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.

Nocturnal scratch and sleep disturbance in patients with atopic dermatitis: Performance of digital outcomes from baricitinib phase III clinical trials

Ju Ji
Eli Lilly and Company
Jonathan I. Silverberg
jonathanisilverberg@gmail.com
George Washington University
Amy S. Paller
Feinberg School of Medicine, Northwestern University
Emma Guttman-Yassky
Mount Sinai Hospital
Hui Zhang
Eli Lilly and Company
Brian Winger
Eli Lilly and Company
Lisa Cirri
Eli Lilly and Company
Najwa Somani
Eli Lilly and Company
Sreekumar Pillai
Eli Lilly and Company

Article

Keywords:

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11514435/v1

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

Additional Declarations: Competing interest reported. • Jonathan I. Silverberg: has received honoraria as a consultant and/or advisory board member for Abbvie, Aldena, Amgen, AObiome, Apollo, Arcutis, Arena, Attovia, Boehringer-Ingelheim, Bristell-Meyers Squibb, Castle Biosciences, Connect Biopharma, Corevitas, Eli Lilly, Enveda, FIDE, Formation Bio, Galderma, Genzyme, GlaxoSmithKline, Immunocore, Incyte, Infinimmune, Imagene, Kowa Pharmaceuticals, Leo Pharma, Merck, Nanook, Nektar, Novartis, Organon, Pfizer, RAPT, Recludix, Regeneron, Sandoz, Sanofi, Shaperon, TARGET-RWE, Teva, Triton, Triveni, UCB, Union, UpToDate; speaker for Abbvie, Arcutis, Eli Lilly, Galderma, Organon, Pfizer, Regeneron, Sanofi-Genzyme; and institution received grants from Galderma. • Amy S. Paller has served as investigator for AbbVie, BioMendics, Dermavant, Eli Lilly, Incyte, Johnson & Johnson Innovative Medicine, Regeneron, UCB; as consultant for Abeona, Arcutis, BioCryst, Boehringer Ingelheim, Castle Creek, Chiesi, Dermavant, Johnson & Johnson Innovative Medicine, Krystal, LEO, Lilly, L'Oreal, MoonLake Immunotherapeutics, Peltheos, Quoin, Regeneron, and Sanofi; and on the data safety monitoring board for AbbVie, Abeona, Biocryst, Daiichi Sankyo, and Galderma. • Emma Guttman-Yassky has, in the past 36 months, received grants or contracts from Regeneron, Sanofi, Eli Lilly, Pfizer, Amgen, GSK, Sanofi, Bristol Meyers Squibb, AnaptysBio, AbbVie, Opsidio, Apollo Therapeutics, Aclaris Therapeutics, Concerto Biosciences, Apogee, Enveda Biosciences, Nektar, Triveni Bio, Sitryx, Concert Pharmaceuticals, and Target PharmaSolutions; consulting fees from AbbVie, Aclaris Therapeutics, Almirall, Amgen, AnaptysBio, Apogee Therapeutics, Apollo Therapeutics, Arcutis, Artax Biopharma, Astria, Bambusa, Biolojic Design, Bionyra Pharma, Boehringer-Ingelheim, Bristol Myers Squibb, Canyon Therapeutics, Celldex, Cue Biopharma, CytoReason, DBV, Dualitas Therapeutics, Egle Therapeutics, Electra Therapeutic, Eli Lilly, EMD Serono, Enveda Biosciences, Evommune, Galderma, Generate Biomedicines, Gilead Sciences, GSK, Imagene, Infimmune, Janssen Biotech, Jasper Therapeutics, Kymera Therapeutics, Leo Pharma, Matchpoint Therapeutics, Merck, Navigator Medicines, Nektar Therapeutics, Opsidio LLC, Pfizer, Propeller Bio, Pierre Fabre, Prologue Medicines, Proteologix, Priovant Therapeutics, RAPT, RayThera, Recludix Pharma, Regeneron, Rubedo Life Sciences, Sanofi, SATO, Sitryx Therapeutics, SPARC, Takeda, Teva, Third Arc Bio (TAB), TRex Bio, Triveni Bio, UCB, Vellera Therapeutics, Ventyx, Veradermics, and VRG Therapeutics. • Ju Ji, Hui Zhang, Brian Winger, Lisa Cirri, Najwa Somani and Sreekumar Pillai are employees of, and minor shareholders in, Eli Lilly and Company.

Nocturnal scratch and sleep disturbance in patients with atopic dermatitis: Performance of digital outcomes from baricitinib phase III clinical trials

Ju Ji[1], Jonathan I. Silverberg[2*], Amy S. Paller[3], Emma Guttman-Yassky[4], Hui Zhang[1], Brian Winger[1], Lisa Cirri[1], Najwa Somani[1], Sreekumar Pillai[1]

[*] Corresponding author (ORCID ID: 0000-0003-3686-7805).
Corresponding author email address: jonathanisilverberg@gmail.com

[1]Eli Lilly and Company, Indianapolis, Indiana, USA
[2]George Washington University School of Medicine and Health Sciences, Washington DC, USA
[3]Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA
[4]Mount Sinai, New York City, New York, USA

ABSTRACT

In clinical trials, itch and sleep disturbances associated with atopic dermatitis (AD) are typically assessed using subjective patient-reported outcomes (PROs). This study explored the clinical validity of algorithms that derive digital scratch and sleep outcomes from actigraphy data in patients with AD. Algorithms that had previously shown good agreement with video annotation/polysomnography ("gold standard") data were applied to actigraphy data from 664 adults with moderate-to-severe AD in two phase III monotherapy studies of baricitinib (BREEZE-AD1 and -AD2). The algorithm-determined percentage of the "total sleep opportunity" (period spent sleeping/attempting to sleep) spent asleep was moderately correlated with nocturnal scratch ($r=-0.45$), with scratch explaining ≈25% of sleep variation. Patients in the lowest quartile for mean hourly nocturnal scratch duration spent 23% more of their total sleep opportunity asleep, and 107 fewer minutes/night awake after initially falling asleep, than those in the highest quartile. Digital outcomes and PROs for itch/scratch and sleep disturbance were weakly correlated ($r<0.30$). Both PROs and digital endpoints significantly improved with baricitinib 4 mg whereas placebo treatment led to improvements in PROs but no improvement of digital outcomes. Findings from this evolving technology indicate that objective digital endpoints can provide clinically relevant insights that complement PROs.

(195 words)

Introduction

Patients with atopic dermatitis (AD) commonly experience intense itching at night, which leads to nocturnal scratch, frequent awakenings and poor sleep quality[1,2]. The combination of itch, scratch, pain, and sleep disturbance creates a profound symptom burden that has a detrimental effect on multiple domains of patient quality of life[3–5]. Identifying and reducing nocturnal scratch is crucial for improving sleep and overall wellbeing in patients with AD.

Patient-reported outcomes (PROs) are frequently used to assess nocturnal scratch and sleep in patients with AD. PROs provide value in clinical research studies because they offer a patient-centric perspective on disease severity and symptom burden[5]. However, the subjective nature of PROs introduces several potential limitations and biases across a range of disease conditions. These include patients' tendency to overestimate treatment effects[6]; fluctuations in patients' mood and expectations[6]; problems with respondent burden and missing data[7]; language barriers[8]; patient recall bias, noise, and errors[9]; and the potential for item bias[3]. In relation to AD specifically, many patients are unaware that they are itching and scratching during their sleep[10], and this limits the reliability and utility of PROs for the assessment of nocturnal itch and scratch.

Digital assessments may avoid many of the limitations of PROs and have the potential to provide complementary valuable information that may contribute to an understanding of the overall disease course. Specifically, within the context of AD, digital outcomes offer an objective, passive, continuous (and typically noninvasive) means of quantifying both the amount of nocturnal scratching (an indirect correlate of itch in pruritic dermatoses[11]) and sleep quality. High-resolution digital wrist actigraphy (three-dimensional accelerometer sensors sampled at ≥20 Hz) is a promising method of collecting such data, and several studies have demonstrated good agreement between nocturnal scratch and sleep outcomes derived from actigraphy data and both video recordings of patients' activities and polysomnography data[12–15]. It should be noted that digital actigraphy devices—as used in the current study—are considerably more accurate than those based on older analog technology (e.g., Schulman et al.[16] as cited in Neishabouri et al.[17]).

To ensure that raw digital actigraphy data are translated into meaningful endpoints, the algorithm used to interpret the data must be able to discriminate among different patient activities with high accuracy, sensitivity, specificity, precision, and F1 score (a measure that encompasses both sensitivity and precision) in real-world situations[18]. To this end, the algorithm developed by Ji et al.[12] addressed some of the challenges that had remained unmet by groups working in this area. These include limitations caused by imbalanced data in the existing literature, failure to capture finger scratches, and limited generalizability to real-world applications. Using actigraphy, highly predictive topological features, and a model-ensembling approach (i.e., combining multiple individual models to create a single, better, predictive model), Ji et al.[12] developed a method of measuring nocturnal scratch duration and intensity that showed good agreement with manually annotated videos, along with the following performance parameters: area under the curve, 0.77; accuracy, 0.76; positive predictive value, 0.35; and negative predictive value, 0.90.

