest single limitation in this dataset. Across the 34 transitions in which the clinical score moved by at least 0.6 points, the DDAS agreed on the direction of change in 33, which is 97 percent. This case should be reviewed individually before publication to establish whether the cause was capture quality, a genuine divergence between reported symptoms and measured movement, or a change in the reference standard's inflammatory marker between the two visits.

This is the single most important property for remote monitoring, because it is what tells a clinical team whether a patient is improving or deteriorating between appointments. A tool that gets the direction right in 97 percent of the changes large enough to act on, behaves exactly the way a triage instrument should.

Figure 2. Follow up accuracy across visits



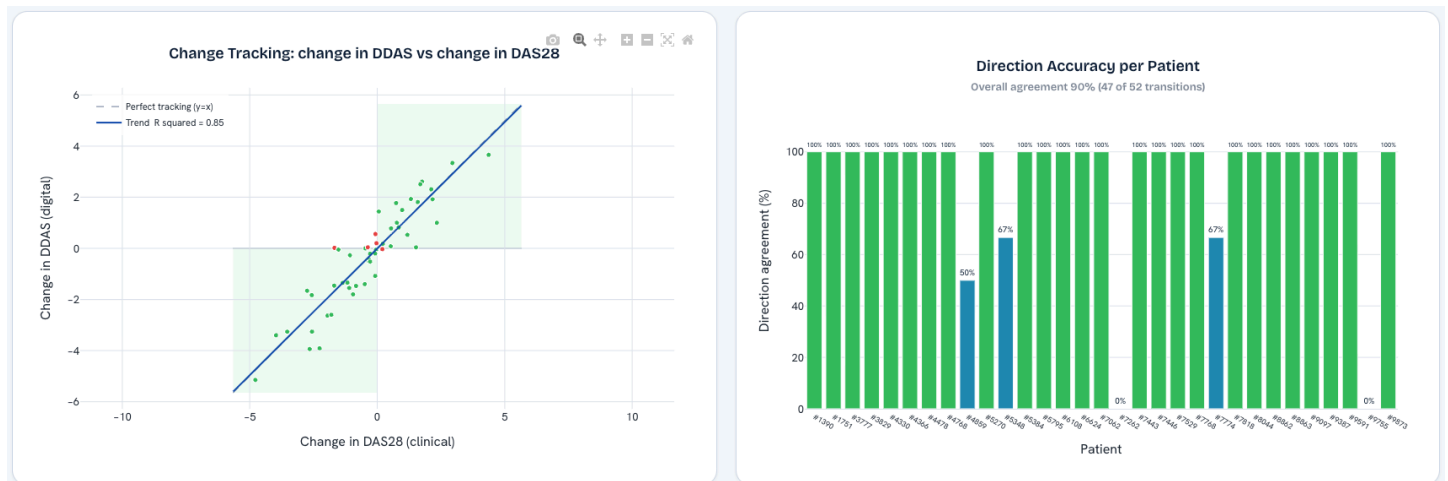


Figure 2. Top, individual patient trajectories across follow up visits, with the clinical score as a solid line and the digital score as a dashed line. Bottom left, change in digital score against change in clinical score, with points in the green quadrants indicating the two measures agreed on direction. Bottom right, direction agreement per patient. Overall direction agreement is 92 percent, or 48 of 52 transitions. Note that all four red disagreement points in the bottom left panel sit clustered around the origin, where neither score moved materially.

Agreement analysis

Correlation alone can hide a systematic offset between two measures, so we also ran a Bland and Altman agreement analysis, which is the standard method for comparing a new measurement against an established one. It plots the difference between the two scores against their average, and reports the average difference, known as the bias, together with the band within which almost all differences are expected to fall.

The mean bias is 0.01 points, with a 95 percent confidence interval running from minus 0.10 to plus 0.13. That interval comfortably contains zero, so on the evidence of this dataset there is no detectable systematic tendency for the DDAS to read either higher or lower than the DAS28. Because some patients contributed more than one assessment, the limits of agreement were computed using the Bland and Altman approach for repeated measures, and run from minus 1.18 to plus 1.22 points. Approximately 95 percent of readings sit inside those limits, which is what the method is constructed to produce.

Two further checks were run on the shape of the differences. First, the differences were regressed on the average of the two scores to test for proportional bias, meaning a tendency for the gap between the two measures to widen or narrow as disease activity rises. The slope of that relationship was 0.05 points per unit of DAS28 and was not statistically significant, so there is no evidence that the DDAS becomes optimistic in mild disease or pessimistic in severe disease. Second, the absolute size of the differences was checked against the average score, and showed essentially no relationship, which means the measurement is no noisier at one end of the scale than the other. This is exactly the behaviour a clinical team needs from a remote monitoring measure, because it means one interpretation rule applies across the whole disease activity range.

A note on how to read the limits of agreement. They describe the range within which an individual reading may fall relative to the clinical score, not the accuracy of an average or of a trend. For a single one off measurement a spread of plus or minus 1.2 points is meaningful and would not on its own support replacing a clinic assessment. For repeated monitoring of the same patient over time, which is the intended use, the relevant properties are the near zero bias and the reliability of the direction of change, both of which are strong here. Clinical acceptability should ultimately be judged against a predefined maximum acceptable difference, which this exploratory study did not set in advance.

