Use of technology for the objective evaluation of scratching behavior: A systematic review

Albert F. Yang, BS,[a,b] Morgan Nguyen, BS,[c] Alvin W. Li, MD,[b] Brad Lee, BS,[c] Keum San Chun, PhD,[d] Ellen Wu, BA,[c] Anna B. Fishbein, MD,[e] Amy S. Paller, MD,[b,f,g] and Shuai Xu, MD[b,g,h,i]
Chicago and Evanston, Illinois and Austin, Texas

Introduction: Pruritus is a common symptom across various dermatologic conditions, with a negative impact on quality of life. Devices to quantify itch objectively primarily use scratch as a proxy. This review compares and evaluates the performance of technologies aimed at objectively measuring scratch behavior.

Methods: Articles identified from literature searches performed in October 2020 were reviewed and those that did not report a primary statistical performance measure (eg, sensitivity, specificity) were excluded. The articles were independently reviewed by 2 authors.

Results: The literature search resulted in 6231 articles, of which 24 met eligibility criteria. Studies were categorized by technology, with actigraphy being the most studied (n = 21). Wrist actigraphy's performance is poorer in pruritic patients and inherently limited in finger-dominant scratch detection. It has moderate correlations with objective measures (Eczema and Area Severity Index/Investigator's Global Assessment: $r_s(\rho)$ = 0.70-0.76), but correlations with subjective measures are poor (r^2 = 0.06, $r_s(\rho)$ = 0.18-0.40 for itch measured using a visual analog scale). This may be due to varied subjective perception of itch or actigraphy's underestimation of scratch.

Conclusion: Actigraphy's large variability in performance and limited understanding of its specificity for scratch merits larger studies looking at validation of data analysis algorithms and device performance, particularly within target patient populations. (JAAD Int 2021;5:19-32.)

Key words: algorithm; atopic dermatitis; disease management; drug development; eczema; general dermatology; itch; machine learning; pediatric dermatology; pruritus; technology.

INTRODUCTION

Pruritus is a common symptom of systemic and dermatologic disorders, and scratching is the innate reflex.[1] The itch-scratch cycle is a hallmark symptom of atopic dermatitis (AD) and perpetuates skin barrier dysfunction. Notably more severe during sleep, itch in AD has been shown to impact sleep quality.[2-5] Historically, itch has been assessed subjectively through visual analog scales (VAS) and numeric rating scales.[6] However, these measures often do not correlate to visually observed scratch, especially in children.[7-9] More recently, studies have explored device-driven methods to objectively measure scratch as a proxy for itch.

Actigraphy is the most commonly tested method and entails the use of accelerometers to monitor wrist movements, a proxy for scratching. Other technologies include acoustic devices,[10,11] strain gauges,[12,13] pressure sensors,[12,14] and vibratory sensors.[12,13,15-18] The commonly accepted gold

From the University of Illinois at Chicago College of Medicine[a]; Department of Dermatology, Northwestern University Feinberg School of Medicine, Chicago[b]; Northwestern University Feinberg School of Medicine, Chicago[c]; Electrical and Computer Engineering, The University of Texas at Austin[d]; Department of Pediatrics (Allergy and Immunology)[e] and Department of Pediatrics (Dermatology), Ann & Robert H. Lurie Children's Hospital, Chicago[f]; Center for Bio-Integrated Electronics, Evanston[g]; Querrey Simpson Institute for Bioelectronics, Evanston[h]; and the Department of Biomedical Engineering, Northwestern University, Evanston.[i]

Funding sources: None.

IRB approval status: Not applicable.

Accepted for publication June 14, 2021.

Correspondence to: Shuai Xu, MD, Simpson Querrey Institute, 303 E. Superior Street, 11-334, Chicago, IL 60611. E-mail: stevexu@northwestern.edu; stevexumd@gmail.com.

2666-3287

https://doi.org/10.1016/j.jdin.2021.06.005

standard is video recording of scratching with manual coding by an observer, which is time-consuming and impractical in clinical settings.[19-21] The purpose of this systematic review is to assess the performance and algorithms of technological methods currently available to evaluate scratching behavior objectively.

CAPSULE SUMMARY

- We assessed ways to quantify itch by measuring scratching behavior via various technological modalities (eg, actigraphy, smartwatch applications, acoustic sensors).
- The overall performance of current objective tools for quantifying itch suffers from low accuracy and variable performance. Further development will allow for more-objective evaluation of disease management and treatment.

METHODS

Search strategy

We queried PubMed, MEDLINE, Embase (Elsevier), Cochrane Library and Cochrane Central Register of Controlled Trials (CENTRAL), Scopus (Elsevier), Web of Science (Clarivate Analytics), and IEEE Xplore Digital Library in October 2020 without limits on publication date. The search strategy is fully detailed in the Supplemental Materials under "Search Strategy" (available via Mendeley at https://data.mendeley.com/datasets/ryg97c26t6/2).

Study selection

Eligibility assessment was performed independently by 2 authors. Included articles must feature critical assessment of a technology designed to measure itch objectively and report at least 1 of the primary outcomes described below. Exclusion criteria included studies of nonhuman subjects, articles without original data, and studies describing technology without assessing its performance.

Quality assessment

Study quality was assessed using a rating scheme (1-5), which was modified from the Oxford Centre for Evidence-Based Medicine[22] for rating levels of evidence. The individual studies assessed are described in Tables I and II and assessment was performed by at least 2 authors.

Data extraction and outcomes

Performance values were extracted using a standardized survey. Primary outcomes included sensitivity, specificity, and positive and negative predictive values of scratch detection methods. Secondary outcomes included correlations of detection methods to other technologies and subjective assessments.

