Mental Health and Physical Activity 29 (2025) 100728

Contents lists available at ScienceDirect

Mental Health and Physical Activity

journal homepage: www.elsevier.com/locate/menpa

Ecological momentary assessment of physical activity and affective responses in healthy adults: a scoping review

Benoit Plante [a,*], Paule Miquelon [a], Pier-Olivier Caron [b]

[a] Department of Psychology, Université du Québec à Trois-Rivières, 3351 boulevard des Forges, Trois-Rivières, QC, G9A 5H7, Canada
[b] Department of Human Sciences, Arts and Communication, Université TÉLUQ, 5800 St Denis, Montréal, QC, H2S 3L4, Canada

A B S T R A C T

Background: Affective experiences, including emotions and moods, are central to well-being. Physical activity (PA) is linked to improved mood, but the acute relationship between objectively measured PA and affect in daily life remains underexplored. Ecological Momentary Assessment (EMA) minimizes biases of retrospective reports and captures real-time effects. This scoping review synthesizes studies using EMA and accelerometers to examine the acute (within 30 min) effects of PA on positive and negative affect in healthy adults, highlighting methodological diversity and future research needs.
Methods: A systematic search of MEDLINE, CINAHL, PsycINFO, and SportDiscus identified studies involving healthy adults that used EMA, accelerometer-based PA measures, and assessed acute PA effects on affect. Data extraction followed the CREMAS protocol, focusing on sample characteristics, PA and affect measures, compliance rates, and moderators.
Results: From 208 identified studies, 14 met the inclusion criteria. Studies varied in sample size, accelerometer placement, affect measurement scales, prompt frequency, and compliance, complicating comparisons. PA was consistently linked to increased arousal; findings for valence and positive affect were mixed. Negative affect tended to decrease after PA, but results were inconsistent. Several studies explored moderators, such as competence, autonomy, and social context.
Conclusions: EMA is a valuable method for studying PA and affect dynamics in everyday life. However, methodological heterogeneity calls for more standardized protocols. Future research should improve transparency in reporting, explore additional moderators, and recruit more diverse samples to enhance generalizability.

1. Introduction

Affective experiences, comprising emotions and moods, constitute fundamental components of human well-being and psychological functioning (Diener, 2009; Keyes, 2002). Physical activity (PA) has been associated with improved mood and emotional well-being (Buecker et al., 2021; Reed & Buck, 2009; Reed & Ones, 2006), thereby offering a promising avenue for promoting mental health.

Traditionally, affect has been assessed through retrospective self-reports, which are vulnerable to recall bias and heuristic-based distortions (Robinson & Clore, 2002; Stevens et al., 2020). These measures often fail to capture the dynamic and context-sensitive nature of emotions.

Over recent decades, several methods offer alternatives for assessing affect as it happens in daily life, such as the Experience Sampling Method (Larson & Csikszentmihalyi, 2014) and Ecological Momentary Assessment (EMA) (Shiffman et al., 2008). These methods provide a window into the variability of human experiences and highlight relationships that remain concealed when relying on summative or average scores from retrospective measures. Moreover, they facilitate the development of just-in-time interventions, tailored to specific triggers such as emotions, motivational states, activities, or environmental contexts (Schneider et al., 2024). Statistically, these methods offer insights at both within- and between-person levels (Curran & Bauer, 2011).

In this scoping review, we use the term EMA to refer to four defining characteristics: 1) data collection in real-world settings, 2) assessment of subjects' current state, 3) strategic selection of assessment time points, and 4) the completion of multiple assessments over time within a day (Shiffman et al., 2008). These characteristics imply that EMA protocols can take various forms depending on the research objectives. Time-contingent sampling designs signal participants to respond to inquiries at fixed or randomly selected times within a predefined timeframe, while event-contingent sampling designs prompt participants to complete questions when specific predefined events occur. Mixed designs combining time and event sampling are also viable, and protocols may vary in duration as well as the frequency of prompts.

To align affective responses with actual behavior, accelerometers provide more objective, high-resolution data on PA (e.g., intensity, frequency, and duration) (Troiano et al., 2008). Self-reports, by

* Corresponding author.
E-mail addresses: benoit.plante@uqtr.ca (B. Plante), paule.miquelon@uqtr.ca (P. Miquelon), pier-olivier.caron@teluq.ca (P.-O. Caron).

https://doi.org/10.1016/j.mhpa.2025.100728
Received 28 April 2025; Received in revised form 31 July 2025; Accepted 5 November 2025
Available online 6 November 2025

comparison, are prone to biases that include recall inaccuracies, social desirability effects, and variability in individual interpretations of activity levels (Prince et al., 2008). They often fail to capture low-intensity activities and tend to aggregate data over broad timeframes, obscuring the dynamic interplay between PA and affective states (Skender et al., 2016).

Currently, two reviews have addressed the association between PA and affect using EMA across diverse adult populations (Liao et al., 2015; Timm et al., 2024). The first conducted by Liao et al. (2015) reviewed 14 studies investigating relationships between affect, physical feeling states, and PA. Among them, only six studies used accelerometers to measure PA. The second review by Timm et al. (2024) systematically analyzed 66 studies focusing on the within-subject association between PA (measured by accelerometers) and affective well-being in daily life. Both reviews indicate that PA predicts higher positive affect and feelings of energy in the hours following activity, while findings for reductions in negative affect are mixed. They emphasized that these inconsistencies were partly due to methodological heterogeneity, such as varying PA durations, affective dimensions assessed, study designs, the use of small, non-representative samples. Liao et al. (2015), as mentioned by Kanning et al. (2013), recommended incorporating accelerometry whereas Timm et al. (2024) suggested to focus on more specific and homogeneous populations which could enhance the reliability of findings by reducing variability related to population characteristics.

In line with these recommendations, a new scoping review is warranted to map the existing literature on the acute effects of PA on affective responses using EMA in combination with accelerometry. Understanding these acute effects is particularly important, as immediate emotional responses are not only central to mental health and well-being. For example, the dual-mode model and the affective-reflective theory emphasize that in-the-moment affective responses during PA are more predictive of future engagement and emotional outcomes than delayed mood assessments (Brand & Ekkekakis, 2018; Ekkekakis, 2009). Although past reviews have documented general associations between PA and affect, none have systematically examined how EMA and accelerometry have been integrated to capture the acute effect in healthy adults.

This review focuses specifically on healthy adults, as EMA requires sustained real-time engagement that may be burdensome for children or clinical populations (Heron et al., 2017). Moreover, PA behavior and affective dynamics vary with age and health status (Carstensen et al., 2000; Caspersen et al., 2000; Vancampfort et al., 2017). A homogeneous adult sample helps improve internal validity and clarity of interpretation.

A scoping review is well suited for mapping this evolving field, given its methodological diversity and conceptual complexity (Peters et al., 2020). This review aims to (1) describe how EMA has been used to examine the acute effects of device-based PA on affect in healthy adults, and (2) offer recommendations to strengthen future research in this area.