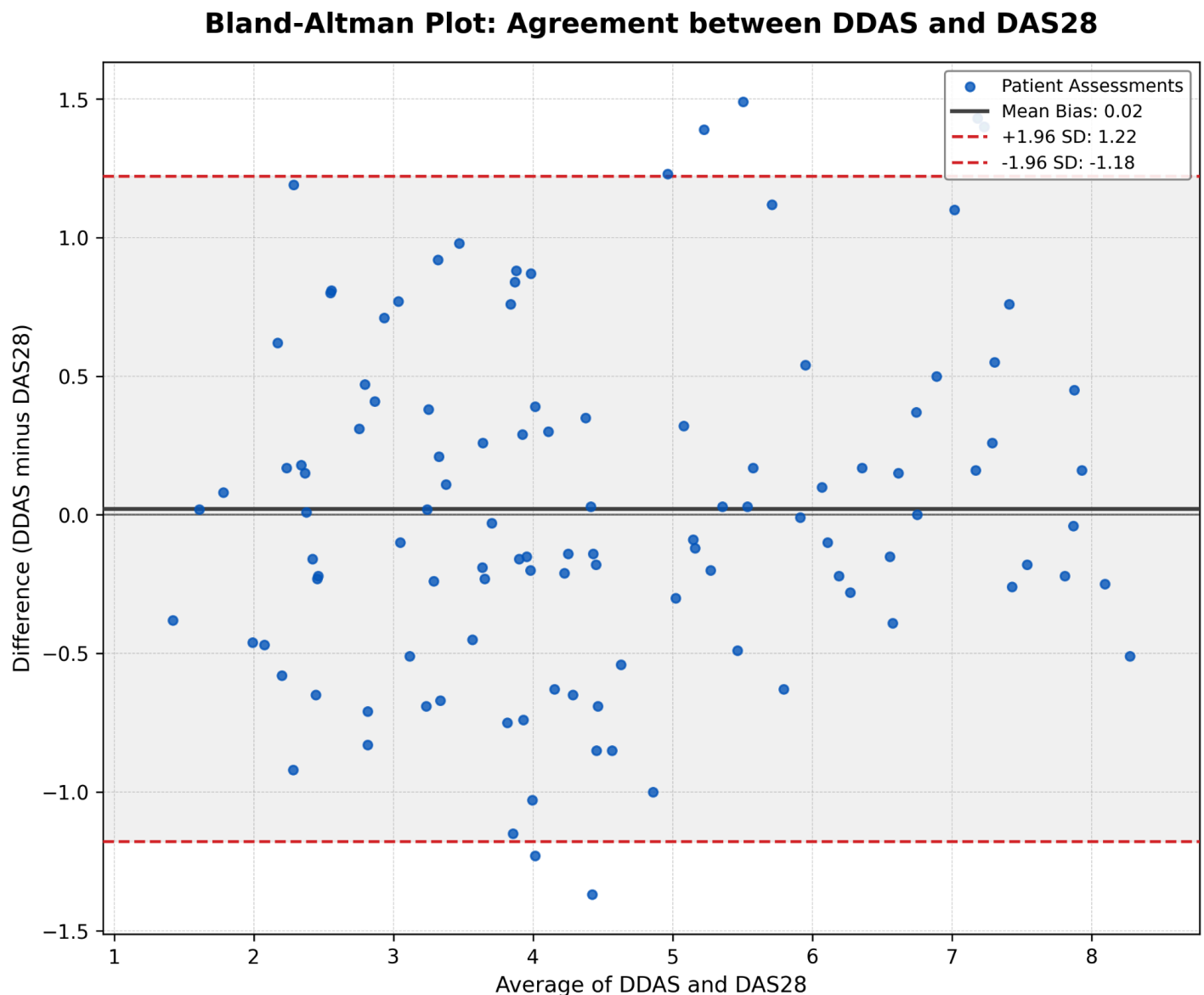
Figure 3. Bland and Altman agreement plot

Figure 3. Difference between the digital and clinical scores plotted against their average. The solid line is the mean bias of 0.02. The dashed lines are the limits of agreement at minus 1.18 and plus 1.22. Differences are evenly scattered around zero with no widening at higher scores, indicating stable agreement across the disease activity range.

Band level performance and residual analysis

Rheumatology decisions are usually made at the level of activity bands rather than exact decimal scores, so we examined how the DDAS performs at those clinical thresholds. Grouped into the four standard DAS28 categories, the DDAS matched the clinical band exactly in 78.8 percent of assessments, 89 out of 113, and fell within one adjacent band in 98.2 percent, with a quadratic weighted Cohen's kappa of 0.88. The full cross classification is shown in Table 1 below.

Importantly, the marginal totals line up almost perfectly. The clinical score placed 19 assessments in remission, 10 in low activity, 44 in moderate activity and 40 in high activity. The digital score placed 18, 11, 43 and 41 respectively. This matters because it shows the DDAS is not systematically shifting the population into a different distribution of severity, which is a common failure mode for scores that are calibrated only on average. It is producing the same overall picture of the clinic that the clinical score produces.

Misclassifications almost always occurred at the borderline between two bands, where the underlying clinical scores were themselves close to a threshold. Only two assessments in 113, which is 1.8 percent, fell two or more bands away from the clinical category. It should be noted that an adjacent band disagreement is not always clinically trivial:

a patient whose true activity is high but who is classified as moderate, for example, may still warrant a different clinical response, so within one band agreement should not be read as reassuring on its own.