In the *post hoc* analysis reported herein, we used data from two independent phase III clinical trials (BREEZE-AD1 [NCT03334396][19] and BREEZE-AD2 [NCT03334422][20]) to further explore clinical validation of the scratch detection algorithm developed by Ji et al.[12] A publicly available sleep detection algorithm[14]

was used to analyze the sleep-related outcomes. The current study had several aims: to evaluate the relationship between actigraphy-derived digital outcome measures of nocturnal scratching and sleep disturbances, to compare these findings with those from PROs that evaluate similar endpoints, and to provide information about the clinical utility of the scratch algorithm employed.

Results

Analysis cohort

Of the 1239 patients enrolled in the BREEZE-AD1 and -AD2 studies, actigraphy data from ≥1 day were available for correlation analyses from 1143 (92%). Of these, 664 (54% of the total clinical trial cohort) had sufficient data for inclusion in the "change from baseline" analyses (see the Methods section for inclusion criteria).

Demographics of study participants

Table 1

shows the baseline demographic and clinical characteristics of the main study population (both trials combined) and the participants with actigraphy data who were included in the "change from baseline" analyses. These two populations were generally similar in terms of demographic and disease severity characteristics, although the proportion of patients of Asian race was numerically higher in participants with actigraphy data.

Relationship between scratch and sleep (digital outcomes)

Digital outcomes for scratch and sleep were derived from 1143 patients. Nocturnal scratch and nighttime sleep were moderately correlated, with Pearson's correlation coefficient (r) values (positive or negative) for correlations between the two scratch outcome measures (hourly scratch duration [absolute value] and scratch intensity) and two of the sleep outcome measures (percent time asleep and wakefulness after sleep onset; WASO) ranging from 0.40 to 0.46 (Figure 1). (Digital outcomes are defined in the Methods section.)

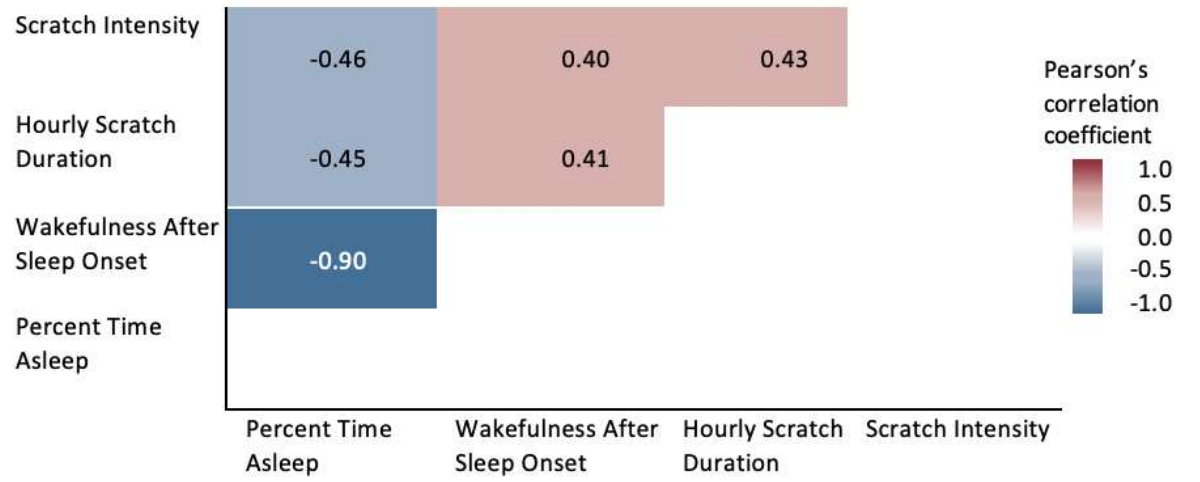


Fig. 1 | Correlations between digital scratch and digital sleep outcomes. Pearson's correlation coefficients for all possible pairs of digital scratch and sleep outcomes (hourly scratch duration, scratch intensity, percent time asleep, and wakefulness after sleep onset) color coded by magnitude and direction.

Dividing the scratch duration data from 1143 patients into quartiles and comparing these with the digital sleep outcomes showed that the percent time asleep was, on average, 23.1% (as an absolute measure) higher in patients with hourly scratch duration in the lowest quartile (Q1) than in those in the highest quartile (Q4) (Table 2). Patients in Q1 also spent an average of 107.4 fewer minutes awake after sleep onset (WASO) than those in Q4.

Using data from 1143 patients, linear regression models of hourly scratch duration (absolute mean values) on percent time asleep and WASO had r^2 values of 0.25 and 0.24, respectively (Figure 2), indicating that approximately 25% of the variation in sleep was explained by scratch duration.

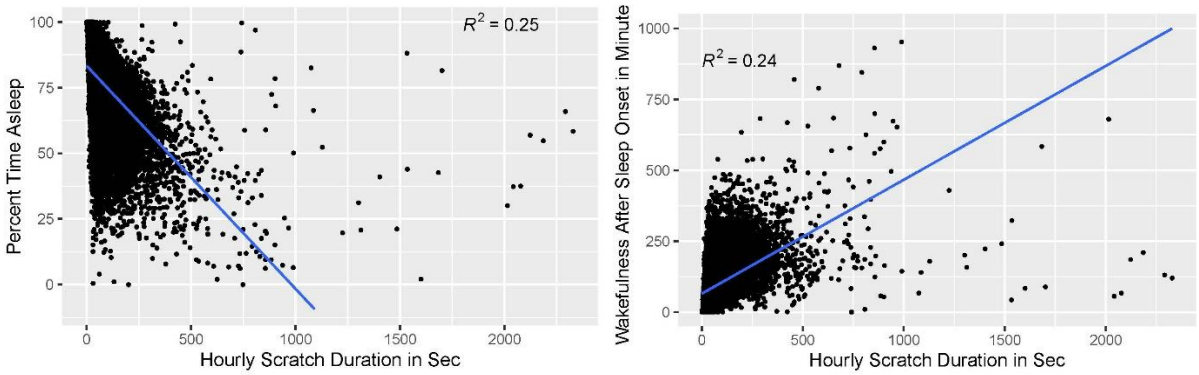


Fig. 2 | Linear regression of digital scratch duration on digital sleep outcomes. Results of linear regression of hourly scratch duration (absolute values) on percent time asleep and wakefulness after sleep onset.

Relationship between digital outcomes and patient- and clinician-reported clinical and quality-of-life outcomes

Correlations between the digital and clinical/quality-of-life outcomes (which include both patient- and clinician-reported outcomes derived from 1143 patients) were weak (Figure 3). Correlations between clinical/quality-of-life outcomes and WASO (0.13–0.24), absolute values of hourly scratch duration (0.17–0.30), and scratch intensity (0.22–0.31) were weakly positive, indicating that higher values of these digital outcome measures corresponded to more severe disease states. In contrast, the percent time asleep findings showed a weak negative correlation with clinical outcomes/quality of life (-0.15 to -0.28), suggesting that an increased percentage of the total sleep opportunity (TSO) spent asleep was associated with less severe disease states.