2. Methods

2.1. Data sources and search

Identification and selection of the studies were conducted according to PRISMA Extension for Scoping review (Tricco et al., 2018). MEDLINE, CINAHL, PsycINFO, and SPORTDiscus databases were selected to retrieve the studies following the advice from a specialized librarian (University of Quebec in Trois-Rivières, Canada). The search strategy was built as a single-line Boolean string, which groups synonyms for each concept with OR (to maximize sensitivity) and links concepts with AND only where essential (to preserve specificity) (Bramer et al., 2018; McGowan et al., 2016). The Boolean string was ((Physical-activity OR Exercise OR Fitness OR Workout or Aerobic) AND (EMA OR Ecological-Momentary-Assessment OR Ambulatory-assessment OR Ambulatory-Monitoring OR ESM OR Experience-Sampling-Method OR Event-Sampling) AND (Affect OR Emotion OR Mood OR Feeling OR Mental-health OR Well-Being) AND accelerometer)) with no date restriction. The search was conducted on July 22, 2024.

2.2. Eligibility criteria and study selection

Studies were eligible for inclusion if they: (1) were published in English, (2) were published in peer-reviewed journals, (3) used EMA methodology, (4) included participants aged 18 years or older, (5) included a measure of PA by an accelerometer, and (6) included at least one measure regarding the acute effect (30 min or less) of PA on affect. Studies were excluded if: (1) participants presented a diagnosis related to mental or physical health, (2) they collected data once per day or less, and (3) they were not conducted in the participant's natural environment.

2.3. Data extraction

A data-charting form was developed by two reviewers (BP & MC) based on the CREMAS, a specific method used to report information from EMA studies (Liao et al., 2016). The two reviewers extracted and filled out the form independently. After extracting the information from three studies, the reviewers met and discussed extraction to make sure they shared a common understanding of the charting process. At the end, extracted information was compared and discussed and a consensus was reached regarding differences in the charting process.

Data extraction included general information about study characteristics including sample size, mean age, outcomes, and measures. Following recommendations from CREMAS, information about the technology used, prompt design, wave duration, monitoring period, prompt frequency, and compliance were extracted. Finally, the relationship between PA and affect was also detailed.

3. Results

3.1. Scientific literature search

The scientific literature search yielded 208 studies. Following the removal of duplicates using EndNote software, two reviewers (BP & MC) screened independently 87 studies against inclusion and exclusion criteria based on title and abstract. Studies that were excluded by only one reviewer were further revised for consensus. Then, the same process was applied to 22 articles based on the full text to reach a consensus on the 12 studies included in this scoping review. During this process, five studies were removed because they did not measure the association between PA and affect, four because they included PA measures over more than 30 min before the prompt and one study was an intervention one. Additionally, we search Google Scholar (10 first pages) with the same key words and found one extra article. Then, a systematic review on similar subject (Timm et al., 2024) was conducted to screen for other articles and, as a result, one article was included. Using this procedure, 14 articles were included in this scoping review. Fig. 1 presents a flow chart of the systematic literature search, according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

3.2. Sampling and measure

The sample characteristics of each included study are presented in Table 1. Half of the studies ($k = 7$), focused on adults from the general population. The other seven focused on a more specific population: five used a sample comprised of university students and two used a sample including only older adults. Most studies ($k = 9$) were conducted in Germany while four were conducted in the United States and one in Scotland. Samples were predominantly from European and North

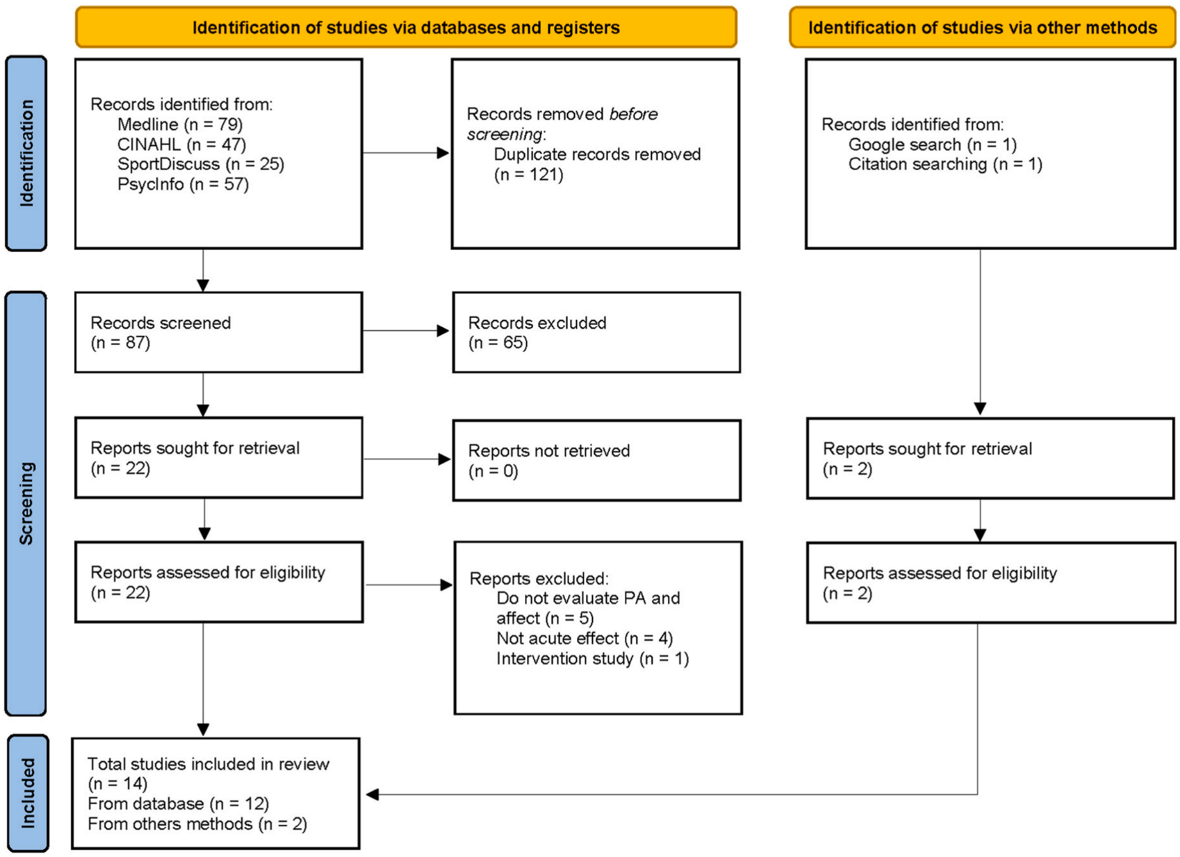

Fig. 1. Study selection flowchart.

Table 1
Characteristics of the sample included in the studies.