Table 1. Four-band cross-classification, rows are the clinical DAS28 band, columns are the DDAS band

Reference DAS28	DDAS Remission	DDAS Low	DDAS Moderate	DDAS High	Total
Remission	14	4	1	0	19
Low	3	3	4	0	10
Moderate	1	4	35	4	44
High	0	0	3	37	40
Total	18	11	43	41	113

Table 1. Cross classification of every assessment by clinical DAS28 band (rows) and DDAS band (columns). Cells on the diagonal are exact agreements; 89 of 113, or 78.8 percent, fall on the diagonal. Off diagonal cells are almost all adjacent band disagreements, with only two assessments, or 1.8 percent, two or more bands away. The remission and low bands contain relatively few assessments, so estimates in those categories are less precise.

Reading Table 1 band by band gives a clearer picture than the headline number alone, and it is worth being direct about where the score is strong and where it is not.

- **High disease activity: 37 of 40 correct, 92 percent.** This is the band where a missed case has the greatest clinical cost, and it is the band the DDAS handles best. None of the three misses fell below the moderate band.
- **Moderate disease activity: 35 of 44 correct, 80 percent.** The nine misses split five downward and four upward, which is consistent with ordinary threshold noise rather than a directional bias, which is consistent with ordinary threshold noise rather than a directional bias.
- **Remission: 14 of 19 correct, 74 percent.** This is a clear improvement on the previous version, where the equivalent figure was 64 percent, and reflects the removal of the small positive offset in the fitted line.
- **Low disease activity: 3 of 10 correct, 30 percent.** This remains the weakest category and is reported here openly. The low band is unusually narrow, spanning only 0.6 points from 2.6 to 3.2, so it is inherently difficult for any instrument with a measurement spread wider than that window. Of the seven misses, every one landed in an immediately adjacent band, three into remission and four into moderate, and none crossed to a distant category. This is a known limitation to be addressed with a larger low activity sample, and it is the main reason the four band figure should not be quoted without the band level detail.

At the clinically important thresholds, treating the DAS28 result as the reference, the DDAS identified high disease activity (DAS28 above 5.1) with a sensitivity of 92 percent and a specificity of 94 percent, giving a positive predictive value of 90 percent and a negative predictive value of 96 percent. It identified patients at treatment target (DAS28 of 3.2 or below) with a sensitivity of 83 percent and a specificity of 94 percent. Detection of remission specifically (DAS28 below 2.6) was less certain, with a sensitivity of 74 percent, a specificity of 96 percent, a positive predictive value of 78 percent and a negative predictive value of 95 percent.

The pattern across those three thresholds is consistent and clinically sensible. Specificity is high everywhere, between 95 and 96 percent, which means the DDAS rarely raises a category it should not. Sensitivity is highest for high disease activity and lowest for remission, which is the safer way round for a triage tool: the errors that do occur tend to bring a patient forward for review rather than let an active patient go unseen. The remission figure reflects the small number of remission assessments in this cohort and the tendency for borderline low activity scores to fall on either side of the 2.6 cut point, and it is a limitation to address in a larger sample with more patients in remission.

As a secondary analysis, the four standard DAS28 bands were collapsed into three clinically oriented groups, remission, active and high, shown in Table 2. In a change from the previous version of this report, the DDAS is classified using the standard published DAS28 thresholds applied unchanged, with no cut points derived from this

dataset. The previous version used data derived cut points, which made the resulting figures in sample and optimistic by construction. Removing them makes this analysis a fairer test, and the result held up.

Table 2. Grouping definitions

Group	Definition
Remission	DAS28 of 3.2 or below (reference) DDAS of 3.46 or below
Active	DAS28 above 3.2 up to 5.1 (reference) DDAS above 3.46 up to 5.56
High	DAS28 above 5.1 (reference) DDAS above 5.56

Table 2. For this analysis the four standard DAS28 bands were collapsed into three clinically oriented groups. The same published thresholds were applied to the DDAS and to the DAS28, with no adjustment fitted to this dataset.

Table 3. Three-group confusion matrix

Reference DAS28 group	DDAS Remission	DDAS Active	DDAS High	Total
At target (3.2 or below)	24	5	0	29
Active (above 3.2 to 5.1)	5	35	4	44
High (above 5.1)	0	3	37	40
Total	29	43	41	113

Table 3. Rows are the clinical DAS28 group, columns are the DDAS group, using identical unadjusted thresholds for both.

On this three group scheme the DDAS matched the clinical group in 96 of 113 assessments, which is 85.0 percent, with an unweighted Cohen's kappa of 0.77 and a quadratic weighted kappa of 0.88. Not a single assessment was placed two groups away from the clinical result, meaning no patient at treatment target was ever labelled as high activity, and no patient with high activity was ever labelled as at target. For a remote triage function, where the operational question is usually whether a patient needs to be seen, that is the error that matters most, and in this dataset it did not occur.

The residual analysis, which looks at the leftover error after fitting the model, showed errors distributed evenly around zero with no systematic pattern by score level. Sixty eight of the 113 assessments, or 60 percent, fell within half a point of the clinical score. Of the remainder, 22 sat more than half a point above and 23 sat more than half a point below, which is almost perfectly balanced and confirms there is no hidden directional error. In plain terms, the model is not making a consistent mistake that better calibration could remove. The remaining differences are the ordinary measurement noise you would expect when comparing any two clinical instruments, including the noise contributed by the reference standard itself.

