JPSIM Journal of Pakistan Society of Internal Medicine

Original Article

Concurrent Validity And Longitudinal Responsiveness of A Smartphone-derived Disease Activity Measure In Rheumatoid Arthritis Using Clinician-assessed Das28 As The Reference Standard

Nasir Saleem Anwar Khan,[1] Nick Hughes,[1] Shehla Javed Akram,[2] Mohammad Faeq[2]

[1]Synovo Ltd, United Kingdom, [2]Aap Ka Clinic / Akram Medical Complex, Lahore, Pakistan

Abstract

Objective: To evaluate the concurrent validity, numerical agreement and longitudinal responsiveness of a smartphone-derived Digital Disease Activity Score (DDAS) against clinician-assessed DAS28 in patients with rheumatoid arthritis.

Methodology: A prospective, comparative study was conducted during routine rheumatology care in Lahore, Pakistan. Participants completed a structured digital questionnaire and four guided smartphone movement assessments. A trained rheumatologist independently completed the DAS28 reference assessment using ESR or CRP. All participants included in the principal analysis had a documented inflammatory marker and a complete DAS28 dataset. Primary outcome was to compare DDAS and DAS28 values and secondary analyses of threshold performance and change between repeat visits.

Results: A total of 113 paired assessments from 61 unique participants, of whom 31 contributed repeat observations, yielding 52 within-participant transitions. DDAS correlated with DAS28 at r=0.94 (95% CI 0.92-0.96), with Spearman rho=0.93 and an intraclass correlation coefficient of 0.94. The fitted relationship was DDAS = 0.99 x DAS28 + 0.07 (R^2=0.89). Mean Bland-Altman bias was 0.02 points, with limits of agreement from -1.18 to +1.22. Exact four-band agreement was 78.8%, and weighted kappa was 0.88.

Conclusion: The smartphone-derived measure demonstrated strong concurrent association with clinician-assessed DAS28, small average numerical bias and substantial categorical and longitudinal concordance in this exploratory single-centre cohort.

Keywords: Rheumatoid arthritis, DAS28, Concurrent validity, Digital health.

How to cite this:

Khan NSA, Hughes N, Akram SJ, Faeq M. Concurrent validity and longitudinal responsiveness of a smartphone-derived disease activity measure in rheumatoid arthritis using clinician-assessed DAS28 as the reference standard. J Pak Soc Intern Med. 2026;7(4): 294-299

Corresponding Author: Dr. Nasir Saleem Anwar Khan **Email:** nasir@synovo.ai

Received: 02-07-2026 **Revised:** 08-09-2026 **Accepted:** 28-10-2026 **DOI:** https://doi.org/10.70302/jpsim.v7i4.2659

2709-8729/© 2026 The Author(s). Published by JPSIM on behalf of Pakistan Society of Internal Medicine, Lahore, Pakistan. This is an open access article under the CC BY 4.0 license https://creativecommons.org/licenses/by/4.0/

Introduction

Rheumatoid arthritis (RA) is a chronic immune-mediated inflammatory disease in which persistent disease activity is associated with progressive joint damage, impaired function and disability. Contemporary management is based on treat-to-target principles, requiring structured and repeated assessment of disease activity and timely adjustment of treatment when the agreed target is not achieved.[1]

The Disease Activity Score in 28 joints (DAS28) is widely used to quantify RA disease activity. It combines tender and swollen joint counts, a patient global assessment and an inflammatory marker, either erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP).[2] Although established in clinical practice and research, repeated DAS28 assessment depends on trained clinical examination and access to laboratory testing. These requirements can limit the frequency with which formal disease-activity assessment is performed between scheduled reviews.[3]

Remote monitoring has therefore become an important research area in rheumatology. Previous studies have evaluated patient-reported outcome collection, smartphone self-assessment, patient-initiated follow-up, hand-function measures and digitally derived biomarkers.[4,5] These approaches suggest that clinically relevant information can be captured outside the traditional consultation, but they also show that remote tools require rigorous

evaluation against accepted clinical measures before being incorporated into care pathways.[6]