Study	Country	Sample	Sample N	Mean age	SD	Female %
Seiferth et al. (2024)	Germany	Adults with higher bodyweight	157	43.7	15.9	68
Li et al. (2022)	Scotland	University students and employes	78	25.5	6.2	73.1
Hevel et al. (2021)	United States	Older adults	103	72.4	7.4	62.5
Kim et al. (2020)	United States	Adults from general population	111	41.3	11.6	76.4
Sudeck et al. (2018)	Germany	Adults from general population	64	35.5	12.3	57.8
Kanning and Hansen (2017)	Germany	Older adults	68	60.1	7.1	49
Liao et al. (2017)	United States	Adults from general population	110	40.4	9.7	70
Reichert et al. (2017)	Germany	Adults from general population	93	23.4	2.7	62.4
Jeckel and Sudeck (2017)	Germany	Adults from general population	46	32	10.2	54.6
Dunton et al. (2015)	United States	Adults from general population	116	40.5	9.5	72.2
Kanning (2013)	Germany	Undergraduate students	87	24.6	3.2	54
von Haaren et al. (2013)	Germany	Undergraduate students	29	21.3	1.7	NR
Bossmann et al. (2013)	Germany	Students	62	21.4	1.8	15
Kanning et al. (2012)	Germany	Undergraduate students	44	26.2	3.2	48

American and may therefore not be representative of other cultures. The number of participants in the studies ranged from 29 to 157 ($M = 83.43$, $SD = 34.47$). The mean age ranged from 21.3 to 72.4 ($M = 38.77$, $SD = 17.25$). The proportion of females ranged from 15 % to 76.4 % with a mean of 58.69 %.

Measures of affect and PA are presented in Table 2. Most studies ($k = 9$) assessed affect using the full six-item Short Mood Scale (Wilhelm & Schoebi, 2007). This scale assesses valence, calmness, and arousal using six bipolar items: *content–discontent*, *unwell–well* (valence); *agitated–calm*, *relaxed–tense* (calmness); and *tired–awake*, *full of energy–without energy* (arousal). The within-person reliability coefficients are 0.70 for both valence and calmness, and 0.77 for arousal. One study measured affect using two dimensions: valence and arousal (Kim et al., 2020). Valence was assessed with two items, *happy* and *sad* (reverse-coded), while arousal was measured using *interested* and *tired* (reverse-coded). No information pertaining to the scale's reliability is reported. The four other studies used different items reflecting positive and negative affect. One assessed positive affect using two items: *I feel happy* and *I enjoy what I am doing* while negative affect was measured with six items: *I feel sad*, *I feel irritable*, *I am restless*, I *feel guilty*, *I feel worthless*, and *I feel hopeless* (Li et al., 2022). The reliability coefficients are 0.80 for positive affect and 0.86 for negative affect, however there is no information if this is between or within reliability. Finally, Liao et al. (2017) assessed positive affect with *happy*, *cheerful*, and *calm or relaxed*, and negative affect with *anxious*, *stressed*, *depressed*, and *angry*. The reliability coefficients are 0.837 for positive affect and 0.865 for negative affect. Dunton et al. (2015) and Hevel et al. (2021) used the same five-item scale: *happy*, *calm/relaxed*, *energetic*, *anxious*, and *stressed*. The reliability coefficients are, respectively, 0.77 and 0.81 for positive affect and 0.76 and 0.78 for negative affect. Those three studies did not report

Table 2
Characteristics of technology used and measures.

Study	Accelerometers	PA Measure	Timing PA	EMA (Hardware)	EMA (Software)	Affect Measure
Seiferth et al. (2024)	Actigraph wGT3X-BT (hip)	Vector magnitude	15 min	Personal smartphone	I-GENDO app	Six-item mood scale (Wilhelm & Schoebi, 2007)
Li et al. (2022)	Move III sensor (chest)	MET	30 min	Android phone (personal or lab)	MovisensXS	Positive/Negative affect items
Hevel et al. (2021)	activPAL (thigh)	Time stepping Time standing	15 min and 30 min	Motoral G4	MovisensXS	Positive/Negative affect items
Kim et al. (2020)	ActiHeart (chest)	Activity counts	5 min,	Palmtop device (unidentified)	NR	Positive/Negative affect items
Sudeck et al. (2018)	Move III sensor (hip)	milli-g	15 min	Google Nexus 5	MovisensXS	Six-item mood scale (Wilhelm & Schoebi, 2007)
Kanning and Hansen (2017)	Varioport-e (hip)	milli-g	10 min	HTC Touch 2	MyExperience	Six-item mood scale (Wilhelm & Schoebi, 2007)
Liao et al. (2017)	Actigraph GT2 M (hip)	Counts per minute	15 and 30 min	HTC Shadow	MyExperience	Positive/Negative affect items
Reichert et al. (2017)	Move II sensor (hip)	milli-g	15 min	Motorola Moto G	MovisensXS	Six-item mood scale (Wilhelm & Schoebi, 2007)
Jeckel and Sudeck (2017)	Move III sensor (chest)	MET	15 min	HTC Touch Diamond	MyExperience	Six-item mood scale (Wilhelm & Schoebi, 2007)
Dunton et al. (2015)	Actigraph GT2M (hip)	Counts per minute	15 min	HTC Shadow	MyExperience	Positive/Negative affect items
Kanning (2013)	Varioport-e (hip)	milli-g	10 min	Palm Tungsten E2	NR	Six-item mood scale (Wilhelm & Schoebi, 2007)
von Haaren et al. (2013)	Move II sensor (chest)	MET	15 min	HTC Touch 2	MyExperience	Six-item mood scale (Wilhelm & Schoebi, 2007)
Bossmann et al. (2013)	Move I sensor (chest)	milli-g	10 min	HTC Touch 2	MyExperience	Six-item mood scale (Wilhelm & Schoebi, 2007)
Kanning et al. (2012)	Varioport-e (hip)	milli-g	10 min	Palm Tungsten E2	Izybuilder	Six-item mood scale (Wilhelm & Schoebi, 2007)

if reliability was assessed at between or within level.

As for PA measurement, the studies varied in how they computed data from accelerometers. Six studies used raw acceleration (milli-g), three studies used the metabolic equivalent of task (MET), and two studies measured counts per minute. Additionally, vector magnitude, activity counts, and time spent stepping/standing were each utilized in one study. Regarding the timing considered for the measure of PA, studies examined the following timeframes: 30 min ($k = 3$), 15 min ($k = 8$), 10 min ($k = 4$) and 5 min ($k = 1$). before the notification. Two studies considered two timing: 15 and 30 min. As for placement, the accelerometer was worn on the hip in eight studies, on the chest in five studies, and on the thigh in one study.

Five studies reported epoch length: one used 4 s (Von Haaren et al., 2013), and four used 60 s (Bossmann et al., 2013; Hevel et al., 2021;

Table 3
Prompting strategy and compliance.