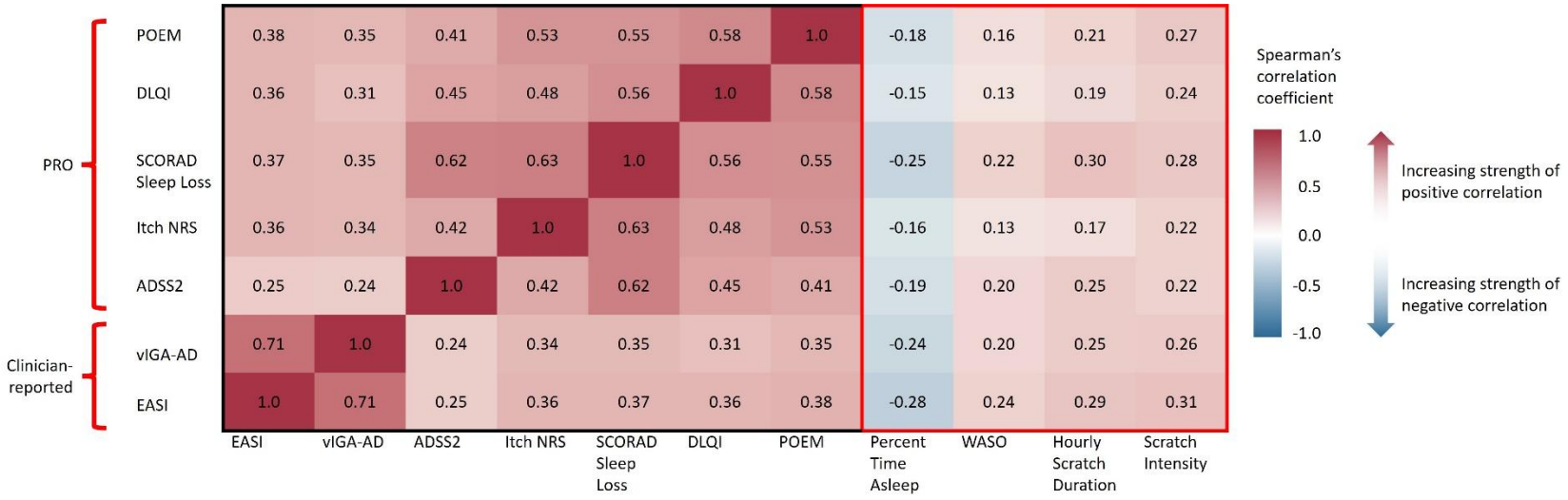


Fig. 3 | Correlations between digital scratch and sleep outcomes and clinical assessments. Spearman's correlation coefficients for all possible pairs of patient- and clinician-reported outcome measures and digital scratch and sleep outcome measures (absolute hourly scratch duration, scratch intensity, percent time asleep, and wakefulness after sleep onset), color coded by magnitude and direction. ADSS2, Atopic Dermatitis Symptom Score item 2; DLQI, Dermatology Life Quality Index; EASI, Eczema Area and Severity Index; NRS, Numerical Rating Scale; POEM, Patient Oriented Eczema Measure; PRO, patient-reported outcome; SCORAD, SCORing Atopic Dermatitis; vIGA-AD, validated Investigator Global Assessment – Atopic Dermatitis; WASO, wakefulness after sleep onset.

Treatment effect estimation with digital endpoints and clinical outcomes

Table 3 shows the values of digital scratch (hourly scratch duration and scratch intensity) and sleep (WASO and percent time asleep) outcomes for the placebo and baricitinib 4 mg treatment groups who participated in actigraphy data collection at baseline and week 4 (N at baseline: placebo, 309; baricitinib 4 mg, 146; N at week 4: placebo, 111; baricitinib 4 mg, 86). At week 4, patients receiving baricitinib 4 mg showed significant improvements versus placebo in all four outcomes: hourly scratch duration, scratch intensity, and WASO were lower by, respectively, a mean ± standard error of 21.6 ± 5.8 seconds, 0.17 ± 0.04 Hz, and 10.8 ± 4.5 minutes, and percent time asleep was higher by 3.3 ± 0.9% ($p<0.05$ for all).

In terms of demonstrating a significant difference between baricitinib 4 mg and placebo in change from baseline at week 4, the mixed-models repeated measures (MMRM) analysis (Figure 4) showed consistency between all the "matched" digital and clinical outcomes for the baricitinib 4 mg treatment arm and for the two paired scratch outcomes in the baricitinib 2 mg treatment arm. However, unlike clinical assessments, significant improvements in digital outcomes were generally not observed in the baricitinib 1 mg treatment arm, and there was no evidence of a placebo effect.

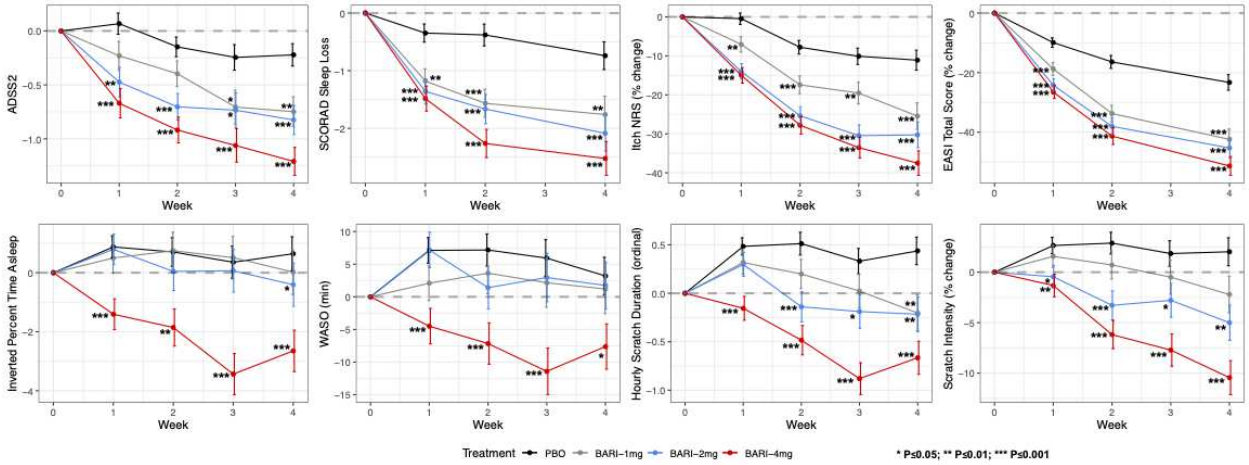


Fig. 4 | Comparison of treatment effect estimation using digital and traditional clinical outcomes. Least square means (± standard error) of the change from baseline in BREEZE-AD1 and BREEZE-AD2 (combined) at weeks 1 through 4, without multiplicity adjustments, for traditional clinical outcomes (upper panel) and digital outcomes (lower panel) for the placebo and active treatment arms. The data shown relate only to study participants who participated in actigraphy data collection. The data were assessed using mixed-models repeated measures analysis incorporating treatment, geographic region, baseline validated Investigator Global Assessment – Atopic Dermatitis, visit, treatment-by-visit interaction, value at baseline, and baseline-by-visit interaction as fixed effects, and subject and subject-by-visit interaction as random effects. P values represent statistical significance vs. placebo. ADSS2, Atopic Dermatitis Symptom Score item 2; BARI, baricitinib; EASI, Eczema Area and Severity Index; NRS, Numerical Rating Scale; PBO, placebo; SCORAD, SCORing Atopic Dermatitis; WASO, wakefulness after sleep onset.

Discussion

Objectively and continuously quantifying both the magnitude (duration and intensity) of nocturnal scratching and the resulting nighttime sleep disturbance in patients with pruritic conditions, including AD, would provide meaningful data to augment the information currently provided by PROs[10,14,18,21]. As a result, the use of wearable devices to measure scratch during sleep has been the subject of considerable research[18]. However, the actigraphy data reported to date are limited in extent and involve relatively small numbers of subjects[18,22,23].

Digital health technologies enable remote, patient-centered data collection and provide real-time insights in clinical settings. As their adoption grows in daily life and clinical care, the need for evidence supporting their use in capturing clinical trial endpoints is also increasing. A framework for reusing existing verification, analytical, and clinical validation data for established digital health technologies was proposed, the aim of which is to maximize the value of prior work[24].

We leveraged our previous work in the development of algorithms for nocturnal scratch and a previously published algorithm for sleep[14] using a wrist-worn digital actigraphy device[12]. For clinical validation, these algorithms were applied to the data collected from two large randomized controlled clinical trials that evaluated the efficacy and safety of an oral selective Janus kinase (JAK) 1/JAK 2 inhibitor in adult patients with AD[25]. These trials offered an opportunity to analyze both pre- and post-treatment data from a large cohort of patients with AD. The objective of the current analysis was to explore the relationship between nocturnal scratching and sleep using digital outcome measures and to examine how these digital metrics correlate with traditional patient- and clinician-reported outcomes.

The accuracy of any digitally measured outcome is substantially dependent on the algorithm used in its derivation. In the setting of AD, of fundamental importance is the algorithm's ability to detect skin contact movements induced by itching. These can include scratching, picking, and rubbing, including rubbing the pruritic site with a leg/foot or, when the patient is in bed, on part of the bed or bedding[12,21,26]. Although some of these movements will not be detected by any wrist-worn actigraphy device[26], actigraphy remains an important and relatively well-studied methodology in this field[18]. The algorithms used in the current study were carefully chosen based on their confirmed analytical and clinical validity. The nocturnal scratching algorithm utilized a combination of predictive models and topological features to enhance the accuracy of scratch detection, particularly in measuring duration and intensity[12]. The algorithm was validated using video recordings as ground truth (gold standard). The sleep algorithm applied in this report used accelerometer and temperature data captured by wrist-worn wearable devices for monitoring sleep[14]. The developers of the algorithm used a multi-step approach—first detecting intended sleep periods, then classifying the sleep–wake state[14]. Digital measures of sleep quantity showed high correlation with corresponding reference measures of sleep quantity obtained from polysomnography[14], which is the gold standard in sleep assessment[27]. This prior validation work provides strong support for the validity of the digital data presented in this study and supports their value as a true reflection of the patient's actual experience without the risk of the multiple forms of bias that are associated with PROs[28].