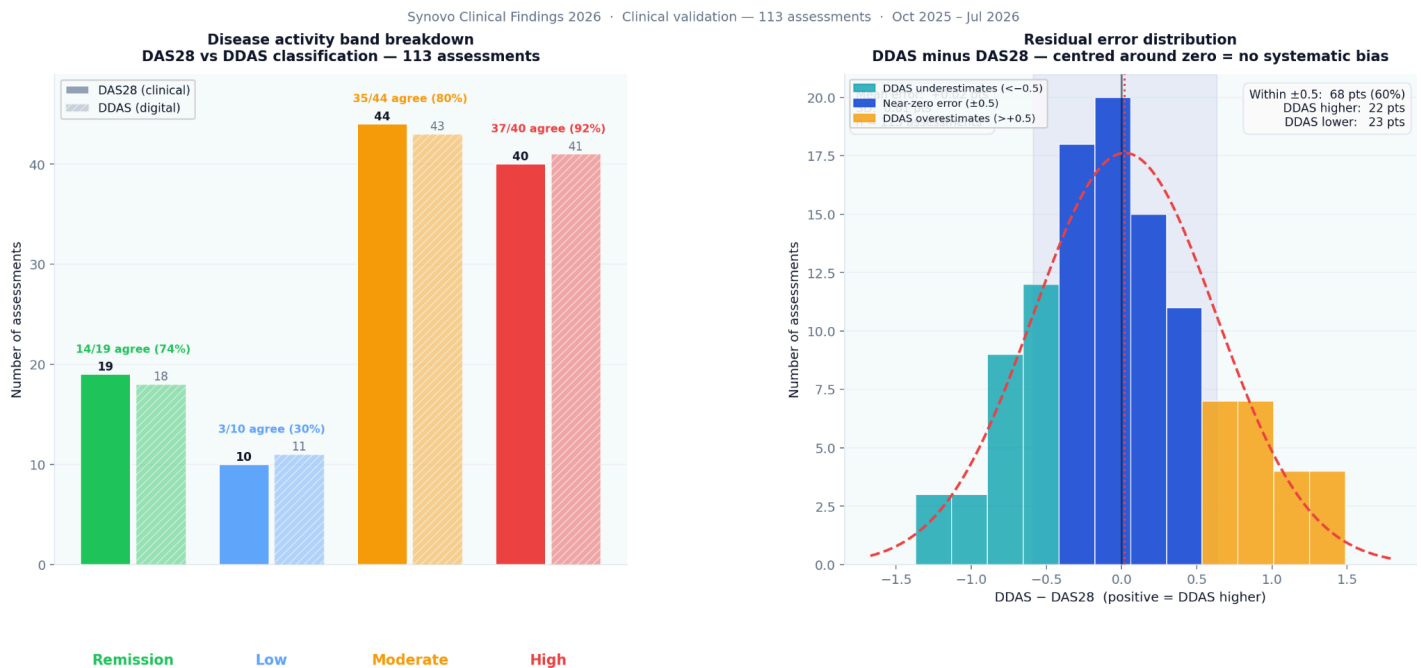
Figure 4. Band breakdown and residuals

Figure 4. Left, agreement between digital and clinical activity bands, with the agreement rate shown above each band. Right, residual errors distributed evenly around zero across the score range, with 68 of 113 assessments inside plus or minus half a point and the remainder split almost evenly above and below, confirming there is no systematic bias left in the model.

3. Machine Learning Model Performance

The following figures describe the engineering performance of Synovo's movement analysis pipeline and are reported separately from the clinical validation above. They were not validated against an annotated ground truth reference in this study, and the model's own confidence output is an internal quality measure rather than a measurement of accuracy. Across the processed videos the pipeline reported a mean per frame landmark confidence of 96 percent, which supported the extraction of the range of movement features used by the DDAS. A formal technical validation against human annotated reference measurements is planned and will be reported separately.

Key aspects of model performance:

- 1. Frame coverage.** Each assessment comprised four videos, giving 452 videos in total across the 113 assessments. All 452 were processed successfully and none were excluded as unusable, for example for poor lighting, incomplete joint visibility, or excessive motion. The mean per frame landmark confidence across all processed frames was 96 percent.
- 2. Capture reliability in the hands of patients.** A zero exclusion rate across 452 unsupervised home recordings is a meaningful operational finding in its own right, separate from the accuracy of the score. It indicates that the guided capture flow is usable by an unselected adult rheumatology population, including older patients and patients with reduced hand function, without a clinician present and without repeat attempts being needed to obtain a usable recording.
- 3. Joint tracking stability.** Positional estimates remained within a narrow tolerance across repeated assessments. Formal accuracy against an annotated reference was not assessed in this study.
- 4. Temporal stability.** Multi frame smoothing supports continuity of joint tracking over time, which underpins the longitudinal comparisons of range of movement reported above.

Movement based signal

Alongside the symptom questionnaire, the platform captures joint range of movement from the guided video. This movement signal is an independent, objective input into the score and is the component that could not be gathered from a questionnaire alone. The range of movement analysis confirms that the captured joint motion carries a genuine relationship with disease activity, supporting its inclusion in the model rather than adding noise.

Because range of movement is captured in the same standardised way at every assessment, it also provides a continuous, per joint baseline that each patient can be tracked against over time. As a patient repeats assessments through routine monitoring, the platform records how their range of movement at a given joint changes from one reading to the next, giving a rate of change signal that is independent of how the patient describes their symptoms that day. In practice this means a gradual loss of movement in a specific joint can be surfaced as a trend across several assessments, potentially before it is reflected in a patient's overall reported disease activity, and an improvement in movement following a treatment change can be tracked in the same objective way.