Study	Prompt Approach	Monitoring Periods	Days per Period	Prompt per Day	Prompt Interval	Compliance (%)
Seiferth et al. (2024)	Interval-design	1	7	8	Fixed - Eight 90 min blocks, 30 min apart. Starting time between 7:30 and 10:30.	NR
Li et al. (2022)	Interval-design	1	14	5	Random - First prompt between 8:00–11:30 a.m., second, third and fourth between 13:00–19:00, last prompt between 21:00–23:00 (minimal gap of 2h between prompt).	76.8
Hevel et al. (2021)	Interval-design	1	7	6	Fixed - Every two and a half hour starting at 9:00 a.m. and ending at 9:30 p.m.	86
Kim et al. (2020)	Interval-design	1	3	6	Random - Within roughly two and a half hour intervals.	NR
Sudeck et al. (2018)	Interval-design	1	4	4	Random - One time per time bloc (9:30–10:30 a.m., 12:30–1:30 p.m., 3:30–4:30 p.m., and 6:30–7:30 p.m.).	76.3
Kanning and Hansen (2017)	Activity-triggered	1	3	Average of 6.4	Prompted when the accelerometer measured a volume of physical activity that surpassed a predefined activity threshold.	NA
Liao et al. (2017)	Interval-design	1	4	8	Random - Within eight pre-programmed windows (between 6:30 a.m. and 10:00 p.m.).	82
Reichert et al. (2017)	Mixed-sampling strategy	1	7	9 to 22	Random - Between 8:00 a.m. and 10:20 p.m., no more often than every 40 min and at least every 100 min. Location had to been more than 500m from last prompt.	81.2
Jeckel and Sudeck (2017)	Interval-design	1	7	3	Random - One time per time bloc (7:30–8:30 a.m., 12:30–1:30 p.m., and 5:30–6:30 p.m.).	76
Dunton et al. (2015)	Interval-design	3	4	8	Random - Within eight pre-programmed windows (between 6:30 a.m. and 10:00 p.m.).	83
Kanning (2013)	Interval-design	1	1	NR	Random - About every 45 min during a defined 14-h daytime period (8:00 a.m. to 10:00 p.m.).	97.5
von Haaren et al. (2013)	Interval-design	1	2	6	Random - Approximately every 2 h between 12:00 a.m. and 10:00 p.m.	NR
Bossmann et al. (2013)	Interval-design	1	1	Average of 10.5	Fixed - Starting when waking up, then every full. Finish when participants cannot fill a particular assessment point.	NA
Kanning et al. (2012)	Interval-design	1	14	an average of 19	Random - About every 45 min during a defined 14-h daytime period (8:00 a.m. to 10:00 p.m.).	NR

Jeckel & Sudeck, 2017; Reichert et al., 2017). Eight studies applied person-mean centering for within-person associations. One used grand-mean centering (Kanning et al., 2013). Five studies did not report any transformation approach.

3.3. Technology and administration

All information regarding technology and administration can be found in Table 2. In most studies, smartphones or handheld devices were provided to participants. In one study, participants could use either a study-issued smartphone or their own; another study had participants used their personal smartphone. The MyExperience software was used in six different studies, while four studies utilized MovisensXS. Two studies employed custom-built platforms, and two did not report the softwa

Regarding accelerometer technology, six studies used different versions of the Move sensor: Move I ($k = 1$), Move II ($k = 2$), Move III ($k = 3$). Other studies used the Varioport-e ($k = 3$), Actigraph ($k = 3$), Actiheart ($k = 1$), and ActivPal ($k = 1$).

3.4. Schedule, prompting strategy and compliance

Table 3 summarizes the information about scheduling, prompting, and compliance. All studies, except one, monitored participants during only one period of time. One study used a three-wave data collection. The duration of the monitoring period was on average 5.92 days ($SD = 4.11$). The frequency of the monitoring periods was: one day ($k = 4$), two days ($k = 1$), three days ($k = 2$), four days ($k = 2$), seven days ($k = 6$), and 14 days ($k = 2$). Prompting frequency was reported in 13 of 14 studies and ranged from 3 to 22 prompts per day ($Mdn = 6.2$).

Most ($k = 12$, 85.7 %) of the studies used an interval-contingent prompting strategy. Of these 12 studies, three used fixed interval contingent and nine used random interval contingent only. One study used an activity-triggered strategy based on a predefined level of PA and one study used a mixed approach using a random interval contingent paired with a location-triggered approach.

Out of the 14 studies included, eight reported a compliance rate. This rate ranges from 76 % to 97.5 %, averaging 82.35 % ($SD = 7.1$). The highest compliance was observed with the shortest protocol (1 day).

3.5. Results of the studies

Table 4 presents the results regarding the relation between PA and affect. PA showed varied relationships with valence, arousal, and calmness. For valence, several studies report non-significant relations (Jeckel & Sudeck, 2017; Kanning & Hansen, 2017; Kim et al., 2020; Reichert et al., 2017; Seiferth et al., 2024). However, positive relations are observed in four studies (Bossmann et al., 2013; Kanning, 2013; Kanning et al., 2012; Sudeck et al., 2018), suggesting that PA can enhance valence under certain conditions. Regarding arousal, PA mostly has a positive effect (Bossmann et al., 2013; Jeckel & Sudeck, 2017; Kanning, 2013; Kanning et al., 2012; Kanning & Hansen, 2017; Reichert et al., 2017; Seiferth et al., 2024; Sudeck et al., 2018), indicating that PA typically increases arousal levels. Two studies did not report significant results (Kim et al., 2020; Von Haaren et al., 2013). In contrast, calmness is often negatively associated with PA (Kanning et al., 2012; M. Kanning, 2013; Kanning & Hansen, 2017; Reichert et al., 2017; Seiferth et al., 2024; Sudeck et al., 2018), suggesting that engaging in PA might reduce feelings of calmness. Three studies did not find a significant association.

When it comes to positive affect, one study found a significant positive relation (Li et al., 2022). However, three other studies reported non-significant effects (Dunton et al., 2015; Hevel et al., 2021; Liao et al., 2017). For negative affect, two studies found a significant negative relation indicating that PA tends to reduce negative feelings (Li et al., 2022; Liao et al., 2017). However, the two other studies (Dunton et al., 2015; Hevel et al., 2021) reported non-significant effects, reflecting the variability in PA's impact on negative emotions.

Table 4
Summary of study results on the association between PA and affective responses.

Study	Results Related to the Affect
Seiferth et al. (2024)	PA → valence = ns PA → arousal = + PA → calmness = -
Li et al. (2022)	PA → positive affect = + PA → negative affect = -
Hevel et al. (2021)	Stepping → positive affect = ns Stepping → negative affect = ns Standing → positive affect = ns Standing → negative affect = ns
Kim et al. (2020)	PA → valence = ns PA → arousal = ns
Sudeck et al. (2018)	PA → valence = + PA x competence → valence = + PA → arousal = + PA x competence → arousal = ns PA → calmness = ns PA x competence → calmness = -
Kanning and Hansen (2017)	PA → valence = ns PA x autonomy → valence = + PA x competence → valence = ns PA x relatedness → valence = ns PA → arousal = + PA x autonomy → arousal = + PA x competence → arousal = - PA x relatedness → arousal = ns PA → calmness = - PA x autonomy → calmness = ns PA x competence → calmness = ns PA x relatedness → calmness = +
Liao et al. (2017)	LPA → positive affect = na LPA → negative affect = - MVPA → positive affect = ns MVPA → negative affect = ns
Reichert et al. (2017)	NEA → valence = ns NEA → arousal = + NEA → calmness = -
Jeckel and Sudeck (2017)	PA → valence = ns PA → arousal = + PA → calmness = ns
Dunton et al. (2015)	PA → positive affect = ns PA → negative affect = ns PA x Alone → positive affect = - PA x Alone → negative affect = ns PA x Outdoors → positive affect = ns PA x Outdoors → negative affect = -
Kanning (2013)	PA → valence = + PA x LW→ valence = ns PA → arousal = + PA x LW→ arousal = ns PA → calmness = - PA x LW → calmness = ns
von Haaren et al. (2013)	PA → valence = ns PA → arousal = ns PA → calmness = ns
Bossmann et al. (2013)	PA → valence = + PA → arousal = + PA → calmness = ns
Kanning et al. (2012)	PA → valence = + PA x autonomy→ valence = ns PA → arousal = + PA x autonomy→ arousal = + PA → calmness = - PA x autonomy → calmness = -