Baseline data served an important role in allowing us to examine the relationship between digital outcomes and their associations with both patient- and clinician-reported outcomes and in evaluating the effect of treatment on the digital outcomes. Baseline values were calculated as the mean of the values from the 7 days for which data were available within the 14-day window prior to the first study

drug administration. Subjects with fewer than 7 days of data from either wrist within the 14-day window were excluded from the analysis. Digital nocturnal scratch and sleep measures showed moderate correlations, with r values between 0.40 and 0.46 (positive or negative) for hourly scratch duration and scratch intensity versus percent time asleep or WASO. Compared with patients in the highest quartile, those in the lowest quartile for hourly scratch duration had a mean percent time asleep that was 23.1% (as an absolute measure) higher and a mean number of minutes awake after sleep onset that was 107.4 minutes lower. This analysis also demonstrated that approximately 25% of the variation in sleep could be explained by scratch duration. These findings highlight the significant burden that nocturnal scratching places on sleep outcomes in patients with moderate to severe AD.

Digital and clinician-/patient-reported outcomes were correlated, showing weak associations. Higher WASO, hourly scratch duration, and scratch intensity weakly correlated positively with clinician-reported assessments of disease severity (r: 0.20–0.31), patient-reported assessments of itch and sleep (r: 0.13–0.30), and patient-reported assessments of quality of life (r: 0.13–0.27), whereas percent time asleep showed weak negative correlations with these outcomes (clinician-reported assessments of disease severity [r: -0.24 to -0.28], patient-reported assessments of itch and sleep [r: -0.16 to -0.25], and patient-reported assessments of quality of life [r: -0.15 to -0.18]). Comparisons between the digital and traditional outcomes in the current study mirror those from previous investigations, with both Yang et al.[18] and Murray and Rees[10] reporting low correlations between digital actigraphy-based evaluation of scratch and subjective (visual analog scale) measures of itch in patients with pruritic conditions. In the current study, the weak correlations between the digital and subjective outcomes (Figure 3) support some degree of consistency between these endpoints. However, several factors may explain why digital measures of scratching and PROs in AD do not always strongly correlate. First, PROs are subjective and can be influenced by the recall period[29] or mood[6], whereas digital tools capture real-time, objective data, often over different timeframes. In addition, PROs that focus on itch may capture impacts of itch beyond physical scratching, such as frustration or sleep disruption[30]. Itch itself is complex and shaped by various personal and psychological factors, including variation in patients' experience and interpretation of itch, so scratching may not always reflect actual itch severity[18]. Finally, digital sensors may miss subtle scratching or misclassify movements, leading to gaps in data[18].

To further explore the apparent weak correlations between the digital endpoints and PROs, we compared baseline digital sleep and scratch outcomes (upper vs. lower tertiles) with PRO-based severity categories in the 664 patients included in the MMRM analysis (see Supplementary table). Patients with better objective sleep (higher percent time asleep and lower WASO) were more likely to fall into the least severe sleep-loss category. This concordance was most evident for SCORAD sleep loss, the most evenly distributed PRO, and suggested that some patients with objectively poor sleep were not captured by the PRO categories. These findings indicate that the digital endpoints provide complementary information beyond PROs rather than simply replicating them. Digital scratch endpoints and itch PROs showed a similar pattern.

Taken together, data from the current study and those published previously indicate that both subjective (PRO) and objective assessment of itch/scratch have value. However, the limitations of PROs—particularly in pediatric patients or when used to record nighttime itch, where recollection can be particularly problematic[31]—support the value of objective measures of scratch for both improving our understanding of itch and evaluating the efficacy of treatments.

Actigraphy data from the BREEZE-AD1 and -AD2 studies enabled validation of scratch and sleep algorithms in a large trial. After 4 weeks of treatment with baricitinib 4 mg, patients showed mean reductions in hourly scratch duration, WASO, and scratch intensity and a mean increase in percent time asleep—all of which would be expected to align with a positive effect of treatment on quality of life[32–34]. The improvements in these outcomes were all significant when compared with placebo, with mean placebo-adjusted decreases from baseline in the baricitinib-treated group of 21.6 seconds in hourly scratch duration, 10.8 minutes in WASO, and 0.17 Hz in scratch intensity and an increase versus placebo of 3.3% in percent time asleep. The baricitinib 2 mg group had similar digital nocturnal scratch outcomes. Although PROs indicated benefits for itch and sleep in the baricitinib 1 mg group, digital endpoints did not show significant changes. Notably, digital outcomes gave no suggestion of a placebo response, whereas PROs suggested some benefit of placebo. Although this observation should be interpreted with caution, it may highlight the ability of digital devices to isolate true treatment effects.

This study has several limitations. First, wrist-based detection only captures some scratching types, missing non-hand and finger-only scratches[18]. Adding a gyroscope may improve accuracy for certain subjects, but its effectiveness varies[12,22] and research with combined sensors has shown mixed results[22]. Additionally, these studies measured scratch rather than itch—the assumption that scratch is a reliable proxy for itch remains questionable[18] and warrants further investigation. Device number and placement may also affect the data, with prior work reporting better performance if data are captured from both hands[22]. In the current study, the use of actigraphy devices on both wrists was optional. The sample size, although relatively large (n=664) when compared with the smaller cohorts in other studies[12,18], could still limit generalizability. It should also be noted that these 664 patients comprised only 54% of the total clinical trial cohort and that there was a bias towards the involvement of Asian patients. The study cohort thus represents a self-selected subgroup which could potentially differ systematically from those who did not opt in to the actigraphy trial. This was not explored in the analysis. In addition, there was substantial attrition between baseline and week 4, a factor that could potentially be associated with bias within the dataset and that will have influenced the statistical power of the analyses. Further to this point, it should be noted that no multiplicity correction was incorporated into the analyses performed in this study. An additional potential limitation is that only 4 of 16 treatment weeks involved actigraphy, limiting our ability to detect treatment effects: in BREEZE-AD1 and -AD2, the proportion of patients achieving the primary efficacy endpoint (validated Investigator Global Assessment of Atopic Dermatitis [vIGA-AD] score of 0 or 1) generally increased, for all baricitinib doses, between weeks 4 and 16[25], and it is assumed that the digital data would show a corresponding increase in effect over this time period. Finally, the digital data showed some variability, a factor that will limit the likelihood of achieving statistical significance. This variability is related to the finer granularity of digital versus patient-reported outcomes in detecting patients' sleep and scratch patterns.

To our knowledge, this is the largest randomized, placebo-controlled dataset to evaluate actigraphy-measured digital nocturnal scratch and sleep outcomes in patients with moderate to severe AD. The study corroborates previous reports on the low correlation between objective digital outcomes and subjective patient-reported outcomes. This highlights the need for multimodal assessment strategies that integrate objective data with subjective experiences to better understand and manage pruritic conditions such as AD. Along with others, this study demonstrates that actigraphy could become a meaningful tool for characterizing nocturnal scratch in patients with AD, with potential utility for clinical research, regulatory purposes, and potential real-world utilization. However, before this potential can be fulfilled, additional evaluation of the psychometric properties of the tools and validation of digital endpoints is necessary. This could be achieved in future clinical trials by incorporating mandatory wearing of actigraphy devices and inclusion of digital outcomes as part of the pre-specified data analysis

plan. Replication of validation of digital outcomes across different placebo-controlled studies in a range of target patient populations with disease that ranges in severity from mild to severe would also be helpful. If clinical use is to be pursued, assessment of feasibility will also be important. Studies such as the one reported here that compare actigraphy-based and PRO data will continue to shed light on the relative merits of these different methods of assessing the severity of disease, the relationship between itch and scratch, the impact of AD on sleep quality, and the effects of interventions on symptoms that are important determinants of quality of life.

Methods

Detailed descriptions of the design, inclusion and exclusion criteria, and methodology of BREEZE-AD1 (NCT03334396)[19] and BREEZE-AD2 (NCT03334422)[20] were reported previously[25]. Briefly, these two clinical trials were identically designed 16-week, randomized, double-blind, parallel-group, placebo-controlled trials that recruited patients aged ≥18 years with moderate-to-severe AD (vIGA-AD score ≥3; Eczema Area and Severity Index [EASI] score ≥16; ≥10% body surface area involvement). Patients were randomized in a 2:1:1:1 ratio to receive once-daily placebo or 1 mg, 2 mg, or 4 mg of the oral selective JAK 1 and JAK 2 inhibitor, baricitinib. Prior to randomization, topical and systemic treatments for AD were discontinued for at least 2 and 4 weeks, respectively.

Both trials were conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines and were approved by the institutional review board or ethics committee at each participating site. All eligible patients provided written informed consent[25]. Details of clinician-reported outcomes and PROs were described in detail in previous reports[25,35]. The aim of this *post hoc* analysis was to evaluate the performance of actigraphy-derived digital outcome measures that focus on nocturnal scratching and sleep disturbances (i.e., to undertake their clinical validation) and to compare them with PRO measures that evaluate related endpoints.

Digital outcomes: Data collection

Digital data were collected from the patients enrolled in BREEZE-AD1[19] and BREEZE-AD2[20] who consented to actigraphy data collection and who wore GENEActive actigraphy devices (Activinsights, Kimbolton, Cambridgeshire, England). These devices collect ambient temperature (°C) and raw acceleration data in x, y, and z directions at a sampling frequency of 20 Hz. In both trials, actigraphy data collection commenced 2 weeks before randomization and continued up to study week 4 (Figure 5).