Performance metrics

Sensitivity is defined as the ability to detect the number of true positives (eg, true scratching) and specificity is the ability to detect the number of true negatives (eg, nonscratching movements). Positive predictive value (PPV, precision) is the proportion of positives that are true positives (eg, movements labeled as scratch that are true scratches). The F1 score encompasses both sensitivity and precision. Root mean square error (RMSE) is the standard deviation of residuals and is effectively an estimation of how well an algorithm predicts the observed data (ie, accuracy).

Algorithms

To efficiently extract and analyze device data, algorithms capable of distinguishing scratch from nonscratch movements are essential. Linear regression modeling is generated from the number of activity counts above a frequency threshold and total scratch time; however, this model is limited by confounding movements (eg, walking, restlessness).[23] Logistic regression modeling is a simple approach to binary classification (eg, scratch vs nonscratch) and analogous to linear regression. Bidirectional recurrent neural networks are a form of machine learning whereby the network can detect patterns directly (eg, scratch waveforms) from raw input data, thereby eliminating precursory extraction of patterns required for other models.[24] The k-means clustering analysis is another approach that involves clustering a set number of subgroups within a data set. The algorithm then allocates device signals into their respective subgroups based on frequency, waveform, or other qualities.[23]

RESULTS

Of the 6231 articles identified, 72 were assessed based on exclusion criteria and 24 fully met eligibility criteria. Most articles looked at AD, although other conditions were also examined (eg, urticaria). Articles reporting performance and correlation measures are summarized in Tables I and II.

Abbreviations used:

AD: Atopic dermatitis
PPV: Positive predictive value
RMSE: Root mean square error
VAS: Visual analog scale
TST%: total scratching time percentage

Sensitivity and specificity ranges of technologies compared to video recording are summarized in Table III. An overview of benefits and limitations is seen in Table IV.

Actigraphy

Performance. Actigraphy is the most studied technology.[5,21,25-27] Twenty-one articles investigated actigraphy devices and data extraction algorithms, with 7 compared to video recording.[12,16,21,24,25,28,29] While all 7 articles looked at healthy subjects, only 2 reported sensitivity values (0.00-0.96; zero values indicate no true positives).[24,29] Specificity was reported by 1 article (0.92).[29] Four articles explored actigraphy in AD subjects, with 1 reporting sensitivity values (0.00-0.89) and PPV values (0.00-0.57).[24] Specificity was not reported in this population. The large ranges likely stem from the various extraction algorithms and actigraphs (eg, PAM-RL, [26,29] Actiwatch Plus,[8,26] DigiTrac[30,31]).

Each algorithm has its limitations. The k-means clustering analysis algorithm of Feuerstein et al[23] yielded high performance values, but required all anticipated movements to be determined a priori. While logistic regression approach from Petersen et al[29] for detecting total nocturnal scratch time yielded comparable performance to the algorithm from Feuerstein et al,[23] the model had significantly decreased performance when tested with a separate data set.[24] The bidirectional recurrent neural networks algorithm proposed by Moreau et al[24] yielded higher sensitivity, PPV, and F1-scores than the logistic regression model; however, it has not been tested in further datasets. Correlation between actigraphy data and video recording was evaluated by Moreau et al,[24] reporting Spearman rank correlation coefficients ($r_s(\rho)$) of 0.95-0.96.[28] Other studies report correlations between actigraphy and video recording for total scratching time percentage (TST%) calculation ($r_s(\rho) = 0.91$),[25] and correlation values between actigraphy and video recording of sleep efficiency were all reported to be greater than 0.92 by Benjamin et al.[21]

Correlations with other objective and subjective measures. Several articles explored correlations between actigraphy and subjective sleep measures, disease severity, AD-associated serum markers, and subjective itch measures. Ten articles compared actigraphy to subjective sleep measures, with 4 reporting correlations. VAS sleep, a patient-reported measure of sleep quality, was examined in 1 article, reporting correlation coefficients of −0.44 in adults and 0.48 in children when compared to average hourly activity scores.[26] The total scoring AD index, which includes both subjective (eg, itch and sleep) and objective (eg, disease severity) measures, had moderate correlations with various activity measures ranging from 0.53-0.64 in adults ($P < .05$) and 0.42-0.62 in children ($P < .05$).[26,27,30] While total and objective total scoring AD indexes both resulted in $r_s(\rho) = 0.52$ ($P < .001$) in children (n = 24) compared to wrist activity, correlations with pruritus and sleep subscores were not significant.[30]

Two articles evaluated other disease severity indices in children and adults, with moderate correlation for objective measures (Eczema and Area Severity Index and Investigator's Global Assessment) compared to actigraphic wake after sleep onset, ranging from 0.70-0.76 ($P < .02$, n = 10).[32] Six area six sign AD was found to have a weak correlation with average nocturnal movement ($r_s(\rho) = 0.15$, $P = .02$, n = 235).[33]

Four articles investigated -serum markers associated with AD. Statistically significant correlations with actigraphy measurements ranged from 0.51-0.93.[17,27,30,31] These studies were not compared to video recording, however, and thus conclusions specifically related to scratch are difficult to make.