Note. + = significant positive association; – = significant negative association; ns = non-significant association. PA = physical activity; LPA = light physical activity; MVPA = moderate-to-vigorous physical activity; NEA = non-exercise activity; Stepping/Standing = specific posture-related activity measured via accelerometer.

Five studies also examined the interaction effect of specific moderators in the relationship between PA and affect. Of these studies, four used valence, arousal and calmness as an outcome. The first study examined perceived competence associated with PA as a moderator and found a significant interaction for valence and calmness but found a non-significant interaction for arousal (Sudeck et al., 2018). The second study examined basic psychological needs, proposed by Self-Determination Theory (Ryan & Deci, 2017) as moderators and found a significant interaction for autonomy with valence and arousal, for competence with arousal, and for relatedness with calmness (Kanning & Hansen, 2017). All other interactions were not significant. The third study only examined the need for autonomy as a moderator and found that its interaction effect was only significant for arousal (Kanning et al., 2012). Finally, one study used the context of PA (leisure/work) as a moderator and did not find any significant interaction (Kanning, 2013).

The fifth study assessed positive and negative affect as an outcome. This study considered being alone or with someone as well as being indoors or outdoors during PA and found a significant interaction for being with someone regarding positive affect (Dunton et al., 2015). However, the authors found no interaction for being indoors or outdoors regarding positive affect while it found a significant interaction between PA and being outdoors for negative affect, but not for positive affect.

4. Discussion

The aim of this scoping review was to explore and describe how EMA is used to examine the relation between PA, as measured with accelerometers, and affective responses. Fourteen studies were selected based on PRISMA guidelines (Tricco et al., 2018) and their data were reported using the CREMAS protocol (Liao et al., 2016). Overall, the studies included were highly heterogeneous, showing an important variation in sample size, technology, duration, and the number of prompts per day. It is therefore difficult to draw firm conclusions about the relationship between PA and affect. That being said, the average compliance rate of 82.35 %, reported in eight studies, seems to indicate that EMA can be an acceptable and feasible method to assess the relationship between PA and affect.

4.1. Sampling

While there is no definitive consensus on the exact sample size required for multilevel model analyses in EMA studies, it largely depends on the complexity of the model and the specific research questions. Simulation studies suggest that under certain conditions, a sample size as small as 30 participants may provide stable and reliable estimates, especially when the models are relatively simple and the number of observations per participant is high (Hox & McNeish, 2020). However, for most multilevel analyses, including those with more complex structures or interactions, a sample size of approximately 50 participants is generally considered sufficient to yield trustworthy and accurate estimates (Maas & Hox, 2005). In this context, all studies, except the one with 29 participants, meet or exceed these recommendations from simulation studies, supporting the reliability of their results. It is important to note that beyond the number of participants, the total number of observations (i.e., measurement occasions per participant) is crucial in ensuring the stability of multilevel model estimates. Therefore, even with slightly smaller sample sizes, an increased number of observations per participant can compensate for the reduced participant count, enhancing the robustness of the findings.

4.2. Affect measurement

Most of the studies rely on the Six-Item Short Mood Scale to index affect. This instrument is brief and has shown good within-person reliability in its original validation (Wilhelm & Schoebi, 2007). A strength of this choice is that the scale covers three core dimensions of the circumplex model (valence, arousal, and calmness) while minimizing participant burden. Nevertheless, most articles quote the original reliability coefficient rather than recalculating it in their own samples. Because reliability is context-dependent failing to re-estimate it can hide attenuation that weakens observed PA–affect associations (Brose et al., 2020). For example, recent work shows that within-person omega can fall by 0.10–0.20 when the same items are delivered six or more times per day, particularly for low-arousal states that vary less moment-to-moment (Haney et al., 2023). Thus, apparent null effects may sometimes reflect measurement error rather than the absence of an underlying relationship.

A second methodological limitation concerns the studies that employed ad-hoc adjective lists to assess positive and negative affect. Although many of these studies reported a reliability coefficient, few clarified whether it reflected within-person or between-person consistency. This distinction matters: a scale can rank individuals reliably across days (high between-person reliability) yet still provide noisy moment-to-moment scores (low within-person reliability), thereby diluting event-level PA effects (Geldhof et al., 2014).

A related concern is the item composition of these ad-hoc lists. Scales can be dominated by high-arousal terms, by low-arousal terms, and or achieve a balanced mix. Composition matters, as it impacts the sensitivity of the scale because high-arousal states show greater within-person variance, increasing power to detect event-level PA effects, whereas low-arousal items vary little over short intervals and thus dilute the signal (Cranford et al., 2006; Haney et al., 2023).

To advance the field, future EMA studies should report both within- and between-person reliability, specifying the statistic employed (e.g., ICC(1), within-person omega, generalizability coefficients); align their measurement strategy with an explicit theoretical model of affect, choosing instruments that map onto the dimensions hypothesised to change; and disclose full item wordings, response scales, and prompts.

4.3. PA measurement

In terms of PA measurement, both the technology and the analytical approaches vary widely across studies. Some studies rely on raw accelerometer data, while others use metrics such as metabolic equivalents (METs), activity counts or time spent stepping or standing to quantify PA. Devices are mounted at the hip, wrist, thigh, or chest, and each site privileges different movement patterns: hip-worn sensors remain the reference for ambulatory locomotion, yet wrist devices capture fine motor activity often missed at the hip, while chest units coupled with heart-rate monitors better index cardiometabolic load (Kerr et al., 2017; Migueles et al., 2017). Accordingly, placement choice should be explicit, with hip, wrist and thigh remaining the most common research sites.

Beyond placement, divergent filtering options, epoch lengths, and non-wear algorithms can shift the same bout of PA into different intensity bands. For example, longer epoch lengths tend to underestimate MVPA and overestimate light activity, different cut-point definitions shift the classification boundaries between intensity levels and alter the proportion of time attributed to each, and non-wear algorithms vary in their sensitivity to inactivity periods (Burchartz et al., 2024; Haddadj et al., 2024). This methodological heterogeneity amplifies between-study variance and complicates the integration of findings on the relationship between PA and affect.