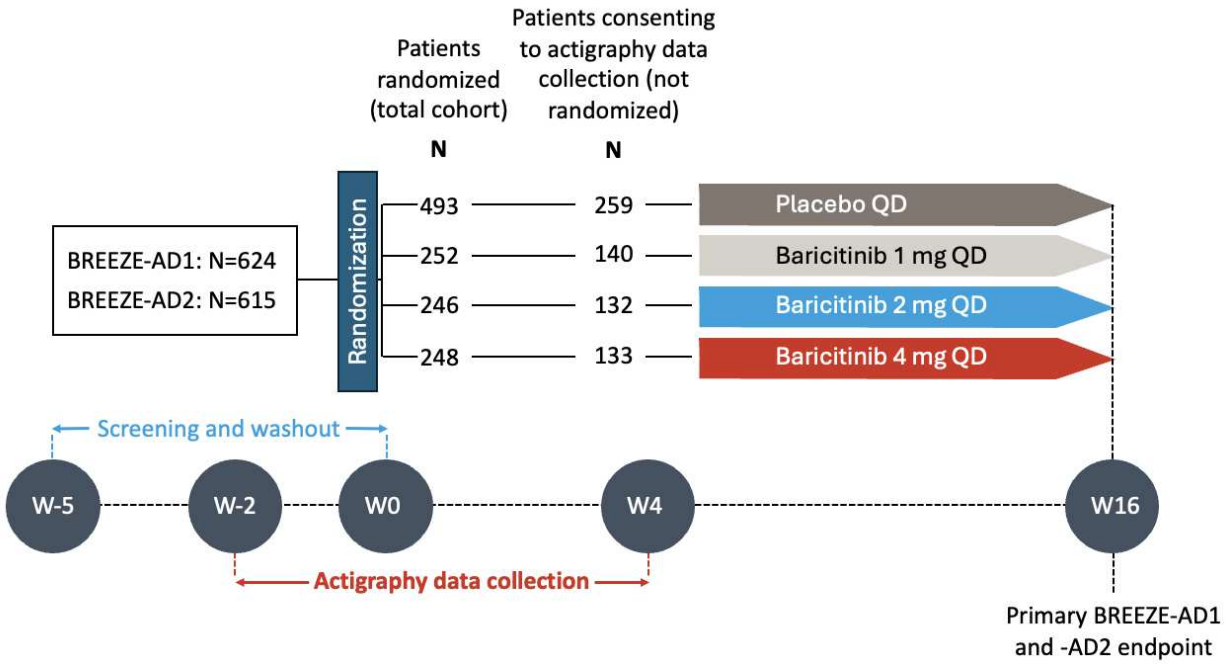


Fig. 5 | **Design of the BREEZE-AD1 and BREEZE-AD2 studies showing the period of actigraphy data collection.** QD, once daily; W, week.

Wearing an actigraphy device was optional during BREEZE-AD1 and BREEZE-AD2. Subjects who chose to participate were provided with two devices and were instructed to wear one on at least one wrist during the night. Only data from subjects who wore at least one actigraphy device were included in the analysis. If the subject wore a device on both wrists, data from both devices were used in the analysis. Overnight periods when the device was not worn were treated as missing and excluded from the analysis without imputation. Treatment effects that required baseline values were only estimated for subjects for whom baseline data were available.

Digital outcomes extraction

The assessed digital endpoints related to scratch duration and intensity and to time spent asleep and awake during the night. The initial step in assessing the digital endpoints was therefore to identify the TSO window. This was defined as the period during which subjects were either sleeping or attempting to sleep. In this manuscript, the term "night" refers specifically to the period of the TSO window.

TSO was derived using open-source packages[36], with "non-wear" periods removed before its derivation. Actigraphy device non-wear periods were identified based on meeting at least one of the following criteria:

1) ≥10 consecutive minutes without any movement[12] and temperature <25°C[14]
2) ≥30 consecutive minutes with temperature <25°C
3) ≥40 consecutive minutes without any movement.

For each subject and each night, the timeseries within the TSO period was segmented into 3-second windows and a pre-trained machine learning model was applied to classify each window as either "scratch" or "no scratch"[12]. These 3-second windows were then aggregated to determine the hourly scratch duration and scratch intensity endpoints. Similarly, a publicly available sleep detection heuristic algorithm was employed to classify each 1-minute window as either "sleep" or "non-sleep"[14]. This classification was based on accelerometer data and a hand motion-based sleep algorithm. If a patient experienced multiple sleep periods within a 24-hour period, the algorithm designated the longest of these periods as their time asleep. Use of this algorithm allowed calculation of the percent time asleep and WASO outcomes (Table 4).

Hourly scratch duration was calculated by summing the scratch durations for each night within the TSO window and dividing by the total number of hours in the TSO window. The resulting hourly scratch duration values could range from 0 to 3600 seconds. To address this large range, the data were transformed to an ordinal scale ranging from 1 to 10, based on percentiles of the original baseline data. This transformation preserved the ability to reflect disease severity while reducing the range of the data. Table 4 shows the digital endpoints extracted and their interpretations. All endpoints were calculated as mean values within each night's TSO, over 7 days within the 14-day baseline period, and over four consecutive 7-day periods within the on-treatment period.

Digital outcomes: Definition of baseline and on-treatment measures

For all digital outcomes, baseline values were computed as the mean of the values from the 7 days for which actigraphy data were available, within the 14-day window before the first study drug administration, that were closest to the date of the first drug administration. These 7 days could be consecutive or non-consecutive. Subjects with fewer than 7 days of data from either wrist within the 14-day window were excluded from the analysis. On-treatment values were reported at weeks 1, 2, 3, and 4 after initiation of treatment (i.e., 7, 14, 21, and 28 days after treatment initiation). Each of these values was computed as the mean of the values obtained in the 7 days prior. Within any 1 week, these values could be derived from 4–7 nights of data collection. If a patient had fewer than 4 nights of data for any digital outcome during any on-treatment week, their data for that outcome were removed from the analysis.

Patient- and clinician-reported outcomes: Definition of baseline and on-treatment measures

The definitions of baseline and on-treatment measures were consistent with the original study analysis plan. The baseline values for each patient's daily diary PRO assessments (Itch Numerical Rating Scale [NRS], skin pain NRS, and Atopic Dermatitis Symptom Score [ADSS]) were—ideally—the means of the assessments in the 7 days before the date of first study drug administration. For patients with fewer than four recorded values for any individual outcome during this 7-day period, the baseline period was extended up to 14 days before study drug administration, and the four most recent assessments were used in calculation of the mean. Patients with fewer than four assessments in the 14-day pre-drug administration period were recorded as having no baseline value for that outcome. In the post-baseline period, the mean of all non-missing assessments made during each of weeks 1, 2, 3, and 4 after initiation of treatment were accepted for analysis, with no lower limit on the number of assessments required for the recording of a value within a week.

Clinician- and patient-assessed outcome measures of disease and symptom severity (Dermatology Life Quality Index, EASI, Patient-Oriented Eczema Measure [POEM], SCORing Atopic Dermatitis [SCORAD], vIGA-AD) were scored at screening, baseline (i.e., on the day of treatment initiation), and at the end of on-treatment weeks 1, 2, and 4[19].

Statistical analysis

All analyses were performed using Python version 3.6+ and R version 4.5.1. Data relating to patient demographics and disease characteristics were analyzed using descriptive statistics. Statistical analysis of the digital endpoint data concentrated on three key areas: the relationship between scratch behavior and sleep as assessed using digital outcome measures; the association between digital outcome measures and traditional clinical assessments; and the clinical significance of digital outcomes. Where correlation analyses were performed, Pearson's correlation was used when all the data were numeric and Spearman's correlation when some of the measures were ordinal. Digital data from each patient-day were only included in the analysis if the TSO from which they were derived was at least 3 hours in duration on that day.

To examine the relationship between scratching behavior and sleep, pairwise Pearson's correlations between digital scratch outcome measures (hourly scratch duration [absolute values] and scratch intensity) and digital sleep outcome measures (percent time asleep and WASO) were calculated. Subsequently, the hourly scratch duration data (absolute values) were divided into quartiles and the mean ± standard deviation percent time asleep and WASO within each quartile was calculated. This approach allowed for a direct comparison of sleep quality across patients with different scratch

durations. To further elucidate the extent to which variations in sleep could be attributed to scratching duration, linear regression analysis was employed, with percent time asleep and WASO as response variables and hourly scratch duration (absolute values) as the covariate.