While there seems to be a moderate correlation between actigraphy and objective measures, this is not the case with subjective measures. Fourteen articles compared actigraphy to subjective itch, with 2 articles reporting correlation coefficients. Comparison between VAS itch and mean actigraphy scores yielded coefficients of determination (r^2) of 0.06 in children and adults with various pruritic conditions (n = 118) and 0.08 in adult AD subjects (n = 20).[8] VAS itch and hourly activity scores yielded $r_s(\rho) = 0.40$ ($P = .049$) in children and 0.18 ($P = .9$) in adults.[26]

Actigraphy-based scratch measurements correlate poorly to VAS itch scores, sleep quality, and other subjective patient-reported outcomes.[8,21,26,27,30] The reasons for this are likely multifactorial. In pediatric populations, proxy measures may be under or overestimated by caregivers. More likely, there are inherent differences between a subject's perception of itch and the objective actions of scratching. An individual may report a high level of subjective itch

Table I. Summary table for studies exploring wrist actigraphs and smartwatch applications

Device types	Study	Sample size and population	Study focus	Video recording? (Yes/No)	Sensitivity	Specificity	Correlation	Accuracy	Study quality (1-5)*
Actigraphy	Feuerstein[23]	Healthy adults (n = 12)	Testing k-means cluster algorithm	No	0.90 ± 0.10	0.98 ± 0.05 (walking) 0.88 ± 0.06 (restlessness)		0.92 (scratch) 0.92 (walking) 0.97 (restless sleep)	3
	Petersen[29]	Healthy adults (n = 12)	Testing logistic regression algorithm	Yes	0.96 (all data) 0.96 (cross-validation, mean)	0.92 (all data) 0.92 (cross-validation, mean)			3
	Almazan[28]	Healthy adults (n = 3), AD adults (n = 9)	Testing BRNN algorithm	Yes			‖r_s (ρ) = 0.96 (actigraphy and video scoring) ‖r_s (ρ) = 0.90 (number of scratching events at home and polysomnography)		3
	Moreau[24]	Healthy adults (n = 6), AD adults (n = 18)	Testing BRNN algorithm compared to logistic regression	Yes	AD: 0.45-0.91 (BRNN) 0.00-0.10 (logistic regression) Healthy: 0.00-0.75 (BRNN) 0.00-0.50 (logistic regression) Total: 0.66 (BRNN) 0.06 (logistic regression)		r^2 = 0.98 r_s (ρ) = 0.95 (BRNN and video recording)	F1 scores: AD: 0.27-0.90 (BRNN) 0.00-0.14 (logistic regression) Healthy: 0.00-0.29 (BRNN) 0.00-0.08 (logistic regression) Total: 0.68 (BRNN) 0.09 (logistic regression)	
	Kurihara[12]	Healthy adults (n = 10)	Actigraphy vs video recording and other devices for TST% calculation	Yes			RMSE = 5.32%-8.12%		2

Murray[8]	Study 1: healthy subjects (n = 24; 12 adults, 12 children), pruritic subjects (n = 118; 68 adults, 50 children) Study 2: AD adults (n = 20)	Actigraphy vs VAS itch	No		Study 1: r^2 = 0.06 Study 2: r^2 = 0.08	3
Shino[37]	Healthy adults (n = 1)	Actigraphy vs video recording and other devices for TST% extraction via novel algorithm	Yes	RMSE = 0.83s (0.64s) TST% error = +5.02% (+4.33%) (parentheses are from visually scoring outputs)		3
Wootton[33]	AD children (n = 336)	Actigraphy vs AD severity (SASSAD, POEM)	No		r_s (ρ): SASSAD = 0.15 (P = .02) POEM = .10 (P = .13)	3
Hon[30]	AD children (n = 24 for subjective surveys, n = 20 chemokines)	Actigraphy vs SCORAD scores and AD-associated chemokines	No		r_s (ρ): †Total SCORAD = 0.52 ‡Objective SCORAD = 0.52 SCORAD pruritus = 0.23 SCORAD sleep loss = 0.36 †CTACK = 0.56 §MDC = 0.63 †TARC = 0.54	3
Hon[31]	AD children (n = 28)	Actigraphy vs BDNF and substance P	No		r_s (ρ): ‖BDNF = 0.83-0.91 ‖Substance P = .83-.87	3

Continued

Table I. Cont'd

Device types	Study	Sample size and population	Study focus	Video recording? (Yes/No)	Sensitivity	Specificity	Correlation	Accuracy	Study quality (1-5)*
	Fujita[27]	AD adults (n = 15)	Actigraphy vs SCORAD, VAS itch, serum cytokines	No			r_s (ρ): †VAS daytime itch = 0.58 †SCORAD = 0.54 †TARC = 0.51 †LDH = 0.65		3
	Bender[43]	Healthy adults (n = 14), AD adults (n = 14)	Actigraphic sleep measures vs VAS itch	No			r_s (ρ): †WASO = 0.35 †Sleep efficiency = 0.38 †Average sleep = 0.46		3
	Benjamin[21]	Healthy children (n = 7), AD children (n = 14)	Video recording (sleep time, scratch time, restlessness) vs actigraphy and VAS itch	Yes			r_s (ρ): ‡Actigraphy, all > 0.92 VAS itch = 0.16-0.30 ($P > .05$)		3
	Bringhurst[26]	Pruritic subjects (n = 33 adults, n = 25 children), healthy subjects (n = 30 adults, n = 17 children)	Actigraphy vs subjective scores (VAS sleep, VAS itch, VAS skin disease), and SCORAD	No			r_s (ρ): Children: †VAS sleep = 0.48 †VAS itch = 0.40 †VAS skin disease = 0.49 ‡SCORAD = 0.62 Adults: †VAS sleep = −0.44 VAS itch = 0.18 †VAS skin disease = 0.15 †SCORAD = 0.53		3
	Ebata[25]	Healthy adults (n = 5), AD adults (n = 29)	Actigraphy vs video recording in TST% calculation	Yes			‖r_s (ρ) = 0.91		3