A further challenge is that RA disease activity is multidimensional. Pain, stiffness, functional restriction, joint movement and patient-reported symptoms may each contribute information about current disease status, while no single digital feature is expected to reproduce the complete clinical examination. A measurement approach that combines structured patient-reported information with guided movement-derived features may therefore provide a clinically relevant estimate of disease activity, but its measurement properties must be established before any proposed use in triage or remote surveillance.[7,8]

The Synovo Digital Disease Activity Score (DDAS) combines a structured questionnaire with four guided smartphone movement assessments and generates a continuous score on the DAS28 scale. The present study was designed as a clinical measurement study rather than an evaluation of a treatment intervention. Its primary objective was to assess concurrent validity between DDAS and clinician-assessed DAS28 obtained during the same clinical encounter. Secondary objectives were to characterise numerical and categorical agreement and to assess whether DDAS tracked within-participant change across routine follow-up visits.

Methodology

This was a prospective, single-centre, observational, non-interventional, blinded method-comparison study combining a cross-sectional paired-assessment component with a longitudinal repeat-visit component. The study was conducted during routine outpatient rheumatology care at Aap Ka Clinic, Lahore, Pakistan, in collaboration with Akram Medical Complex, between October 2025 and July 2026, in accordance with Protocol Version 1.2.1 and the associated Statistical Analysis Plan. The study did not assign or modify treatment.

The study protocol, Participant Information Sheet and Informed Consent Form were approved by the Aap Ka Clinic Institutional Review Board on 21 October 2025. Written informed consent was obtained before study-specific procedures. DDAS results were not used independently to guide diagnosis, treatment or other clinical decisions.

One participant, a 17-year-old female enrolled on 31 October 2025, was subsequently identified as having been recruited below the protocol minimum age of 18 years. This was recorded as a protocol deviation. The participant had a complete clinical dataset, including an inflammatory marker. The final manuscript will report the prespecified primary analysis together with a sensitivity analysis excluding this participant, and the governance disposition of the deviation will be documented in the study file.

The protocol specified recruitment of up to 75 unique participants with clinician-confirmed RA receiving routine care at the study site. Eligible participants were required to provide written informed consent and to complete the digital assessment independently or with reasonable non-clinical assistance. Participants could contribute additional paired assessments when returning for routine clinical follow-up. Apart from the single age-eligibility deviation described above, the analysis population was assembled according to the protocol-defined eligibility framework.

At each included assessment, participants completed a structured questionnaire covering patient-reported health, functional status, fatigue, morning stiffness, flare-related symptoms and joint-specific information. They also completed four guided smartphone movement assessments involving the hands and wrists, shoulders and elbows, and knees. Movement features derived from the videos were combined with questionnaire responses by the fixed DDAS scoring model to generate a continuous disease-activity value calibrated to the DAS28 scale.

The model version used throughout the study was SFL1.1.2. The model was locked before study processing, remained unchanged during the validation period, and no participant data from the validation cohort were used for model training, tuning or calibration. The proprietary internal scoring architecture is not disclosed, but model version, input domains, validation separation and study-level analytical methods are documented.

The reference assessment was clinician-derived DAS28 performed by a trained rheumatologist during the same clinical encounter. The assessment comprised 28-joint tender and swollen joint counts, patient global health and either ESR or CRP according to routine clinical practice. Standard disease-activity categories were used: remission <2.6, low disease activity 2.6-3.2, moderate disease activity >3.2-5.1 and high disease activity >5.1.

Although the protocol permitted marker-free clinical assessments to be retained for exploratory analyses where laboratory testing was not clinically indicated, no participant without an inflammatory-marker result was included in the final primary analysis. All 61 participants included in the current dataset had a documented ESR or CRP result supporting a complete clinician-derived DAS28 assessment. The

inflammatory-marker formulation used for each assessment was recorded.

Patient global health was obtained using a structured five-level question asking how active the participant's arthritis had been during the preceding week. Responses were mapped to the 0-100 global-health range. This discretised approach is a measurement characteristic of the present study and is considered when interpreting agreement with conventional DAS28 implementations.