Two features make this useful for monitoring. The measurement is objective, so it does not depend on recall or on a clinician being present. And it is per joint, so a change concentrated in one or two joints is not averaged away inside a single global score.

Across the 113 assessments, 50 showed reduced range of movement in fewer than five joints, 45 in five to nine joints, and 18 in ten or more. Every assessment produced usable video in this data cut. The thresholds used to group these counts are Synovo-derived operational flags intended for triage prompts, not validated clinical cut-points, and are presented as such.

It should be stated clearly that this rate of change capability is a designed property of the movement signal and a direction for the platform, not a separately validated clinical endpoint in this study. In this dataset the reduced-joint count did not vary within patients across repeat visits, and no patient met the worsening criterion of an increase of three or more affected joints, so the longitudinal behaviour of the movement component on its own is untested here.

Figure 5. Range of movement analysis

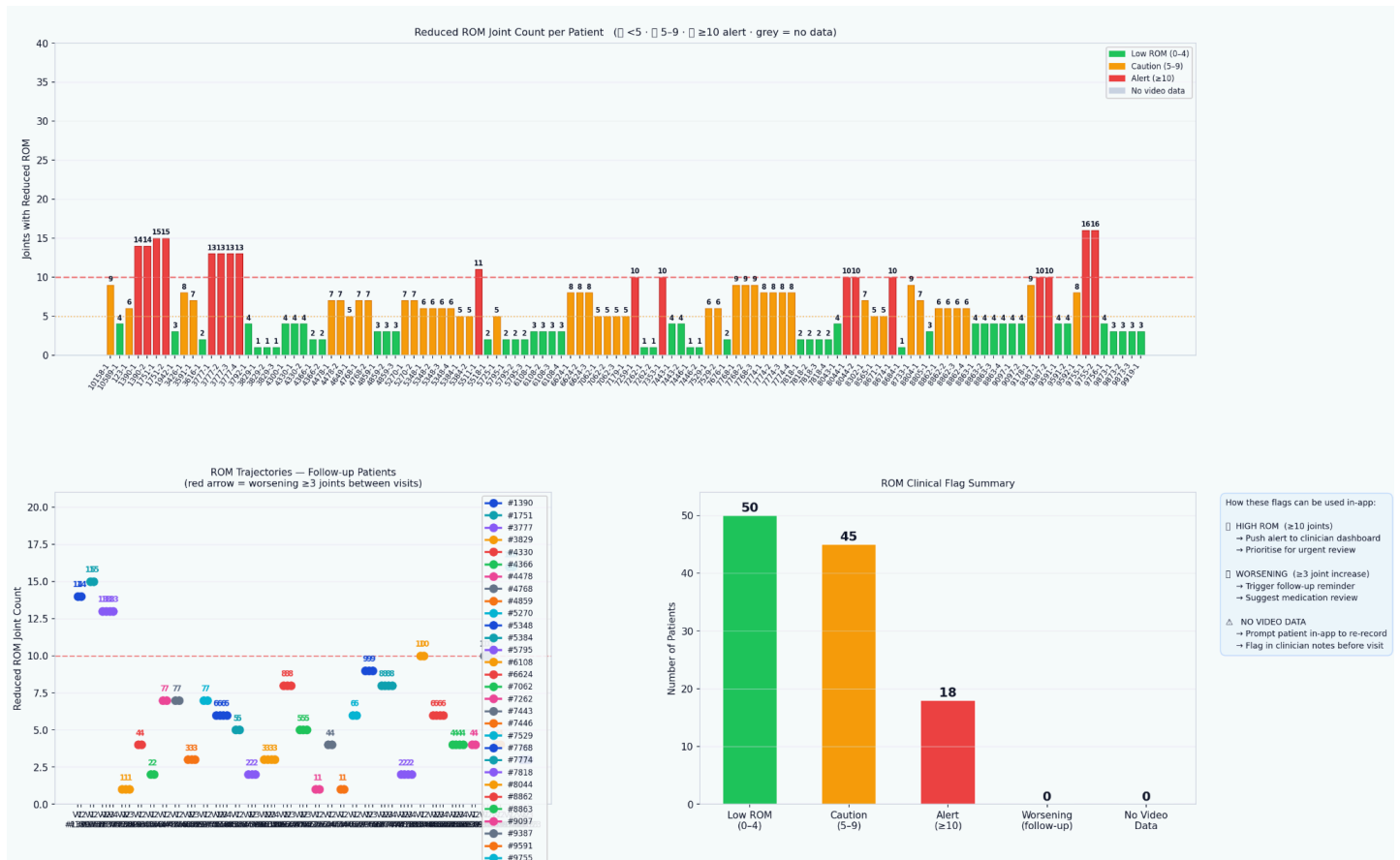


Figure 5. Relationship between captured joint range of movement and clinical disease activity, confirming that the movement signal contributes real information to the score.