	Sandoval[32]	AD adults (n = 10)	Actigraphic WASO vs IGA and EASI at baseline and after 5-day fluocinonide 0.1% cream	No			r_s (ρ): †baseline EASI = 0.75 †baseline IGA = 0.76 †end treatment EASI = 0.70 †end treatment IGA = 0.73		3
	Kaburagi[16]	Healthy adults (n = 12)	TST% estimation algorithm for various devices	Yes			RMSE = 4.29% (4.85%) (parentheses are from visual scoring of outputs for TST%)		4
Smartwatch applications	Ikoma[36]	AD adults (n = 5)	"ItchTracker" (now "DermaTrack") testing for scratch detection	Yes	0.85 ± 0.10		R = 0.85-0.90		4
	Lee[34]	Healthy adults (n = 3)	"Itchtector" prototype testing	Yes	dominant hand = 0.98-1.00 nondominant hand = 0.63-0.82	dominant hand = 0.98-1.00 nondominant hand = 0.99		dominant hand = 0.985-0.99 nondominant hand = 0.933-0.976	3
	Lee[35]	Pruritic subjects (n = 13)	"Itchtector" testing in pruritic subjects	Yes	0.75			0.90	3

AD, Atopic dermatitis; *BDNF*, brain-derived neurotrophic factor; *BRNN*, bidirectional recurrent neural network; *CTACK*, cutaneous T-cell-attracting chemokine; *EASI*, Eczema Area and Severity Index; *IGA*, Investigator's Global Assessment; r_s (ρ), Spearman's rank correlation coefficient; *LDH*, lactate dehydrogenase; *MDC*, macrophage-derived chemokine; r^2, coefficient of determination; *RMSE*, root mean square error; *POEM*, Patient-Oriented Eczema Measure; *SASSAD*, Six Area, Six Sign Atopic Dermatitis; *SCORAD*, SCORing Atopic Dermatitis; *TARC*, thymus and activation-regulated chemokine; *TST%*, total sleep time percentage; *VAS*, visual analog scale; *WASO*, wake after sleep onset.

*Study quality was assessed using a rating scheme modified from the Oxford Centre for Evidence-Based Medicine for ratings of individual studies: (1) properly powered and conducted randomized clinical trial or systematic review with meta-analysis; (2) well-designed controlled trial without randomization or prospective comparative cohort trial; (3) case-control study or retrospective cohort study; (4) case series with or without intervention or cross-sectional study; and (5) opinion of respected authorities or case reports.[22]

†$P < .05$.

‡$P < .01$.

§$P < .005$.

‖$P < .001$.

Table II. Summary table for studies exploring acoustic, vibratory, pressure, and strain gauge devices. Note that no specificity values are reported for any of the studies listed

Device type	Study	Sample size and population	Study focus	Video recording? (Yes/No)	Sensitivity	Correlation	Accuracy	Study quality (1-5)*
Acoustic	Kurihara[12]	Healthy adults (n = 10)	Finger-mounted microphone vs video recording and other devices for TST% calculation	Yes			RMSE = 1.09%	2
	Noro[10]	Healthy adults (n = 8), AD adults (n = 4)	Wristwatch-type piezoelectric device for scratching rate compared to video recording	Yes		r^2 = 0.98 (nocturnal scratching rate by acoustic device vs video recording)		3
Vibratory	Kurihara[18]	Healthy adults (n = 12)	Validation of piezoceramic disk devices placed under bed legs vs video recording for scratch and nonscratch	Yes			RMSE (staying calmly) = 0.35-0.72s RMSE (moving hand, turning over, moving foot) = 0.94-1.26s RMSE (scratching) = 0.56-1.29s	3
	Kurihara[12]	Healthy adults (n = 10)	Piezoceramic disk bed devices placed under bed legs vs video recording and other devices for TST% calculation	Yes			RMSE = 0.87 = 6.31%	3
	Shino[37]	Healthy adults (n = 1)	Piezoceramic bed devices vs video recording and other devices for TST% extraction via novel algorithm	Yes			RMSE = 0.68-0.79s (0.40-0.94s) TST% error = 2.13-4.11% (−6.51-0.82%) (parentheses are from visually scoring outputs)	3
	Kaburagi[16]	Healthy adults (n = 12)	TST% estimation algorithm for various devices	Yes			RMSE (left bed head) = 1.51% (1.84%) RMSE (right bed head) = 0.92% (1.86%) RMSE (left bed foot) = 6.58% (6.27%)	4

							RMSE (right foot bed) = 3.97% (6.83%) (parentheses are from visual scoring of outputs for TST%)	
	Kogure[17]	AD subjects (n = 20)	Evaluation of sheet-shaped body vibrometer vs wrist actigraphy for measurement of scratching, activity count, and sleep efficiency	No		r_s (ρ): activity count per minute = 0.63-0.82 † sleep efficiency = 0.82-0.91‡		3
Pressure Sensor	Endo[14]	Healthy adults (n = 10), AD adults (n = 20 total; 10 male, 10 female)	Evaluation of "Scratch Monitor" device on dorsal hand	No	0.74 (overall) 0.65 (male) 0.83 (female)			3
	Kurihara[12]	Healthy adults (n = 10)	Ceramic sheet placed on dorsal hand vs video recording and other devices for TST% calculation	Yes			RMSE = 0.73%	3
Strain Gauge	Kurihara[12]	Healthy adults (n = 10)	Strain gauge on index finger vs video recording and other devices for TST% calculation	Yes			RMSE = 2.41%	3
	Shino[37]	Healthy adults (n = 1)	Strain gauge on index finger vs video recording and other devices for TST% extraction via novel algorithm	Yes			RMSE = 0.53s (0.37s) TST% error = +1.38% (−1.54%) (parentheses are from visually scoring outputs)	3
	Kaburagi[16]	Healthy adults (n = 12)	TST% estimation algorithm for various devices	Yes			RMSE = 1.29% (1.63%) (parentheses are from visual scoring of outputs for TST%)	4