The study used blinded measurement procedures. The clinician completing the DAS28 assessment did not have access to the DDAS result, and DDAS was generated without access to the clinician-derived DAS28. Each measure was finalised independently before linkage for analysis.

The prespecified primary outcome was Pearson correlation between paired DDAS and DAS28 measurements, reported with a 95% confidence interval. Secondary analyses included Spearman correlation, ordinary least-squares regression, intraclass correlation for absolute agreement, repeated-measures Bland-Altman analysis, four-band cross-classification with quadratic-weighted Cohen kappa, threshold sensitivity and specificity, and longitudinal comparison of the direction and magnitude of change between consecutive assessments.

Because repeated assessments were contributed by some participants, repeated-measures methods and participant-level sensitivity analyses were specified. Activity thresholds applied to DDAS were the published DAS28 thresholds and were not optimised on the validation dataset. A change of at least 0.6 DAS28 points was used as a clinically meaningful change threshold in the longitudinal descriptive analysis. Independent statistical review or reproduction is planned before journal submission.

Results

The current analysis comprised 61 unique participants contributing 113 paired DDAS-DAS28 assessments. Thirty participants contributed one assessment and 31 contributed two or more, generating 52 consecutive within-participant transitions. All 113 included paired assessments had complete DDAS and clinician-derived DAS28 data with a documented ESR or CRP result. All 452 guided videos, four per assessment, were processed without technical exclusion.

A complete participant-level baseline table, including age, sex, disease duration, treatment and distribution of ESR- and CRP-based DAS28 assessments, should be populated from the final locked dataset before submission. The single participant aged 17 years is included in the current descriptive count but will also be addressed through a sensitivity analysis excluding that participant.

Table 1: *Characteristics of Study Subjects*

Characteristic	Current value
Unique participants	61
Paired assessments	113
Participants with one assessment	30
Participants with repeat assessments	31
Within-participant transitions	52
Assessments with complete ESR/CRP-supported DAS28	113 (100%)
DAS28 range, mean	1.60-8.54± 4.49
DDAS range, mean	1.23-8.10±4.51

DDAS showed a strong positive association with clinician-assessed DAS28 (Pearson r=0.94, 95% CI 0.92-0.96; Spearman rho=0.93). The intraclass correlation coefficient for absolute agreement was 0.94. The fitted regression equation was DDAS = 0.99 x DAS28 + 0.07 (R^2=0.89), indicating similar scaling across the observed range of disease activity.

Repeated-measures Bland-Altman analysis showed a mean difference of 0.02 points between DDAS and DAS28, with 95% limits of agreement from -1.18 to +1.22. The mean absolute difference was 0.47 points and the median absolute difference was 0.39 points. Sixty percent of paired measurements were within 0.5 DAS28 points and 89% were within 1.0 point.

Table 2: *Repeated Measures outcomes of Data*

Measure	Result
Pearson correlation	0.94 (95% CI 0.92-0.96)
Spearman rank correlation	0.93
Intraclass correlation, absolute agreement	0.94
Regression	DDAS = 0.99 x DAS28 + 0.07
R^2	0.89
Mean bias	0.02 points
95% limits of agreement	-1.18 to +1.22
Mean absolute difference	0.47 points
Median absolute difference	0.39 points

Using the standard four DAS28 categories, DDAS and clinician-assessed DAS28 were in the same category in 89 of 113 assessments (78.8%). Quadratic-weighted kappa was 0.88. Most disagreements occurred between adjacent categories. Two assessments were separated by two or more categories. Exact agreement was highest in the high-disease-activity category and lowest in the narrow low-disease-activity interval between 2.6 and 3.2.