To understand the relationship between digital and clinical assessments, pairwise Spearman's correlations were calculated between all possible pairs of the following outcomes using data from the participants who underwent actigraphy:

- Digital outcome measures: percent time asleep, WASO, hourly scratch duration (absolute values), scratch intensity
- PROs: Dermatology Life Quality Index, POEM, SCORAD sleep loss, Itch NRS, ADSS item 2 (number of nighttime awakenings due to itch)
- Clinician-reported outcome measures: vIGA-AD, EASI total score.

Baseline data were included in all the analyses described above. However, the absence of baseline data for any individual patient did not preclude use of their on-treatment data.

Finally, to evaluate the treatment effects as estimated by the digital assessments, a methodology akin to that used in the original study by Simpson et al.[25] was adopted. Specifically, a mixed-models repeated measures analysis was employed, using data only from participants who underwent actigraphy. The model incorporated treatment, geographic region, baseline vIGA-AD, visit, treatment-by-visit interaction, value at baseline, and baseline-by-visit interaction as fixed effects and subject and subject-by-visit interaction as random effects[37]. This analysis evaluated change from baseline in the clinical assessments that aligned most closely in evaluation purpose with the digital outcome measures (change from baseline in ADSS item 2 [number of nighttime awakenings due to itch] and SCORAD sleep loss, and percentage change from baseline in Itch NRS, and EASI total score) and the digital outcome measures used in the previous analyses (percentage change from baseline in scratch intensity and change from baseline in percent time asleep, WASO, and hourly scratch duration [ordinal values]). Ordinal (rather than absolute) values of hourly scratch duration were used because these data showed lower variability and hence a better ratio of signal (treatment-associated change in scratch duration) to noise. When shown graphically, the percent time asleep data were inverted. Statistical significance was ascribed if $p<0.05$.No correction was made for multiplicity.

Funding

This study was sponsored by Eli Lilly and Company, under license from Incyte Corporation.

Data and code availability

The datasets generated during and/or analyzed during the current study, as well as the code used to generate the data, are available from the corresponding author upon reasonable request. Data are available to request after publication acceptance. No expiration date of data requests is currently set once data are made available. Access is provided subject to review, after receipt of a signed data sharing agreement. Data and documents will be provided in a secure data sharing environment. For details on submitting a request, see the instructions provided at www.vivli.org.

References

1. Sugiyama, A. et al. Nocturnal scratching and quality of sleep in children with atopic dermatitis. *Acta Derm. Venereol.* **103,** adv12345 (2023). doi: 10.2340/actadv.v103.12345.

2. Bawany, F., Northcott, C. A., Beck, L. A. & Pigeon, W. R. Sleep disturbances and atopic dermatitis: relationships, methods for assessment, and therapies. *J. Allergy Clin. Immunol. Pract*. **9,** 1488-1500 (2021). doi: 10.1016/j.jaip.2020.12.007

3. Vakharia, P. P., Cella, D. & Silverberg, J. I. Patient-reported outcomes and quality of life measures in atopic dermatitis. *Clin. Dermatol*. **36,** 616-630 (2018). doi: 10.1016/j.clindermatol.2018.05.011.

4. Pavlis, J. & Yosipovitch, G. Management of itch in atopic dermatitis. *Am. J. Clin. Dermatol*. **19,** 319-332 (2018). doi: 10.1007/s40257-017-0335-4.

5. Silverberg, J. I. et al. Psychometric properties of the itch numeric rating scale, skin pain numeric rating scale, and atopic dermatitis sleep scale in adult patients with moderate-to-severe atopic dermatitis. *Health Qual. Life Outcomes*. **19,** 247 (2021). doi: 10.1186/s12955-021-01877-8.

6. Kluzek, S., Dean, B. & Wartolowska, K. A. Patient-reported outcome measures (PROMs) as proof of treatment efficacy. *BMJ Evid. Based Med.* **27,** 153-155 (2022). doi: 10.1136/bmjebm-2020-111573.

7. Aiyegbusi, O. L. et al. Key considerations to reduce or address respondent burden in patient-reported outcome (PRO) data collection. *Nat. Commun.* **13**, 6026 (2022). doi: 10.1038/s41467-022-33826-4

8. Long, C., Beres, L. K., Wu, A. W. & Giladi, A. M. Patient-level barriers and facilitators to completion of patient-reported outcomes measures. *Qual. Life Res*. **31,** 1711-1718 (2022). doi: 10.1007/s11136-021-02999-8.

9. Duverdier, A., Hurault, G., Thomas, K. S., Custovic, A. & Tanaka, R. J. Evaluation of measurement errors in the Patient-Oriented Eczema Measure (POEM) outcome. *Clin. Exp. Allergy*. **54,** 207-215 (2024). doi: 10.1111/cea.14441.

10. Murray, C. S. & Rees, J. L. Are subjective accounts of itch to be relied on? The lack of relation between visual analogue itch scores and actigraphic measures of scratch. *Acta Derm. Venereol.* **91,** 18-23 (2011). doi: 10.2340/00015555-1002.

11. Ebata, T., Iwasaki, S., Kamide, R. & Niimura, M. Use of a wrist activity monitor for the measurement of nocturnal scratching in patients with atopic dermatitis. *Br. J. Dermatol*. **144,** 305-309 (2001). doi: 10.1046/j.1365-2133.2001.04019.x.

12. Ji, J. et al. Assessing nocturnal scratch with actigraphy in atopic dermatitis patients. *NPJ Digit. Med.* **6,** 72 (2023). doi: 10.1038/s41746-023-00821-y.

13. Moreau, A. et al. Detection of nocturnal scratching movements in patients with atopic dermatitis using accelerometers and recurrent neural networks. *IEEE J. Biomed. Health Inform*. **22,** 1011-1018 (2018). doi: 10.1109/JBHI.2017.2710798.

14. Mahadevan, N. et al. Development of digital measures for nighttime scratch and sleep using wrist-worn wearable devices. *NPJ Digit. Med*. **4,** 42 (2021). doi: 10.1038/s41746-021-00402-x.

15. Ikoma, A. et al. Measurement of nocturnal scratching in patients with pruritus using a Smartwatch: initial clinical studies with the itch tracker app. *Acta Derm. Venereol*. **99,** 268-273 (2019). doi: 10.2340/00015555-3105.

16. Schulman, J. L. & Reisman, J. M. An objective measure of hyperactivity. *Am. J. Ment. Defic.* **64,** 455–456 (1959).

17. Neishabouri, A. et al. Quantification of acceleration as activity counts in ActiGraph wearable. *Sci. Rep.* **12,** 11958 (2022). https://doi.org/10.1038/s41598-022-16003-x.

18. Yang, A. F. et al. Use of technology for the objective evaluation of scratching behavior: A systematic review. *JAAD Int*. **5,** 19-32 (2021). doi: 10.1016/j.jdin.2021.06.005.

19. Eli Lilly and Company. Protocol I4V-MC-JAHL(a). A multicenter, randomized, double-blind, placebo-controlled, phase 3 study to evaluate the efficacy and safety of baricitinib in adult patients with moderate to severe atopic dermatitis. NCT03334396 https://cdn.clinicaltrials.gov/large-docs/96/NCT03334396/Prot_000.pdf (2017).

20. Eli Lilly and Company. Study of baricitinib (LY3009104) in patients with moderate to severe atopic dermatitis (BREEZE-AD2). NCT03334422 https://clinicaltrials.gov/study/NCT03334422 (2022).

21. Ke Wang, W., Cesnakova, L., Goldsack, J. C. & Dunn, J. Defining the digital measurement of scratching during sleep or nocturnal scratching: review of the literature. *J. Med. Internet Res.* **25,** e43617 (2023). doi: 10.2196/43617.

22. Xing, Y. et al. Quantifying nocturnal scratch in atopic dermatitis: a machine learning approach using digital wrist actigraphy. *Sensors (Basel)*. **24,** 3364 (2024). doi: 10.3390/s24113364.

23. Merola, J.F. et al. Dupilumab significantly improves sleep in adults with atopic dermatitis: results from the 12-week placebo-controlled period of the 24-week phase IV randomized double-blinded placebo-controlled DUPISTAD study. *Br. J. Dermatol*. **189,** 685-694 (2023). doi: 10.1093/bjd/ljad284.

24. Bertha, A. et al. Incorporating digitally derived endpoints within clinical development programs by leveraging prior work. *NPJ Digit. Med*. **6,** 139 (2023). doi: 10.1038/s41746-023-00886-9.

25. Simpson, E.L. et al. Baricitinib in patients with moderate-to-severe atopic dermatitis and inadequate response to topical corticosteroids: results from two randomized monotherapy phase III trials. *Br. J. Dermatol*. **183,** 242-255 (2020). doi: 10.1111/bjd.18898.

26. Smith, M. P. et al. Emerging methods to objectively assess pruritus in atopic dermatitis. *Dermatol. Ther. (Heidelb).* **9,** 407-420 (2019). doi: 10.1007/s13555-019-0312-3.

27. Ancoli-Israel, S. et al. The SBSM guide to actigraphy monitoring: clinical and research applications. *Behav. Sleep Med*. **13 Suppl 1**, S4-S38 (2015). doi: 10.1080/15402002.2015.1046356.