AD, Atopic dermatitis; r^2, coefficient of determination; *RMSE*, root mean square error; r_s (ρ), Spearman's rank correlation coefficient; *TST%*, total sleep time percentage.
*Study quality was assessed using a rating scheme modified from the Oxford Centre for Evidence-Based Medicine for ratings of individual studies: (1) properly powered and conducted randomized clinical trial or systematic review with meta-analysis; (2) well-designed controlled trial without randomization or prospective comparative cohort trial; (3) case-control study or retrospective cohort study; (4) case series with or without intervention or cross-sectional study; and (5) opinion of respected authorities or case reports.[22]
†$P < .005$.
‡$P < .001$.

Table III. Reported sensitivity of algorithms for scratch detection in studies focused on subjects with atopic dermatitis, which used video recording as comparison

Performance metric	Actigraphy	Smartwatch applications
Sensitivity (range)	0.45-0.91 (BRNN)[24] 0.00-0.10 (logistic regression)[24]	0.75-0.85[34,36]

BRNN, Bidirectional recurrent neural network.

but exhibit an equally high level of scratching restraint. In contrast, some individuals with chronic itch are habituated to it and report low scores despite frequent scratching. Ultimately, scratch measurements with objective tools and subject-reported outcomes are interrelated outputs that provide complementary information.

Smartwatch applications

Applications leveraging smartwatches and their accelerometers show comparable performance in detecting scratch when compared to actigraphs. Three articles examined smartwatch applications compared to video recording. In preliminary testing of their "Itchtector" app, Lee et al[34] reported sensitivity (0.63-1.00), specificity (0.98-1.00), PPV (0.83-0.98), negative predictive value (0.93-1.00), and accuracy (93.3%-99.0%) in healthy adults (n = 3). When cross-validated in pruritic subjects (n = 13), the app yielded lower sensitivity (0.75), PPV (0.74), and accuracy (90%), which may be due to the small initial sample size, different subject populations, and different smartwatches.[35]

Ikoma et al[36] also tested the "ItchTracker" app in adult AD subjects (n = 5) and reported a sensitivity of 0.85 and PPV of 0.90. They reported a correlation between the app and video recording for an hourly scratch duration of $r_s(\rho)$ = 0.851-0.901 ($P < .001$). The authors further compared scratching duration percentage to current and 7-day itch in healthy and AD adults, reporting $r_s(\rho)$ = 0.36-0.43 ($P < .001$). Similar findings were reported regarding self-reported sleep disturbance ($r_s(\rho)$ = 0.45) and daytime disturbance ($r_s(\rho)$ = 0.42). Disease severity measured by the Eczema and Area Severity Index was significantly correlated to scratching duration percentage ($r_s(\rho)$ = 0.60).[36] However, they excluded finger-only scratching movements. Additionally, the small sample size should be taken into consideration. Although smartwatch applications show good sensitivity, there are no reported specificity ranges for pruritic subjects, making it difficult to assess their ability to distinguish between scratch and nonscratch movements.

Acoustic

Acoustic devices detect sound waves generated from scratching. Two articles studied healthy subjects and compared the performance of their respective devices to that of video recording. No sensitivity or specificity values were reported. The finger-mounted microphone presented by Kurihara et al[12] yielded an RMSE of 1.09% for TST% calculation when compared to video recording. Noro et al[10] reported r^2 = 0.98 when comparing scratching rate captured by their acoustic sensor and scratching rate obtained from video observation. While the devices show strong accuracy in detecting fine finger movements, the technology is not widely available and follow-up studies have not been conducted since first reported in 2014.

Vibratory

Vibratory devices allow for nonintrusive monitoring of body movements and mitigate lesion exacerbation by devices that require skin contact. Four articles studied bed vibratory sensors compared to video recording.[12,13,18,37] Accuracy was measured by RMSE, ranging 0.56-1.29s for scratching time[18] and 0.87%-6.31% for TST% calculation.[12] Shino et al[37] reported comparable RMSE values for their TST% algorithm (0.68-0.79s) when compared to visually scored device outputs (0.40-0.94s) (n = 1). For both studies, the vibratory RMSE values were among the lowest when compared to other technologies. While vibratory devices have comparable accuracy to actigraphy and are largely burden-free once installed, their cost and setup may be deterrents.

Pressure sensors

Pressure sensors placed on the dorsal hand detect pressure changes with hand movements. Only 1 of 2 articles was compared to video recording. Kurihara et al[12] compared a ceramic sheet to other devices in healthy subjects, and reported a RMSE of 0.72% for TST% calculation when compared to video recording, the lowest among the devices tested. Although not compared to video recording, the Scratch Monitor pressure sensor presented by Endo et al[14] was tested in healthy adults and yielded sensitivity ranging from 0.65-0.83.