Table 3: *Comparison of DAS28 & DDAS Outcomes*

Clinical DAS28 category	DDAS remission	DDAS low	DDAS moderate	DDAS high
Remission (n=19)	14	4	1	0
Low (n=10)	3	3	4	0
Moderate (n=44)	1	4	35	4
High (n=40)	0	0	3	37

For high disease activity defined as DAS28 >5.1, DDAS had sensitivity of 92% and specificity of 95%. For the treatment target threshold of DAS28 <=3.2, sensitivity was 83% and specificity 94%. For remission, DAS28 <2.6, sensitivity was 74% and specificity 96%. These estimates are exploratory and should be interpreted with regard to the modest numbers within some disease-activity categories.

Table 4: *Clinical Threshold of DDAS*

Clinical threshold	Sensitivity	Specificity	PPV	NPV
High disease activity (>5.1)	92%	95%	90%	96%
Treatment target (<=3.2)	83%	94%	[final report]	[final report]
Remission (<2.6)	74%	96%	78%	95%

Across 52 consecutive within-participant transitions, DDAS and DAS28 changed in the same direction in 47 transitions (90%). Among the 34 transitions in which clinician-assessed DAS28 changed by at least 0.6 points, direction agreed in 33 (97%). Change in DDAS explained 85% of the variation in change in DAS28 (R^2=0.85). One transition showed a clinically meaningful DAS28 improvement of 1.68 points with minimal change in DDAS (0.02 points); this observation should be reviewed clinically and retained transparently in the final report.

Discussion

This study evaluated a smartphone-derived RA disease-activity measure as a measurement instrument rather than as a treatment or autonomous decision system. In a real-world outpatient cohort, DDAS demonstrated strong concurrent association with clinician-assessed DAS28, small average numerical bias, substantial categorical concordance and high directional agreement during clinically meaningful change.[9,10] These findings are encouraging but do not establish clinical interchangeability.

The statistical pattern is relevant because correlation alone is insufficient for method comparison. The regression slope close to one and the small mean Bland-Altman bias indicate similar average scaling, while the limits of agreement show that individual measurements can still differ materially. For clinical use, the acceptable magnitude of such disagreement must be defined prospectively with rheumatology input rather than inferred from the present dataset.[11]

For remote monitoring, the clinical consequences of classification error are at least as important as aggregate correlation. Sensitivity was highest for high disease activity, while exact classification was weakest in the narrow low-disease-activity band. This pattern is clinically plausible because the 2.6-3.2 interval spans only 0.6 DAS28 points, making boundary shifts more likely when two methods have non-zero measurement error. A remote measurement intended to support surveillance or triage would be expected to prioritise sensitivity to meaningful deterioration while avoiding false reassurance.[12,13] In this dataset, no participant at treatment target was classified as high activity and no high-activity assessment was classified as at target when the three clinically oriented groups were used. However, these findings require confirmation in a larger cohort enriched for remission, low disease activity and transitions across clinically important thresholds.

The observed limits of agreement, approximately +/- 1.2 DAS28 points around the mean difference, are important. Although average bias was minimal, individual differences of this magnitude may cross treatment-relevant category boundaries.[14] The present study did not prespecify a maximum clinically acceptable difference; establishing such a margin should be a priority for independent rheumatology-led validation.

Previous studies have shown that smartphone self-assessment, patient-reported outcome monitoring, patient-initiated follow-up and device-derived functional signals can contribute to RA monitoring. The present study differs in combining patient-reported information with guided movement-derived features and expressing the resulting measurement on

the DAS28 scale. Its contribution is therefore primarily one of measurement validity and longitudinal responsiveness rather than demonstration of a clinical pathway benefit.[15,16] Implementation studies in digital rheumatology also show that analytical validity is only one component of successful adoption. Adherence, usability, workflow integration, clinical safety, escalation rules and health-economic consequences require separate prospective evaluation. The current study should be interpreted as an early validation step within that broader evidence pathway.

DAS28 is an established composite clinical reference but is not a direct biological measure of synovitis. ESR- and CRP-based formulations are related but not numerically identical, and disease-activity categories can differ near thresholds. In this study the treating rheumatologist selected ESR or CRP according to routine practice, and every included assessment had a documented inflammatory marker. Separate ESR- and CRP-based sensitivity analyses should therefore be reported in the final statistical package where sample size permits. The patient-global component was collected using a five-level scale mapped to the conventional 0-100 range rather than a continuous 100-mm visual analogue scale.[17,18] This pragmatic approach should be retained as a measurement limitation because it may reduce granularity in the reference score.