28. Zini, M. L. L. & Banfi, G. A Narrative literature review of bias in collecting patient reported outcomes measures (PROMs). *Int. J. Environ. Res. Public Health*. **18,** 12445 (2021). doi: 10.3390/ijerph182312445.

29. Stull, D. E., Leidy, N. K., Parasuraman, B., & Chassany, O. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. *Curr. Med. Res. Opin*. **25,** 929–942 (2009). Doi: 10.1185/03007990902774765.

30. Rams, A. et al. Sleep-loss related to itch in atopic dermatitis: assessing content validity and psychometric properties of a patient-reported sleep-loss rating scale. *J. Patient Rep. Outcomes*. **8,** 77 (2024). doi: 10.1186/s41687-024-00764-2.

31. Au, C. Y., Manazir, N., Kang, H. & Saleem Bhagat, A. A. Development of objective measurements of scratching as a proxy of atopic dermatitis – a review. *Sensors (Basel).* **25,** 4316 (2025). doi: 10.3390/s25144316.

32. Kong, T. S., Han, T. Y., Lee, J. H. & Son, S. J. Correlation between severity of atopic dermatitis and sleep quality in children and adults. *Ann. Dermatol*. **28,** 321-326 (2016). doi: 10.5021/ad.2016.28.3.321.

33. Snyder, A. M. et al. Atopic dermatitis: a qualitative study on the burdens of living with itchiness. *J. Clin. Aesthet. Dermatol.* **16,** 22-25 (2023).

34. Boehncke, W.-H. et al. Itch improvement has a major and comparable effect on the Dermatology Life Quality Index in psoriasis and atopic dermatitis patients. *EADV Clin. Pract*. **3,** 1509–1517 (2024).

35. Reich, K. et al. Baricitinib improves symptoms in patients with moderate-to-severe atopic dermatitis and inadequate response to topical corticosteroids: patient-reported outcomes from two randomized monotherapy phase III trials. *J. Dermatolog. Treat*. **33,** 1521-1530 (2022). doi: 10.1080/09546634.2020.1839008.

36. van Hees, V. T. et al. Estimating sleep parameters using an accelerometer without sleep diary. *Sci. Rep*. **8,** 12975 (2018). doi: 10.1038/s41598-018-31266-z.

37. Ji, J. et al. 192 Assessing treatment effect of baricitinib in atopic dermatitis using digital endpoints derived from actigraphy data: nocturnal scratch and sleep disturbance. *J. Invest. Dermatol.* **143,** S364 (2023).

38. Leshem, Y. A., Hajar, T., Hanifin, J. M. & Simpson, E. L. What the Eczema Area and Severity Index score tells us about the severity of atopic dermatitis: an interpretability study. *Br. J. Dermatol.* **172,** 1353-1357 (2015). doi: 10.1111/bjd.13662.

39. Oranje, A. P., Glazenburg, E. J., Wolkerstorfer, A. & de Waard-van der Spek, F. B. Practical issues on interpretation of scoring atopic dermatitis: the SCORAD index, objective SCORAD and the three-item severity score. *Br. J. Dermatol*. **157,** 645-648 (2007). doi: 10.1111/j.1365-2133.2007.08112.x.

Acknowledgements

This design of this study, the collection, analysis, and interpretation of the data, and the writing of the manuscript were funded by Eli Lilly and Company. Medical writing support was provided by Janet Douglas and Caroline Spencer (Rx Communications Ltd., Mold, UK).

Author contributions

LC: resources; project administration; writing – review & editing. **EG-Y**: writing – review & editing. **JJ**: conceptualization; investigation; methodology; formal analysis; data curation; software; validation; visualization; writing – original draft preparation; writing – review & editing. **SP**: resources; project administration; conceptualization; investigation; methodology; writing – original draft preparation; writing – review & editing. **JIS**: conceptualization; methodology; writing – review & editing. **NS**: writing – review & editing; supervision. **ASP**: writing – review & editing. **BW**: resources; conceptualization; writing – review & editing; supervision. **HZ**: conceptualization; methodology; data curation; software; validation; visualization; writing – review & editing. All authors read and approved the final manuscript.

Competing interests

Jonathan I. Silverberg has received honoraria as a consultant and/or advisory board member for Abbvie, Aldena, Amgen, AObiome, Apollo, Arcutis, Arena, Attovia, Boehringer-Ingelheim, Bristell-Meyers Squibb, Castle Biosciences, Connect Biopharma, Corevitas, Eli Lilly, Enveda, FIDE, Formation Bio, Galderma, Genzyme, GlaxoSmithKline, Immunocore, Incyte, Infinimmune, Imagene, Kowa Pharmaceuticals, Leo Pharma, Merck, Nanook, Nektar, Novartis, Organon, Pfizer, RAPT, Recludix, Regeneron, Sandoz, Sanofi, Shaperon, TARGET-RWE, Teva, Triton, Triveni, UCB, Union, UpToDate; speaker for Abbvie, Arcutis, Eli Lilly, Galderma, Organon, Pfizer, Regeneron, Sanofi-Genzyme; and institution received grants from Galderma. Amy S. Paller has served as investigator for AbbVie, BioMendics, Dermavant, Eli Lilly, Incyte, Johnson & Johnson Innovative Medicine, Regeneron, UCB; as consultant for Abeona, Arcutis, BioCryst, Boehringer Ingelheim, Castle Creek, Chiesi, Dermavant, Johnson & Johnson Innovative Medicine, Krystal, LEO, Lilly, L'Oreal, MoonLake Immunotherapeutics, Peltheos, Quoin, Regeneron, and Sanofi; and on the data safety monitoring board for AbbVie, Abeona, Biocryst, Daiichi Sankyo, and Galderma. Emma Guttman-Yassky has, in the past 36 months, received grants or contracts from Regeneron, Sanofi, Eli Lilly, Pfizer, Amgen, GSK, Sanofi, Bristol Meyers Squibb, AnaptysBio, AbbVie, Opsidio, Apollo Therapeutics, Aclaris Therapeutics, Concerto Biosciences, Apogee, Enveda Biosciences, Nektar, Triveni Bio, Sitryx, Concert Pharmaceuticals, and Target PharmaSolutions; consulting fees from AbbVie, Aclaris Therapeutics, Almirall, Amgen, AnaptysBio, Apogee Therapeutics, Apollo Therapeutics, Arcutis, Artax Biopharma, Astria, Bambusa, Biolojic Design, Bionyra Pharma, Boehringer-Ingelheim, Bristol Myers Squibb, Canyon Therapeutics, Celldex, Cue Biopharma, CytoReason, DBV, Dualitas Therapeutics, Egle Therapeutics, Electra Therapeutic, Eli Lilly, EMD Serono, Enveda Biosciences, Evommune, Galderma, Generate Biomedicines, Gilead Sciences, GSK, Imagene, Infimmune, Janssen Biotech, Jasper Therapeutics, Kymera Therapeutics, Leo Pharma, Matchpoint Therapeutics, Merck, Navigator Medicines, Nektar Therapeutics, Opsidio LLC, Pfizer, Propeller Bio, Pierre Fabre, Prologue Medicines, Proteologix, Priovant Therapeutics, RAPT, RayThera, Recludix Pharma, Regeneron, Rubedo Life Sciences, Sanofi, SATO, Sitryx Therapeutics, SPARC, Takeda, Teva, Third Arc Bio (TAB), TRex Bio, Triveni Bio, UCB, Vellera Therapeutics, Ventyx, Veradermics, and VRG Therapeutics. Ju Ji, Hui Zhang, Brian Winger, Lisa Cirri, Najwa Somani and Sreekumar Pillai are employees of, and minor shareholders in, Eli Lilly and Company.