Strain gauge

Strain gauges placed on the index finger to measure finger bending were evaluated in 2 studies, both of which were compared to video recording

Table IV. Comparison of various technologies used to detect scratching

Device type	Benefits/pros	Limitations/cons	Algorithms for scratch detection
Actigraphy	• Most studied, has a large literature base • Validated against video recording • High sensitivity for wrist-dominant scratching movements in healthy subjects • Statistically significant moderate correlation with other objective measures (eg, SCORAD, IGA, EASI)	• Very poor correlation with subjective assessment tools for itch • Varied performance regarding scratch detection • Poor sensitivity for finger-dominant scratching movements • Deterioration of performance in pruritic subjects • Poor specificity given difficulty distinguishing wrist movements from scratching • False positives with similar waveforms (eg, walking) • Larger studies in target populations (eg, AD subjects) needed for algorithm development	• The k-means cluster analysis algorithm has good performance, but impractical in clinical setting given required determination of all movements a priori[23] • The BRNN model has good performance in pruritic subjects (albeit poorer than healthy subjects) and moderate F1-scores[24] • Logistic regression model in the study by Petersen et al[29] has comparable performance to k-means cluster analysis, but poorer performance in separate data set by Moreau et al[24] • Note that all of the aforementioned algorithms are for determination of TST
Smartwatch applications	• Similar to actigraphy in that it utilizes the smartwatch's built-in accelerometer, more convenient for current smartwatch owners • Bluetooth and cloud capabilities make accessing data easy for both patients and health care providers	• Few applications available • Some applications (eg, "DermaTrack", formerly called "ItchTracker") do not show raw data output • Smartwatches may be cumbersome for pediatric subjects, with no currently reported pediatric data	• Algorithm proposed by Lee et al[34,35] reveals good accuracy in pruritic subjects, but authors report false negatives (eg, nonperiodic scratching) and false positives (nonscratching periodic movements such as arm shaking)
Acoustic	• Greater specificity with detection of scratch-generated sounds; will have different pattern than restlessness or turning over • Able to detect both finger and wrist scratching	• Limited research • Privacy concerns/risk • Unable to use in patients who do not sleep alone or have OSA	• Able to estimate TST% in healthy subjects with high accuracy (low RMSE compared to video recording) in healthy subjects
Vibratory	• Noninvasive • Able to localize scratching based on different waveforms	• Subject must use specific bed and unable to be used in patients who do not sleep alone	• Able to estimate TST% with variable accuracy depending on distance between the sensor and scratch site in healthy subjects
Pressure sensor	• Able to detect finger scratching if placed on dorsal hand along metatarsal bone (ceramic sheet)	• Performance dependent on technology • Eg, false positives from any hand movement that causes changes in pressure • Limited research	• Able to estimate TST% with high accuracy (low RMSE) due to distinct waveforms in healthy subjects
Strain gauge	• Higher sensitivity for finger-dominant scratching when placed on index finger compared to actigraphy	• False positives with non-scratch finger bending movements • Limited research	• Able to estimate TST% with good accuracy in healthy subjects

AD, Atopic dermatitis; *BRNN*, bidirectional recurrent neural network; *EASI*, Eczema Area and Severity Index; *IGA*, Investigator's Global Assessment; *N/A*, not available; *OSA*, obstructive sleep apnea; *RMSE*, root mean square error; *SCORAD*, SCORing Atopic Dermatitis; *TST*, total scratch time; *TST%*, total sleep time percentage.

and tested in healthy subjects. The devices yielded an RMSE of 2.41% for TST% calculation, half that of wrist actigraphy.[12] The devices also yielded a TST% error of 1.38% when automatically extracted via an algorithm proposed by Shino et al,[37] which was compared to a TST% error of −1.54% when the data were visually scored. No sensitivity or specificity values were reported. It should be noted that strain gauges may be more susceptible to false positives (eg, nonscratching finger movements).

DISCUSSION

While the development of existing and novel devices has progressed tremendously, their performances reveal large areas in need of improvement. Actigraphy-based algorithms appear to have good sensitivity and specificity in healthy subjects; however, their performance deteriorates considerably when applied to pruritic subjects. This may be due to a lack of algorithm generalizability and failure to capture finger scratching. Additionally, most data used for establishing scratch parameters were obtained from small healthy samples. While there have been cross-validation studies with data from small AD samples, testing in larger samples of pruritic patients has not been performed. The same principle applies to newer scratch technologies, whereby further testing in both populations is needed for robust algorithms. While certain devices have demonstrated greater sensitivity for detecting finger scratching, the studies do not explicitly mention their abilities to detect rubbing or use of other scratching tools (eg, back scratchers). Rubbing, like scratching, is a natural reaction to itch; if devices are unable to distinguish rubbing or use of scratching tools from other motions, they may be underestimating itch. Further development of these technologies may help provide a more comprehensive picture of itch.

Performance metrics and algorithms

With advances in machine learning, data-driven approaches for objective scratch monitoring have gained significant interest. Various metrics have been employed to evaluate performance. While specificity and accuracy are useful, they need to be used with caution as they can be prone to class imbalances. Under typical situations, scratching arises sporadically, each over a brief period, ranging from a few seconds to several minutes depending on symptom severity. Thus, the majority of data collected features nonscratching behaviors; only a small amount of data feature scratching, resulting in a significant class imbalance. For example, a poor classification algorithm that predicts nonscratch all the time will, most likely, produce excellent accuracy and specificity. Given this problem, other metrics, such as sensitivity, precision, and F1-score are deemed more appropriate to quantify performance.