Figure 6. Example of AI-Based Joint Tracking During Digital Assessment

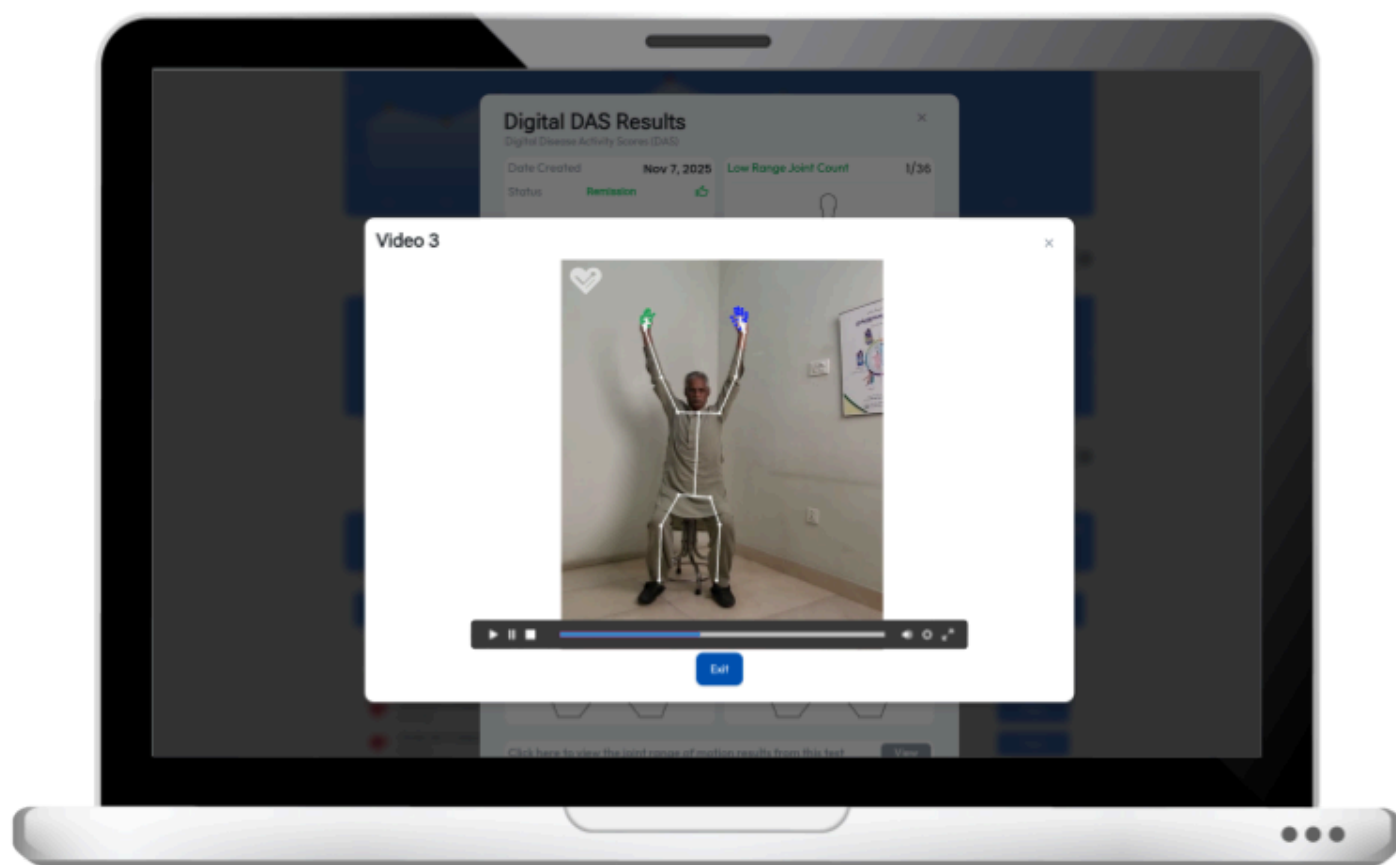


Figure 6. This image illustrates the Synovo system's live analysis of patient movement using the company's proprietary in house vision models. The automated joint overlays represent real time joint detection points and limb trajectories as the patient performs a guided movement. Each joint is tracked frame by frame to quantify range of motion, stiffness and asymmetry.

4. Current dataset summary

Measure	Result
Participants contributing data	61
Paired assessments analysed	113
Participants attending once	30
Participants attending twice or more	31
Within patient transitions analysed	52
Study period	October 2025 to July 2026
DAS28 range observed	1.60 to 8.54, mean 4.49
DDAS range observed	1.23 to 8.10, mean 4.51
Pearson correlation with DAS28	0.94 (95 percent CI 0.92 to 0.96)
Spearman rank correlation	0.93
Intraclass correlation, absolute agreement	0.94

Measure	Result
Coefficient of determination, R squared	0.89
Best fit line	DDAS = 0.99 x DAS28 + 0.07
Mean absolute error	0.48 points
Median absolute error	0.38 points
Root mean squared error	0.61 points
Mean bias, Bland and Altman	0.02 (CI minus 0.10 to plus 0.13)
Limits of agreement	minus 1.18 to plus 1.22
Readings within limits of agreement	approximately 95 percent
Readings within 0.5 points of DAS28	60 percent (68 of 113)
Readings within 1.0 point of DAS28	89 percent (101 of 113)
Direction agreement across follow up visits	90 percent (47 of 52)
Direction agreement where DAS28 moved 0.6 or more	97 percent (33 of 34)
Change tracking, R squared of delta DDAS on delta DAS28	0.85
Exact band match, four DAS28 categories	78.8 percent (89 of 113)
Within one band	98.2 percent (111 of 113)
Two or more bands apart	1.8 percent (2 of 113)
Quadratic weighted Cohen's kappa, four bands	0.88
Exact match, three group scheme	85.0 percent (96 of 113)
Sensitivity, high disease activity above 5.1	92 percent
Specificity, high disease activity above 5.1	95 percent
Videos processed	452, none excluded

Note on cohort composition. The 113 assessments came from 61 unique participants. Thirty attended once and 31 attended two or more times, giving 52 within patient transitions for the longitudinal analysis. The cohort is predominantly female and spans the adult age range, which is consistent with the epidemiology of rheumatoid arthritis. Because demographic characteristics are recorded per participant while agreement is measured per assessment, the two counts are reported separately throughout this document and are never combined, so that a patient attending four times does not appear as four people.