Table 1 | Patient demographics and disease characteristics at baseline

Characteristics	BREEZE-AD1 and -AD2 combined				Participants with actigraphy data collected who were included in the "change from baseline" analyses			
	PBO (N=493)[†]	Bari 1 mg (N=252)[†]	Bari 2 mg (N=246)	Bari 4 mg (N=248)	PBO (N=259)	Bari 1 mg (N=140)	Bari 2 mg (N=132)	Bari 4 mg (N=133)
Age, years	35.4 ± 12.8	34.3 ± 11.4	35.2 ± 13.4	35.5 ± 13.6	34.5 ± 11.8	33.5 ± 11.2	36.1 ± 12.9	35.4 ± 13.8
Female sex	191 (38.7)	94 (37.3)	99 (40.2)	83 (33.5)	103 (40.0)	49 (35.3)	55 (41.7)	44 (33.1)
Race								
White	316 (64.4)	159 (63.1)	160 (65.0)	152 (61.5)	154 (59.5)	82 (58.6)	84 (63.6)	83 (62.4)
Asian	145 (29.5)	76 (30.2)	72 (29.3)	79 (32.0)	93 (35.9)	50 (35.7)	47 (35.6)	46 (34.6)
Duration since AD diagnosis, years	25.6 ± 14.8	25.2 ± 14.0	24.6 ± 14.2	23.7 ± 14.9	24.9 ± 13.3	23.4 ± 13.5	25.6 ± 13.9	23.0 ± 13.9
Weight, kg	72.6 ± 15.6	74.4 ± 16.9	73.3 ± 16.3	73.5 ± 16.1	71.9 ± 16.3	73.8 ± 17.3	71.6 ± 15.0	73.5 ± 16.1
BMI, kg/m^2	25.0 ± 4.4	25.3 ± 4.9	25.3 ± 5.1	25.1 ± 4.3	24.9 ± 4.5	24.9 ± 5.0	24.7 ± 4.2	25.1 ± 4.4
Geographical region								
EU	257 (52.1)	131 (52.0)	127 (51.6)	132 (53.2)	128 (49.4)	68 (48.6)	67 (50.8)	68 (51.1)
Japan	90 (18.3)	45 (17.9)	43 (17.5)	45 (18.1)	58 (22.4)	29 (20.7)	31 (23.5)	27 (20.3)
Other	146 (29.6)	76 (30.1)	76 (30.9)	71 (28.7)	73 (28.2)	43 (30.7)	34 (25.8)	38 (28.6)
vIGA-AD								
Score of 3	267 (54.2)	135 (53.8)	132 (53.7)	134 (54.0)	145 (56.0)	84 (60.0)	75 (56.8)	76 (57.1)
Score of 4	226 (45.8)	116 (46.2)	114 (46.3)	114 (46.0)	114 (44.0)	56 (40.0)	57 (43.2)	57 (42.9)
EASI								
All patients	32.3 ± 12.9	31.0 ± 12.4	32.8 ± 14.1	32.5 ± 12.7	32.4 ± 13.2	29.9 ± 12.2	32.7 ± 14.0	31.9 ± 12.5
Moderate (7.1–21.0)	115 (23.3)	68 (27.0)	55 (22.4)	56 (22.6)	56 (21.6)	41 (29.3)	30 (22.7)	33 (24.8)

Severe (21.1–50.0)	319 (64.7)	162 (64.3)	155 (63.0)	164 (66.1)	170 (65.6)	89 (63.6)	84 (63.6)	86 (64.7)
Very severe (50.1–72.0)	59 (12.0)	21 (8.3)	36 (14.6)	28 (11.3)	33 (12.7)	10 (7.1)	18 (13.6)	14 (10.5)
SCORAD								
All patients	67.9 ± 13.3	66.5 ± 13.7	68.6 ± 13.1	68.0 ± 13.2	67.2 ± 13.2	65.2 ± 13.6	67.6 ± 13.6	67.3 ± 12.4
Moderate (25–50)	47 (9.5)	32 (12.7)	18 (7.3)	21 (8.5)	25 (9.7)	20 (14.3)	12 (9.1)	12 (9.0)
Severe (>50)	445 (90.3)	219 (86.9)	228 (92.7)	227 (91.5)	234 (90.3)	120 (85.7)	120 (90.9)	121 (91.0)
Body surface area affected, %	52.4 ± 22.4	50.9 ± 21.8	52.2 ± 24.2	53.0 ± 21.6	52.5 ± 22.6	49.0 ± 20.9	52.0 ± 23.2	52.3 ± 22.3
Itch NRS	6.8 ± 2.1	6.2 ± 2.1	6.5 ± 2.2	6.5 ± 2.1	6.7 ± 2.1	6.0 ± 2.1	6.4 ± 2.1	6.5 ± 2.2
Skin Pain NRS	6.1 ± 2.5	5.6 ± 2.5	6.0 ± 2.6	5.9 ± 2.5	6.0 ± 2.5	5.3 ± 2.5	5.9 ± 2.4	5.9 ± 2.6
ADSS item 2	2.6 ± 4.0	2.0 ± 2.8	2.2 ± 3.5	2.6 ± 4.1	2.5 ± 4.1	1.9 ± 3.1	2.1 ± 3.6	2.3 ± 3.8

Data are provided as mean ± standard deviation or n (%) unless otherwise stated.

[†]Severity data missing at baseline for EASI (one patient in the baricitinib 1 mg-treated group) and SCORAD (one patient each in the placebo- and baricitinib 1 mg-treated groups).

For all the disease characteristic outcomes reported, a higher score indicates greater disease severity. vIGA-AD measures the Investigator's Global Assessment of disease severity based on a static 5-point scale from 0 (clear skin) to 4 (severe disease). EASI assesses the extent (area) and severity of disease; scores range from 0 to 72 (severity categories per Leshem et al.[38]). Itch NRS ranges from 0 (no itch) to 10 (worst itch imaginable). Skin Pain NRS ranges from 0 (no skin pain) to 10 (worst skin pain imaginable). ADSS item 2 assesses the frequency of night-time awakenings due to itch the previous night on a scale of 0–29. SCORAD is a combined score of investigator-reported disease severity, affected body surface area, and patient-reported symptoms of itch and sleep dysfunction; scores range from 0 to 103 (severity categories per Oranje et al.[39]).

AD, atopic dermatitis; ADSS, Atopic Dermatitis Sleep Scale; Bari, baricitinib; BMI, body mass index; EASI, Eczema Area and Severity Index; EU, European Union; NRS, Numerical Rating Scale; PBO, placebo; SCORAD, SCORing Atopic Dermatitis; vIGA-AD, validated Investigator's Global Assessment of Atopic Dermatitis.

Table 2 | Hourly scratch duration (quartiles) vs. digital sleep data

Hourly scratch duration quartile	Mean hourly scratch duration (sec)	Percent time asleep	WASO (min)
Q1 [<25th]	20.1 ± 5.6	88.2 ± 9.5	44.8 ± 38.7
Q2 [25th, 50th]	37.8 ± 5.5	80.6 ± 12.0	76.4 ± 50.7
Q3 [50th, 75th]	63.5 ± 10.2	74.8 ± 12.8	102.9 ± 57.9
Q4 [>75th]	167.2 ± 131.4	65.1 ± 14.5	152.2 ± 82.1

All data are presented as mean ± standard deviation.

min, minutes; Q, quartile; sec, seconds; WASO, wakefulness after sleep onset.

Table 3 | Digital sleep and scratch outcomes for placebo- and baricitinib 4 mg-treated patients at baseline and week 4

Outcome	HSD (sec†)	SI (Hz†)	WASO (min†)	PTA (%†)
Placebo				
Baseline	67.6 ± 45.1	0.95 ± 0.23	97.2 ± 52.6	76.3 ± 12.9
Week 4	71.0 ± 53.7	0.92 ± 0.25	88.3 ± 55.4	77.9 ± 14.3
Week 4: Change from baseline, mean ± SE	12.2 ± 3.8	0.00 ± 0.01	-0.5 ± 3.1	-0.1 ± 0.7
Week 4: Percentage change from baseline, mean	18.0	0.23	-0.5	-0.2
Baricitinib 4 mg				
Baseline	67.1 ± 41.3	1.00 ± 0.33	101.3 ± 55.0	74.9 ± 13.6
Week 4	56.7 ± 39.7	0.86 ± 0.27	88.5 ± 55.0	78.6 ± 12.7
Week 4: Change from baseline, mean ± SE	-9.4 ± 4.4	-0.17 ± 0.03	-14.2 ± 4.7	4.3 ± 0.9
Week 4: Percentage change from baseline, mean	-14.0	-17.24	-14.0	5.7
Between-group difference in change from baseline at week 4				
Baricitinib 4 mg – placebo, mean ± SE§	-21.6 ± 5.8***	-0.17 ± 0.04***	-10.8 ± 4.5*	3.3 ± 0.9***
Difference in mean percentage change, baricitinib – placebo	-32.0	-17.47	-13.5	5.9

Data are presented as mean ± standard deviation unless otherwise indicated.

* $p \leq 0.05$ vs. placebo; *** $p \leq 0.001$ vs. placebo.

† Units relate to absolute values at baseline, week 4, and change from baseline data only.

§ SI, WASO, and PTA data derived from MMRM analysis; HSD mean and SE values based on aggregated mean, p-value derived from MMRM analysis.

HSD, hourly scratch duration; MMRM, mixed-models repeated measures; PTA, percent time asleep; SD, standard deviation; SE, standard error; SI, scratch intensity; WASO, wakefulness after sleep onset.

Table 4. Digital outcomes

Outcome	Interpretation
Hourly scratch duration (seconds)	Mean scratch duration per hour during the TSO
Hourly scratch duration (ordinal)	Ordinal measure (range 1–10); transformed from original value based on percentile within the baseline data
Scratch intensity (Hz)	Mean dominant frequency of hand movement in all scratch events during the TSO
Percent time asleep	Percent of TSO spent asleep
Wakefulness after sleep onset (minutes)	Total number of minutes awake during the TSO after initially falling asleep

TSO, total sleep opportunity.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Baricitinibdigitalendpointsanalysissuppmaterial.docx