In addition to hardware and data-processing differences, variation in statistical modeling procedures such as whether or not person-mean centering was applied to PA predictors constitutes another important source of heterogeneity. Centering PA at the individual level is essential when the goal is to estimate how deviations from one's typical activity level relate to momentary affect, rather than capturing stable between-person differences. Without person-mean centering, models risk conflating within- and between-person variance, potentially leading to biased or attenuated estimates of the PA–affect relationship (Curran &

Bauer, 2011; McNeish & Kelley, 2019). Given that EMA is designed to detect temporal fluctuations, failing to center PA appropriately may obscure the real-time dynamics of this relationship. Clear reporting and justification of centering practices should therefore be standard in EMA studies seeking to model acute effects of PA on affect.

For these reasons, it is crucial for researchers to provide comprehensive information on accelerometer make and model, sampling frequency, placement site, epoch length, wear-time algorithm, filtering settings, and statistical procedures because each of these decisions can alter the quantity and intensity of activity recorded (Curran & Bauer, 2011; Kerr et al., 2017; Migueles et al., 2017). Standardizing such reporting practices would enhance the interpretability of PA–affect findings and foster a more cohesive, cumulative body of literature. Moreover, the use of raw accelerometer data and open-source metrics (e.g., ENMO, MAD) processed through open-source software, such as GGIR (Migueles et al., 2019), could provide a transparent option for processing accelerometer data.

Emerging AI-based models offer promising tools for standardizing PA classification across devices and placements. Such models can learn to infer activity types or intensity levels directly from raw tri-axial acceleration signals, often outperforming traditional cut-point methods in free-living conditions (Farrahi & Rostami, 2024; Jones et al., 2021). While not yet widely adopted in EMA studies, these techniques may reduce inter-study variability and support more nuanced behavior recognition (e.g., differentiating fidgeting from purposeful walking), thus enhancing PA–affect modelling.

Furthermore, even if no study in this scoping review used commercial wearables such as Fitbit, Apple Watch, or Garmin, it is important to acknowledge that these devices can introduce both opportunities and challenges. On one hand, these tools allow large-scale, passive data collection with high ecological validity and user compliance. On the other hand, their proprietary algorithms and opaque data processing limit reproducibility and comparability with research-grade accelerometry (Evenson et al., 2015; Fuller et al., 2020). As EMA research continues to grow, striking a balance between scalability and measurement validity will be important.

4.4. Schedule, prompting strategy and compliance

Given the varied research questions regarding the relationship between PA and affect, studies in the review used diverse scheduling and prompting strategies. However, this heterogeneity, coupled with the fact that only seven studies (50 %) reported compliance rates, makes it difficult to draw any firm conclusions about how study design affects compliance.

However, recent studies have provided recommendations to maximize compliance in EMA studies, and it appears crucial to balance data collection frequency with participant burden. Moderate sampling frequencies of 3–6 prompts per day are ideal, as more frequent prompting increases burden without significantly improving data quality. If higher frequencies are necessary, limiting the study duration to around one week can help maintain adherence (Eisele et al., 2022; Wrzus and Neubauer, 2023). Questionnaire length is another important factor. Longer questionnaires (over 60 items) reduce compliance and data quality, as participants may experience fatigue or respond carelessly. To avoid this, shorter questionnaires of 30–40 items per prompt are recommended to maintain data quality (Eisele et al., 2022; Hasselhorn et al., 2021).

Study duration also impacts compliance. As adherence tends to decrease in studies lasting more than two weeks, shorter studies are preferable. For longer studies, introducing break days or reducing the number of prompts later in the study can help sustain participation (Wrzus and Neubauer, 2023). Monetary incentives and personalized feedback have been shown to increase compliance. Offering rewards tied to compliance and providing regular feedback can keep participants engaged (Eisele et al., 2022; Hasselhorn et al., 2021). Lastly, protocol flexibility, such as allowing participants a 15–30 min window to respond, reduces perceived burden and improves adherence without compromising too much data quality. Adapting the protocol to fit participants' daily routines minimizes disruptions and encourages continued participation (Hasselhorn et al., 2021).

4.4.1. Affective responses to PA: Synthesis of findings

The results regarding the relationship between PA and affect reveal a complex and varied pattern shaped by differences in how PA and affect are measured. To interpret the heterogeneity in outcomes, we organize the findings according to the circumplex model of affect, separating valence, arousal, and calmness and positive and negative affect model.

Findings for valence were mixed and appear to depend on both methodological and measurement features. Of the 10 studies examining valence, six reported no significant effect, while four found positive associations. Notably, the studies with significant effects commonly employed hip-mounted accelerometers and relied on raw acceleration data (milli-g) rather than processed metrics like METs or activity counts. Those choices may have enhanced the sensitivity to capture PA movement and minimized distortion from post-processing algorithms (Kerr et al., 2017; Migueles et al., 2017).

By contrast, arousal demonstrated more consistent patterns: 8 out of 10 studies reported significant within-person increases following PA, regardless of methodological considerations. Null effects in two studies may stem from methodological limitations: Kim et al. (2020) was the only one not to use the Six-Item Mood Scale, while Von Haaren et al. (2013) had the smallest sample size (N = 29), limiting power. These findings align with physiological models linking PA to increased sympathetic activation and catecholamine release, mechanisms directly tied to heightened arousal (Ekkekakis, 2003).

The calmness dimension revealed a more context-dependent pattern. Of the nine studies that assessed calmness, four reported a significant post-PA decrease, while five did not. All studies with significant effects used hip-worn sensors and more frequent sampling (>6 prompts/day), suggesting these factors may enhance detection of calmness.

Interestingly, calmness showed an inverse pattern to arousal in several studies: 7 of the 9 that measured both found that as arousal increased, calmness decreased. This appears consistent with the circumplex model of affect (Feldman Barrett & Russell, 1998), which places these states on opposite ends of the activation axis. Yet, the studies used the short mood scale by Wilhelm and Schoebi (2007), which conceptualizes calmness and arousal as distinct but related dimensions. This suggests PA can shift both dimensions simultaneously, increasing alertness while also affecting relaxation depending on context, helping explain the more variable effects on calmness.

To complement the findings based on the circumplex model dimensions, several studies assessed affect using composite positive and negative affect scales. Of the four studies using this approach, one found significant PA-related increases in positive affect and decreases in negative affect (Li et al., 2022); one found a significant decrease on negative affect only for light PA (Liao et al., 2017); and the remaining two studies reported null findings (Dunton et al., 2015; Hevel et al., 2021). These inconsistencies may be driven more by differences in affect measurement than by study design.

All three studies that reported null or partial findings used a similar list of adjectives combining both high- and low-arousal terms. Hevel et al. (2021) and Dunton et al. (2015) both used five items (*happy, calm, energetic, stressed, anxious*), which blend opposite poles of the valence and arousal axes. Liao et al. (2017) used similar list of items only removing *energetic* and adding *cheerful*, *depressed*, and *angry*. These scales may reduce sensitivity to PA effects by averaging signals from emotional states with divergent reactivity.

By contrast, Li et al. (2022) employed a more focused set of items (*I feel happy* and *I enjoy what I am doing* for positive affect, and six negative items that are more consistent in arousal and valence. This narrower operationalization may enhance the ability to detect subtle within-person associations with PA.