Strengths include prospective paired assessment, independent completion of the digital and clinician-derived measures before linkage, use of a fixed model version, complete inflammatory-marker data for the principal analysis, inclusion of repeated routine-care observations, application of published DAS28 thresholds rather than validation-set-optimised cut-points, and use of complementary continuous, categorical and longitudinal analyses.[19,20]

The study was conducted at a single centre with a modest sample and limited numbers in remission and low disease activity. Formal inter-rater reliability of the clinician reference assessment was not evaluated. The study was designed and funded by the developer of DDAS, and the initial analyses were performed by the sponsor; independent statistical review and independent clinical review are therefore important before submission. One participant was enrolled below the protocol minimum age and will be addressed through sensitivity analysis and transparent protocol-deviation reporting. The scoring model is proprietary, limiting full algorithmic reproducibility.[21,22] Finally, the study evaluates measurement performance and does not establish that DDAS can replace clinician assessment, direct treatment decisions, reduce visits or improve outcomes.

The appropriate next step is independent multicentre validation involving rheumatologists and, where relevant, rehabilitation or musculoskeletal functional-assessment expertise.[23,24] Future studies should prospectively define acceptable measurement error, examine performance separately in ESR- and CRP-based DAS28, enrich recruitment around treatment-relevant thresholds, evaluate responsiveness during true flare and improvement, and test whether use within a clinical pathway can safely support prioritisation of review. Clinical utility, patient acceptability and health-economic impact should be evaluated separately from measurement validity.

Conclusion

In this exploratory prospective cohort, a smartphone-derived RA disease-activity measure showed strong concurrent association with clinician-assessed DAS28, minimal average bias and substantial categorical and longitudinal concordance. The results support further independent rheumatology-led multicentre validation. They do not establish clinical interchangeability, replacement of clinician assessment or autonomous clinical use.

Ethical Approval: The IRB/EC approved this study via letter no. AKC/25/SYN0726 dated October 21, 2025.

Conflict of Interest: None

Funding Source: The study was sponsored and funded by Synovo Ltd, developer of the DDAS platform.

Authors' Contribution

NSAK: Conception.

SJA: Design of the work.

NH, MF: Data acquisition, analysis, or interpretation.

NH, SJA, MF: Draft the work.

NSAK: Review critically for important intellectual content.

All authors approve the version to be published.

All authors agree to be accountable for all aspects of the work.