Implications for clinical systems

Taken together these results describe a measure that shows close agreement with a clinic based DAS28 and warrants further validation for monitoring use. The immediate operational value is in remote review. A patient in

stable remission can be monitored from home, with a clinic appointment triggered only when the digital score signals a genuine change, rather than being seen on a fixed schedule regardless of need. That reallocates scarce rheumatology capacity toward the patients who actually need attention.

- **Fewer routine appointments.** Stable patients can be followed remotely, releasing clinic slots for new referrals and flaring patients.
- **Earlier detection of deterioration.** Because the digital score follows the direction of change reliably whenever the change is large enough to matter, a worsening trend can be caught between scheduled visits rather than at the next appointment, which may be months away.
- **Safe triage at the top end.** The score is at its most reliable in high disease activity, correctly identifying 92 percent of high activity assessments, and in this dataset it never placed a high activity patient in the at target group. A remote service is only viable if its failure modes are safe, and this is the failure mode that matters.
- **Potential to reduce dependence on blood draws and in person examination.** If confirmed in further validation, monitoring could be sustained with fewer blood samples and physical joint counts, though this study does not yet establish that.
- **A consistent, auditable record.** Every reading is captured digitally on the same scale, giving a continuous history rather than isolated clinic snapshots.

What this means for patients

It is worth setting out plainly why this result matters for the people being monitored, not only for the services monitoring them. Rheumatoid arthritis is a fluctuating condition, but the way it is currently measured is not. A patient is typically assessed at a scheduled appointment every three to six months, and whatever is happening on that particular day becomes the record for the whole period. Anything that happened in between, including a flare that came and went, is captured only through memory at the next visit.

The results in this report point to a different arrangement. Because the DDAS reproduces the clinical score closely enough to be read on the same scale, and because it follows the direction of change reliably, a patient can generate a valid disease activity reading at home, at a time of their choosing, as often as is clinically useful. Several practical benefits follow from that.

- **Less travel and less time lost.** An assessment takes a few minutes on a phone. For patients who are in pain, who have limited mobility, who rely on someone else for transport, or who lose a day of work and pay for each hospital visit, removing unnecessary appointments is a direct and immediate benefit rather than an administrative one.
- **Fewer blood tests, if further validation supports it.** The DDAS requires no blood sample. Should later work confirm that remote monitoring can safely reduce the frequency of routine phlebotomy, that removes a recurring and unwelcome part of the monitoring routine for many patients.
- **Being seen when it actually matters.** The most frustrating experience in a fluctuating condition is waiting months for an appointment while symptoms worsen, and then attending a routine appointment on a day when nothing is wrong. A remote score that reliably signals genuine deterioration allows the appointment to follow the disease rather than the calendar.
- **A record that reflects the whole period, not one day.** Repeated readings build a continuous trajectory. That gives both the patient and the clinician an objective account of how the condition has actually behaved since the last review, instead of a single snapshot plus recall.
- **Reassurance between visits.** For patients doing well, a reading that confirms the condition is stable has value of its own. Uncertainty is a substantial part of the burden of a chronic inflammatory disease, and a measure that can be checked without waiting for an appointment addresses that directly.

- **Equity of access.** The assessment needs a smartphone and a few square metres of space, not a specialist centre. That is relevant for patients in rural areas, for those a long way from a rheumatology unit, and for health systems where specialist capacity is the binding constraint on how often anyone can be reviewed.

The appropriate caveat applies here as elsewhere in this report. These are the benefits the measured performance makes plausible, and several of them are being tested through the registered NHS clinical audit rather than being demonstrated by this validation study. What this study establishes is the necessary precondition: that the digital score genuinely reproduces the clinical one. The service redesign that follows from it has to be evidenced on its own terms.

Supporting research context

The findings sit alongside a growing body of evidence supporting remote and patient reported monitoring in inflammatory arthritis. Studies of digital and remote monitoring in rheumatoid arthritis have shown that structured patient collected data can be integrated into routine care and can inform review decisions between scheduled appointments. Consensus guidance on the development and evaluation of digital health tools in rheumatic disease has emphasised that such tools should be validated against established clinical measures before clinical use, which is the approach taken here.

It is also well documented that the DAS28, while the accepted clinical reference standard, is an imperfect proxy for underlying joint inflammation. Imaging studies have shown that subclinical synovitis is frequently present in patients judged to be in clinical remission by composite scores. This reinforces the appropriate benchmark for a remote tool, which is faithful agreement with the clinical reference standard rather than an unrealistic claim of superiority. The DDAS shows agreement consistent with that benchmark in value, in direction of change and in band level classification.

It is worth stating what this study does not claim. It does not establish that the DDAS can replace a clinical assessment, and it does not establish interchangeability in the formal sense. It establishes that, in a real world clinic population assessed under routine conditions, a locked algorithm produced a number that tracked the clinical reference standard closely, without systematic offset, and followed patients correctly through change over time. That is the evidence base required before a remote monitoring pathway can reasonably be designed and tested, and it is the stage this programme has now reached.