These findings highlight a broader concern in EMA-based affect research: the use of aggregated affect scores that conflate distinct emotional processes can obscure true variation. Scales that fail to account for arousal levels may miss critical affective shifts, especially those linked to PA, which often modulates arousal more than valence alone (Barrett, 2006; Kuppens et al., 2013).

Taken together, these findings suggest that while PA consistently enhances arousal in daily life, its effects on valence, calmness, and composite affect scores are more variable and context-sensitive. This variability appears driven in part by methodological differences in both PA and affect measurement. Yet beyond measurement, these inconsistencies may also reflect genuine differences in how individuals experience PA depending on psychological, social, and contextual factors.

4.4.2. Moderating factors and theoretical perspectives

Five studies investigated moderators that might explain variability in affective responses to PA. These offer key insights into the psychological and contextual factors shaping emotional experiences during daily activity and help clarify the heterogeneity observed in the core affective findings.

Sudeck et al. (2018) examined affect regulation competence, an individual's ability to regulate emotions via PA, and found stronger increases in valence and calmness among participants with higher competence. This self-regulatory trait, grounded in social-cognitive theory, suggests that individuals differ in their capacity to use PA as a means of emotional regulation. Affective benefits may therefore hinge not just on the activity itself but on one's perceived ability to derive psychological gains from it.

Three studies by Kanning and colleagues applied Self-Determination Theory (SDT), which posits that emotional functioning is enhanced when basic psychological needs are fulfilled (Ryan & Deci, 2017). Kanning and Hansen (2017) found that perceived autonomy moderated PA's impact on valence and arousal, and relatedness moderated its effect on calmness. Kanning et al. (2012) similarly reported that autonomy amplified arousal responses, while Kanning (2013) did not observe significant interactions. Collectively, these findings suggest that PA is more likely to elicit affective benefits when it is perceived as self-chosen, meaningful, and socially embedded. This aligns with broader literature showing that intrinsically motivated PA tends to enhance affective experiences more reliably than externally imposed or constrained activities (Ryan & Deci, 2017).

Taken together, these moderator findings illustrate that affective responses to PA are shaped not only by physiological components but also by how individuals interpret the activity in context. For example, a brisk walk may be energizing and enjoyable for one person, yet stressful and fatiguing for another, depending on their prior experience, current goals, or perceived control. These subjective appraisals operate alongside biological processes and help explain why even methodologically similar studies report divergent affective outcomes.

From a theoretical standpoint, these patterns support integrative models such as the Affective-Reflective Theory (Brand & Ekkekakis, 2018), which proposes that affective responses are shaped by both automatic emotional reactions (e.g., bodily activation) and reflective appraisals (e.g., perceived purpose or self-congruence of the activity). By embedding PA within a motivational and cognitive-affective framework, these models help account for when and for whom PA enhances well-being.

Beyond the frameworks tested in the included studies, other theoretical perspectives offer promising avenues for explaining the heterogeneity in affective responses to PA. For exemple, the Dual-Mode Theory (Ekkekakis, 2003, 2009) posits that affective responses vary systematically with exercise intensity, with cognitive factors playing a greater role at moderate levels. This model may help explain why similar PA intensities yield divergent affective outcomes across individuals. In addition, the Process Model of Emotion Regulation conceptualizes PA as a regulation opportunity (e.g., attentional deployment/distraction, cognitive change/reappraisal, response modulation), whereby heightened arousal can be converted into pleasant activation and negative affect dampened in situ (Gross, 1998, 2015). Finally, the Broaden-and-Build Theory of Positive Emotions (Fredrickson, 2001) suggests that repeated experiences of positive affect during PA may build personal resources over time, enhancing future motivation and well-being. Future studies would benefit from integrating these complementary perspectives to capture both immediate and longer-term affective mechanisms.

4.5. Limitations

A major limitation of this scoping review is that it might have missed published studies that met the inclusion criteria. Although the authors included multiple terms that are used to refer to EMA methodology, it is possible that some researchers used terms that were not included. Also, the authors selected the most predominant database in psychology and PA, but they cannot rule out that some studies have been published somewhere else. Moreover, the authors included only studies published in English.

Future research should enhance methodological transparency by rigorously adhering to established reporting guidelines, such as CREMAS, and by providing comprehensive details on affect and PA activity, including both within- and between-person reliability metrics (Brose et al., 2020; Liao et al., 2016; Wilhelm & Schoebi, 2007). Standardizing accelerometer protocols is critical for improving cross-study comparability; this includes clear reporting on device placement, data processing methods, and standardized cut points for activity intensity (Aadland & Ylvisåker, 2015; Kerr et al., 2017; Migueles et al., 2017; Troiano et al., 2014). Moreover, optimizing EMA designs is essential to balance data richness with participant burden, by calibrating prompt frequencies, shortening questionnaire length, and limiting study duration to maintain high compliance (Eisele et al., 2022; Hasselhorn et al., 2021; Wrzus and Neubauer, 2023). Finally, incorporating potential moderators, such as contextual factors and individual psychological constructs, can further elucidate the complex interplay between PA and affect, thereby informing more targeted and effective interventions (Kanning et al., 2012; Kanning & Hansen, 2017; Sudeck et al., 2018).

5. Conclusion

This scoping review summarized the current knowledge regarding the use of EMA to assess the relationship between PA and affect among healthy adults from the general population. The result shows EMA to be both feasible and useful as it can contribute to gaining a better understanding of the effect of PA in daily life and the well-being of individuals. However, the reported design and results are heterogeneous, and thus, more studies are needed to reach a consensus on the relationship between PA and affect in healthy adults.

CRediT authorship contribution statement

Benoit Plante: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Paule Miquelon:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Pier-Olivier Caron:** Writing – review & editing, Supervision.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the library services of the Université du Québec à Trois-Rivières for their assistance in developing search strategies for the database queries. We are also grateful to our respective institutions for providing the necessary academic and logistical support to complete this work.

Data availability

No data was used for the research described in the article.