Future considerations

While patient history and examination remain important tools in assessing itch, there remains an ongoing need for adjunctive objective and precise, tools to quantify itch, such as in the case of subconscious habitual scratching. Many technologies and algorithmic strategies have been studied, though their performances are highly variable, with validation studies rarely extending beyond small samples. In addition, most studies focus on nocturnal scratching. Given that the perception of itch varies during the day, daytime scratching remains an important behavior that is largely unstudied.

In this review, very few studies reported specificity values. While this is understandable in nocturnal scratching, during which the targeted behavior scratching is rare overall, daytime wear introduces other confounders, such as texting or walking. Thus, specificity may hold greater relevance in daytime wear, during which the wristwatch-based systems may struggle to differentiate scratching from other movements. Our group has introduced a novel mechano-acoustic skin device that incorporates actigraphy and acoustic detection of scratching by conforming to the dorsal hand and sampling at higher frequencies (~1600 Hz) compared to actigraphy (20-100 Hz). Scratch algorithm development performed in healthy subjects yielded high sensitivity and specificity with comparable performance among AD datasets using an IR camera gold standard, even with confounders.[38,39] A comparison of data outputs for scratch from actigraphy, smartwatch application, and mechano-acoustic device is shown in Supplemental Fig 1.

CONCLUSION

While actigraphy remains the most frequently studied modality in clinical studies, performance is variable with no assessment of daytime performance. Further testing of these technologies will be needed before used in the clinical setting. A reliable technological modality would allow for objective support of drug development outcomes,[40-42] guide disease management, and assess treatment response.

Conflicts of interest

Drs Yang, Nguyen, Li, Lee, Chun, Wu, Fishbein, and Paller have no conflicts of interest to declare. Dr Xu has equity in a private company with a commercial interest in

scratch sensors and inventorship interest in patents related to a scratch sensor.

REFERENCES

1. Yosipovitch G, Greaves MW, Schmelz M. Itch. *Lancet*. 2003; 361(9358):690-694.
2. Fishbein AB, Vitaterna O, Haugh IM, et al. Nocturnal eczema: review of sleep and circadian rhythms in children with atopic dermatitis and future research directions. *J Allergy Clin Immunol*. 2015;136(5):1170-1177.
3. Jeon C, Yan D, Nakamura M, et al. Frequency and management of sleep disturbance in adults with atopic dermatitis: a systematic review. *Dermatol Ther (Heidelb)*. 2017;7(3):349-364.
4. Lavery MJ, Stull C, Kinney MO, Yosipovitch G. Nocturnal pruritus: the battle for a peaceful night's sleep. *Int J Mol Sci*. 2016;17(3):425.
5. Bender BG, Ballard R, Canono B, Murphy JR, Leung DY. Disease severity, scratching, and sleep quality in patients with atopic dermatitis. *J Am Acad Dermatol*. 2008;58(3):415-420.
6. Pereira MP, Ständer S. Measurement tools for chronic pruritus: assessment of the symptom and the associated burden: a review. *Itch*. 2019;4(4):e29.
7. Price A, Cohen DE. Assessment of pruritus in patients with psoriasis and atopic dermatitis: subjective and objective tools. *Dermatitis*. 2014;25(6):334-344.
8. Murray CS, Rees JL. 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*. 2011;91(1):18-23.
9. de Jong AE, Bremer M, Schouten M, Tuinebreijer WE, Faber AW. Reliability and validity of the pain observation scale for young children and the visual analogue scale in children with burns. *Burns*. 2005;31(2):198-204.
10. Noro Y, Omoto Y, Umeda K, et al. Novel acoustic evaluation system for scratching behavior in itching dermatitis: rapid and accurate analysis for nocturnal scratching of atopic dermatitis patients. *J Dermatol*. 2014;41(3):233-238.
11. Okuyama T, Hatakeyama K, Tanaka M. Measurement of human scratch behavior using compact microphone. *Int J Appl Electrom*. 2014;45(1):731-737.
12. Kurihara Y, Kaburagi T, Watanabe K. Development of a non-contact sensing method for scratching activity measurement. *IEEE Sens J*. 2013;13(9):3325-3330.
13. Kurihara Y, Kaburagi T, Watanabe K, Tanaka H. Development of vibration sensing system with wide dynamic range: monitoring of scratching and turning-over motions during sleep. *Artif Life Robotics*. 2015;20(4):372-378.
14. Endo K, Sumitsuji H, Fukuzumi T, Adachi J, Toshiyuki A. Evaluation of scratch movements by a new scratch-monitor to analyze nocturnal itching in atopic dermatitis. *Acta Derm Venereol (Stockh)*. 1997;77:432-435.
15. Felix R, Shuster S. A new method for the measurement of itch and the response to treatment. *Br J Dermatol*. 1975;93(3):303-312.
16. Kaburagi T, Kurihara Y. Algorithm for estimation of scratching time. *IEEE Sens J*. 2017;PP(99):1.
17. Kogure T, Ebata T. Activity during sleep measured by a sheet-shaped body vibrometer and the severity of atopic dermatitis in adults: a comparison with wrist actigraphy. *J Clin Sleep Med*. 2018;14(2):199-204.
18. Kurihara Y, Kaburagi T, Watanabe K. Sensing method of patient's body movement without attaching sensors on the patient's body: evaluation of ''scratching cheek,'' ''turning over and scratching back,'' and ''scratching shin.''. *IEEE Sens J*. 2016; 16(23):1.
19. Ebata T, Aizawa H, Kamide R. An infrared video camera system to observe nocturnal scratching in atopic dermatitis patients. *J Dermatol*. 1996;23(3):153-155.
20. Ebata T, Aizawa H, Kamide R, Niimura M. The characteristics of nocturnal scratching in adults with atopic dermatitis. *Br J Dermatol*. 1999;141(1):82-86.
21. Benjamin K, Waterston K, Russell M, Schofield O, Diffey B, Rees JL. The development of an objective method for measuring scratch in children with atopic dermatitis suitable for clinical use. *J Am Acad Dermatol*. 2004;50(1):33-40.
22. Phillips B, Ball C, Sackett D, et al. Oxford Centre for Evidence-Based Medicine: levels of Evidence (March 2009). Centre for Evidence-Based Medicine (CEBM), University of Oxford. Accessed November 11, 2020. https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009
23. Feuerstein J, Austin D, Sack R, Hayes TL. Wrist actigraphy for scratch detection in the presence of confounding activities. *Annu Int Conf IEEE Eng Med Biol Soc*. 2011;2011:3652-3655.
24. Moreau A, Anderer P, Ross M, et al. Detection of nocturnal scratching movements in patients with atopic dermatitis using accelerometers and recurrent neural networks. *IEEE J Biomed Health Inform*. 2018;22(4):1011-1018.
25. 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*. 2001;144(2): 305-309.
26. Bringhurst C, Waterston K, Schofield O, Benjamin K, Rees JL. Measurement of itch using actigraphy in pediatric and adult populations. *J Am Acad Dermatol*. 2004;51(6):893-898.
27. Fujita H, Nagashima M, Takeshita Y, Aihara M. Correlation between nocturnal scratch behavior assessed by actigraphy and subjective/objective parameters in patients with atopic dermatitis. *Eur J Dermatol*. 2014;24(1):120-122.
28. Almazan T, Craft N, Torres J, et al. High-resolution actigraphy and advanced signal processing objectively quantifies nocturnal scratching events in patients with atopic dermatitis. *J Am Acad Dermatol*. 2016;74(5):AB87.
29. Petersen J, Austin D, Sack R, Hayes TL. Actigraphy-based scratch detection using logistic regression. *IEEE J Biomed Health Inform*. 2013;17(2):277-283.
30. Hon KL, Lam MC, Leung TF, et al. Nocturnal wrist movements are correlated with objective clinical scores and plasma chemokine levels in children with atopic dermatitis. *Br J Dermatol*. 2006;154(4):629-635.
31. Hon KL, Lam MC, Wong KY, Leung TF, Ng PC. Pathophysiology of nocturnal scratching in childhood atopic dermatitis: the role of brain-derived neurotrophic factor and substance P. *Br J Dermatol*. 2007;157(5):922-925.
32. Sandoval LF, Huang K, O'Neill JL, et al. Measure of atopic dermatitis disease severity using actigraphy. *J Cutan Med Surg*. 2014;18(1):49-55.
33. Wootton CI, Koller K, Lawton S, O'Leary C, Thomas KS; SWET study team. Are accelerometers a useful tool for measuring disease activity in children with eczema? Validity, responsiveness to change, and acceptability of use in a clinical trial setting. *Br J Dermatol*. 2012;167(5):1131-1137.
34. Lee J, Cho D, Song S, Kim S, Im E, Kim J. Mobile system design for scratch recognition. In: Begole B, Kim J, Inkpen K, Woo W, eds. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors Computing Systems*. Association for Computing Machinery; 2015:1567-1572.
35. Lee J, Cho D, Kim J, et al. Itchtector: a wearable-based mobile system for managing itching conditions. In: Mark G, Fussell S,