References

1. Wells G, Becker JC, Teng J, Dougados M, Schiff M, Smolen . Validation of the 28-joint Disease Activity Score (DAS28) and EULAR response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis. Ann Rheum Dis. 2009;68(6):954-60.
2. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham C. rheumatoid arthritis classification criteria: an ACR/EULAR collaborative initiative. Arthritis Rheum. 2010;62(9):2569-81.
3. Moreland LW, O'Dell JR, Paulus HE, Curtis JR, Bathon JM, St. Clair EW, et al. A randomized comparative effectiveness study of oral triple therapy versus etanercept plus methotrexate in early aggressive rheumatoid arthritis: the TEAR trial. Arthritis Rheum. 2012;64(9):2824-35.
4. Smolen JS, Aletaha D, Bijlsma JWJ. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis. 2010;69(4):631-7.
5. Orbai AM, Bingham CO III. Patient reported outcomes in rheumatoid arthritis clinical trials. Curr Rheumatol Rep. 2015;17(4):28.
6. Nishiguchi S, Ito H, Yamada M, Yoshitomi H, Furu M, Ito T, et al. Self-assessment tool of disease activity of rheumatoid arthritis by using a smartphone application. Telemed J E Health. 2014;20(3):235-240.
7. Nishiguchi S, Ito H, Yamada M, Yoshitomi H, Furu M, Ito M, et al. Self-assessment of rheumatoid arthritis disease activity using a smartphone application: development and 3-month feasibility study. Methods Inf Med. 2016;55(1):65-9.
8. Espinoza F, Le Blay P, Coulon D, Lieu S, Munro J, Jorgensen C, Pers YM. Handgrip strength measured by a dynamometer connected to a smartphone: a new applied health technology solution for self-assessment of rheumatoid arthritis disease activity. Rheumatol (Oxford). 2016;55(5):897-901.
9. van Riel P, Alten R, Combe B, Abdulganieva D, Bousquet P, Courtenay M, et al. Improving inflammatory arthritis management through tighter monitoring and innovative electronic tools. RMD Open. 2016;2(2):e000302.
10. Seppen BF, L'ami MJ, Duarte Dos Santos Rico S. Feasibility of self-monitoring rheumatoid arthritis with a smartphone app: results of two mixed-methods pilot studies. JMIR Form Res. 2020;4(9):e20165.
11. Seppen B, Wiegel J, ter Wee MM. Smartphone-assisted patient-initiated care versus usual care in rheumatoid arthritis and low disease activity: a randomized controlled trial. Arthritis Rheumatol. 2022;74(11):1737-45.
12. Colls J, Lee YC, Xu C, Corrigan C, Lu F, Marquez-Grap, et al. Patient adherence with a smartphone app for patient-reported outcomes in rheumatoid arthritis. Rheumatology (Oxford). 2021;60(1):108-12.
13. Lee YC, Lu F, Colls J, Luo D, Wang P, Dunlop DD, et al. Outcomes of a mobile app to monitor patient-reported outcomes in rheumatoid arthritis: a randomized controlled trial. Arthr Rheumatol. 2021;73(8):1421-9.
14. Najm A, Nikiphorou E, Kostine M. EULAR points to consider for development, evaluation and implementation of mobile health applications aiding self-management in rheumatic and musculoskeletal diseases. RMD Open. 2019;5:e001014.
15. Creagh AP, Hamy V, Yuan H, Mertes G, Tomlinson R, Chen WH, et al. Digital health technologies and machine learning augment patient-reported outcomes to remotely characterise rheumatoid arthritis. NPJ Digit Med. 2024;7(1):33.
16. Arumalla N, Chan CK, Gibson M, Man YL, Adas MA, Norton S, et al. Clinical impact of electronic patient-reported outcome measures in remote monitoring of inflammatory arthritis: systematic review and meta-analysis. Arthritis Rheumatol. 2023;75(11):1892-903.
17. Austin L, Sharp CA, van der Veer SN. Providing the bigger picture: benefits and feasibility of integrating remote monitoring from smartphones into the electronic health record. Rheumatol (Oxford). 2020;59(2):367-78.
18. Watson K, Sheldon H, Pallari E, Arumalla N, Olive RR, Boiko O, et al. Remote monitoring of patients with rheumatoid arthritis in a low disease activity state: a mixed-methods evaluation across six hospitals in London, UK. Rheumatol (Oxford). 2024;63(10):2721-33.
19. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-282.
20. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS). J Clin Epidemiol. 2011;64(1):96-106.
21. Collins GS, Dhiman P, Andaur Navarro CL. Protocol for development of a reporting guideline for artificial intelligence-based clinical prediction models: TRIPOD+AI. BMJ. 2024;385:e078378.
22. Naeem F, Khan SE, Saeed MA, Farman S. Diagnostic and therapeutic delay in rheumatoid arthritis patients: impact on disease outcome. Pak J Med Sci. 2021;37(4):1001-7.
23. 32. Wagan AA, Raheem A, Bhatti A. Fatigue assessment by FACIT-F scale in Pakistani cohort with rheumatoid arthritis: the FAF-RA study. Pak J Med Sci. 2021;37(4):1025-30.
24. 33. Jalil F, Yaqoob M, Khaliq A. Progression pattern of rheumatoid arthritis: a study of 500 Pakistani patients. Pak J Pharm Sci. 2017;30(4):1219-23.