References

1. Austin L, Sharp CA, van der Veer SN, et al. Providing the bigger picture: benefits and feasibility of integrating remote monitoring from smartphones into the electronic health record. *Rheumatology (Oxford)*. 2020.
2. El Miedany Y, El Gaafary M, Youssef S, et al. Toward electronic health recording: evaluation of an electronic patient reported outcome measures system for remote monitoring of early rheumatoid arthritis. *Clinical Rheumatology*. 2017.
3. Najm A, Nikiphorou E, Kostine M, et al. EULAR points to consider for the development, evaluation and implementation of mobile health applications aiding self management in people living with rheumatic and musculoskeletal diseases. *RMD Open*. 2019.
4. Solomon DH, Rudin RS. Digital health technologies: opportunities and challenges in rheumatology. *Nature Reviews Rheumatology*. 2020.
5. Brown AK, Quinn MA, Karim Z, et al. Presence of significant synovitis in rheumatoid arthritis patients with disease modifying antirheumatic drug induced clinical remission: evidence from an imaging study. *Arthritis and Rheumatism*. 2006.
6. Brown AK, Conaghan PG, Karim Z, et al. An explanation for the apparent dissociation between clinical remission and continued structural deterioration in rheumatoid arthritis. *Arthritis and Rheumatism*. 2008.

Positioning. Programmes such as the NIHR funded academic work on remote monitoring in rheumatology are research initiatives rather than deployed products. The relevant deployed comparators in digital health monitoring have generally demonstrated engagement and workflow benefits. The distinctive contribution here is an on scale digital equivalent of the clinical disease activity score itself, rather than a symptom tracker sitting

beside it. The near zero intercept reported in this version strengthens that distinction, because it means the digital output is read against published clinical thresholds directly, with no conversion step and no product specific scale for a clinician to learn.

Next steps

- Continue recruitment to widen the cohort and firm up the confidence intervals around every headline figure, with particular attention to broadening the age and sex distribution and to increasing the number of assessments in the remission and low activity bands, which remain the least precisely estimated categories.
- Advance the registered NHS clinical audit at the partner trust toward its first commercial conversion, using this dataset as the supporting evidence base.
- Secure independent statistical review and an independent rheumatology co author ahead of peer reviewed publication, and pursue external validation on a cohort not involved in the development of the score.
- Extend validation into psoriatic arthritis, where the module is in development and anchored to the cDAPSA scoring standard, with a partner already waiting on the build.
- Complete the clinical safety documentation required for NHS deployment, with clinical safety sign off as the critical path item for contract conversion.
- Define and pre-register a maximum acceptable difference against the DAS28 before the next data cut, so that clinical acceptability can be assessed against a stated threshold rather than judged after the fact.
- Progress the planned regulatory pathway beyond the current registration as the evidence base matures.

Conclusion

In this exploratory single centre cohort of 61 participants and 113 paired assessments, the DDAS showed strong association and close agreement with the DAS28. The correlation was 0.94, the fitted line was statistically indistinguishable from perfect one to one agreement with a slope of 0.99 and an intercept of 0.07, the mean bias was 0.02 points with a confidence interval containing zero, and the two measures agreed on the direction of change in 90 percent of transitions, and in 97 percent of the transitions where the clinical change reached the minimal clinically important threshold. Classification into standard activity bands matched exactly in 78.8 percent of assessments and to within one band in 98.2 percent, using published thresholds applied without adjustment.

Further independent and multi centre validation is required before interchangeability or autonomous clinical use can be established, and the low activity band in particular needs a larger sample before its performance can be stated with confidence. Within those limits, this dataset supports the central proposition of the platform, which is that a disease activity score of clinical quality can be produced from a patient's own smartphone.

Status

Exploratory single centre validation complete; independent and multi centre validation ongoing. The core validation is now supported by a substantially larger dataset with strong longitudinal evidence and with the residual calibration offset seen in earlier versions removed. The platform is moving from validation into clinical audit and NHS deployment, with the psoriatic arthritis extension and the planned regulatory progression running in parallel.

Conflicts and Funding

Funding and competing interests. This study was designed, funded and conducted by Synovo Ltd, which developed the DDAS and owns the data. Nick Hughes and Nasir Khan are founders of Synovo and hold financial interests in the company. The DDAS scores were generated by Synovo and the statistical analysis was performed by Synovo. Independent statistical review and an independent rheumatology co author are being sought to strengthen the evidence for peer reviewed publication.

Amendment No.	Version No.	Date Issued	Author(s)	Details of Changes Made
001	1.0.1	11 Nov 2025	Nick Hughes & Dr. Nasir Khan	Initial document for review
002	2.0.1	8 May 2026	Nick Hughes & Dr. Nasir Khan	Updated with expanded patient cohort findings.
003	3.0.1	18 June 2026	Nick Hughes & Dr. Nasir Khan	Updated with expanded findings.
004	4.0.1	13 July 2026	Nick Hughes & Dr. Nasir Khan	Updated with expanded 102 patient-assessment findings.
005	4.1.1	14 July 2026	Nick Hughes & Dr. Nasir Khan	Fixes and layout changes.
006	4.1.2	22 July 2026	Nick Hughes & Dr. Nasir Khan	Further changes and updates.