Lampe C, et al., eds. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery; 2017:893-905.

36. Ikoma A, Ebata T, Chantalat L, et al. Measurement of nocturnal scratching in patients with pruritus using a smartwatch: initial clinical studies with the Itch Tracker app. *Acta Derm Venereol*. 2019;99(3):268-273.
37. Shino T, Kurihara Y, Nukaya S, Watanabe K, Tanaka H. Signal processing method for extracting scratching time. In: Ao SI, Castillo O, Douglas C, Feng DD, Lee JA, eds. *Proceedings of the International MultiConference of Engineers and Computer Scientists*. Newswood Limited; 2012:1141-1146.
38. Chun KS, Kang YJ, Lee JY, et al. A skin-conformable wireless sensor to objectively quantify symptoms of pruritus. *Sci Adv*. 2021;7(18):eabf9405.
39. Jo HH, Kim J, Lee JY, et al. OP25: using motion, sound, and machine learning to measure scratch with a skin-mounted, soft, wireless and flexible acoustomechanic sensor: performance with confounding activities. *Itch Abstracts*. 2019;4:1-62.
40. Wollenberg A, Howell MD, Guttman-Yassky E, et al. Treatment of atopic dermatitis with tralokinumab, an anti-IL-13 mAb. *J Allergy Clin Immunol*. 2019;143(1):135-141.
41. Guttman-Yassky E, Brunner PM, Neumann AU, et al. Efficacy and safety of fezakinumab (an IL-22 monoclonal antibody) in adults with moderate-to-severe atopic dermatitis inadequately controlled by conventional treatments: a randomized, double-blind, phase 2a trial. *J Am Acad Dermatol*. 2018;78(5):872-881.e6.
42. Silverberg JI, Simpson EL, Thyssen JP, et al. Efficacy and safety of abrocitinib in patients with moderate-to-severe atopic dermatitis: a randomized clinical trial. *JAMA Dermatol*. 2020;156(8):863-873.
43. Bender BG, Leung SB, Leung DY. Actigraphy assessment of sleep disturbance in patients with atopic dermatitis: an objective life quality measure. *J Allergy Clin Immunol*. 2003;111(3):598-602.