References

Aadland, E., & Ylvisåker, E. (2015). Reliability of objectively measured sedentary time and physical activity in adults. *PLoS One, 10*(7), Article e0133296. https://doi.org/10.1371/journal.pone.0133296

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Bossmann, T., Kanning, M., Koudela-Hamila, S., Hey, S., & Ebner-Priemer, U. (2013). The association between short periods of everyday life activities and affective states: A replication study using ambulatory assessment. *Frontiers in Psychology, 4*, 102. https://doi.org/10.3389/fpsyg.2013.00102

Bramer, W. M., Jonge, G. B. de, Rethlefsen, M. L., Mast, F., & Kleijnen, J. (2018). A systematic approach to searching: An efficient and complete method to develop literature searches. *Journal of the Medical Library Association, 106*(4), 531–541. https://doi.org/10.5195/jmla.2018.283

Brand, R., & Ekkekakis, P. (2018). Affective–reflective theory of physical inactivity and exercise: Foundations and preliminary evidence. *German Journal of Exercise and Sport Research, 48*, 48–58. https://doi.org/10.1007/s12662-017-0477-9

Brose, A., Schmiedek, F., Gerstorf, D., & Voelkle, M. C. (2020). The measurement of within-person affect variation. *Emotion, 20*(4), 677–690. https://doi.org/10.1037/emo0000583

Buecker, S., Simacek, T., Ingwersen, B., Terwiel, S., & Simonsmeier, B. A. (2021). Physical activity and subjective well-being in healthy individuals: A meta-analytic review. *Health Psychology Review, 15*(4), 574–592. https://doi.org/10.1080/17437199.2020.1760728

Burchartz, A., Kolb, S., Klos, L., Schmidt, S. C., Haaren-Mack, B. von, Niessner, C., & Woll, A. (2024). How specific combinations of epoch length, non-wear time and cut-points influence physical activity: Processing accelerometer data from children and adolescents in the nationwide MoMo study. *German Journal of Exercise and Sport Research, 54*(2), 169–178. https://doi.org/10.2196/14370

Caspersen, C. J., Pereira, M. A., & Curran, K. M. (2000). Changes in physical activity patterns in the United States, by sex and cross-sectional age. *Medicine & Science in Sports & Exercise, 32*(9), 1601–1609. https://doi.org/10.1097/00005768-200009000-00013

Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin, 32*(7), 917–929. https://doi.org/10.1177/0146167206287721

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Diener, E. (2009). Subjective well-being. In E. Diener (Ed.), *The science of well-being* (pp. 11–58). Springer. https://doi.org/10.1007/978-90-481-2350-6_2.

Dunton, G. F., Liao, Y., Intille, S., Huh, J., & Leventhal, A. (2015). Momentary assessment of contextual influences on affective response during physical activity. *Health Psychology, 34*(12), 1145–1155. https://doi.org/10.1037/hea0000223

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment, 29*(2), 136–151. https://doi.org/10.1177/1073191120957102

Ekkekakis, P. (2003). Pleasure and displeasure from the body: Perspectives from exercise. *Cognition & Emotion, 17*(2), 213–239. https://doi.org/10.1080/02699930302292

Ekkekakis, P. (2009). The dual-mode theory of affective responses to exercise in metatheoretical context II: Bodily-constructed meaning and the exercise-affect-adherence relationship. *International Review of Sport and Exercise Psychology, 2*(2), 139–160. https://doi.org/10.1080/17509840902829323

Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity, 12*(1), 159. https://doi.org/10.1186/s12966-015-0314-1

Farrahi, V., & Rostami, M. (2024). Machine learning in physical activity, sedentary, and sleep behavior research. *Journal of Activity, Sedentary and Sleep Behaviors, 3*(1), 5. https://doi.org/10.1186/s44167-024-00045-9

Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology, 74*(4), 967. https://doi.org/10.1037/0022-3514.74.4.967

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist, 56*(3), 218. https://doi.org/10.1037//0003-066x.56.3.218

Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M. A., Simango, B., … Taylor, N. G. (2020). Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: Systematic review. *JMIR mHealth and uHealth, 8*(9), Article e18694. https://doi.org/10.2196/18694

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*(1), 72–91. https://doi.org/10.1037/a0032138

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*(3), 271–299. https://doi.org/10.1037/1089-2680.2.3.271

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry, 26*(1), 1–26. https://doi.org/10.1080/1047840X.2014.940781

Haddadj, R., Verdot, C., Salanave, B., Deschamps, V., & Vanhelst, J. (2024). Impact of varying accelerometer epoch length on physical activity patterns in adults: Considerations for public health. *PLoS One, 19*(12). https://doi.org/10.1371/journal.pone.0316176

Haney, A. M., Fleming, M. N., Wycoff, A. M., Griffin, S. A., & Trull, T. J. (2023). Measuring affect in daily life: A multilevel psychometric evaluation of the PANAS-X across four ecological momentary assessment samples. *Psychological Assessment, 35*(6), 469–483. https://doi.org/10.1037/pas0001231

Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2021). The effects of assessment intensity on participant burden, compliance, within-person variance, and within-person relationships in ambulatory assessment. *Behavior Research Methods, 53*, 1875–1894. https://doi.org/10.3758/s13428-021-01592-2

Heron, K. E., Everhart, R. S., McHale, S. M., & Smyth, J. M. (2017). Using mobile-technology-based ecological momentary assessment (EMA) methods with youth: A systematic review and recommendations. *Journal of Pediatric Psychology, 42*(10), 1087–1107. https://doi.org/10.1093/jpepsy/jsx078

Hevel, D. J., Dunton, G. F., & Maher, J. P. (2021). Acute bidirectional relations between affect, physical feeling states, and activity-related behaviors among older adults: An ecological momentary assessment study. *Annals of Behavioral Medicine, 55*(1), 41–54. https://doi.org/10.1093/abm/kaaa027

Hox, J., & McNeish, D. (2020). Small samples in multilevel modeling. In R. Van de Schoot, & M. M. Miočević (Eds.), *Small sample size solutions* (pp. 215–225). Routledge. https://doi.org/10.4324/9780429273872-18.

Jeckel, S., & Sudeck, G. (2017). Physical activity and affective well-being in everyday life. *Zeitschrift für Gesundheitspsychologie, 25*, 19–31. https://doi.org/10.1026/0943-8149/a000163

Jones, P. J., Catt, M., Davies, M. J., Edwardson, C. L., Mirkes, E. M., Khunti, K., … Rowlands, A. V. (2021). Feature selection for unsupervised machine learning of accelerometer data physical activity clusters–A systematic review. *Gait & Posture, 90*, 120–128.

Kanning, M. (2013). Using objective, real-time measures to investigate the effect of actual physical activity on affective states in everyday life: Differentiating the contexts of working and leisure time in a sample with students. *Frontiers in Psychology, 4*, 602. https://doi.org/10.3389/fpsyg.2012.00602

Kanning, M., Ebner-Priemer, U., & Brand, R. (2012). Autonomous regulation mode moderates the effect of actual physical activity on affective states: An ambulant assessment approach to the role of self-determination. *Journal of Sport & Exercise Psychology, 34*(2), 260–269. https://doi.org/10.1123/jsep.34.2.260

Kanning, M., & Hansen, S. (2017). Need satisfaction moderates the association between physical activity and affective states in adults aged 50+: An activity-triggered ambulatory assessment. *Annals of Behavioral Medicine, 51*(1), 18–29. https://doi.org/10.1007/s12160-016-9824-6

Kanning, M. K., Ebner-Priemer, U. W., & Schlicht, W. M. (2013). How to investigate within-subject associations between physical activity and momentary affective states in everyday life: A position statement based on a literature overview. *Frontiers in psychology, 4*, 187. https://doi.org/10.3389/fpsyg.2013.00187

Kerr, J., Marinac, C. R., Ellis, K., Godbole, S., Hipp, A., Glanz, K., Mitchell, J., Laden, F., James, P., & Berrigan, D. (2017). Comparison of accelerometry methods for estimating physical activity. *Medicine & Science in Sports & Exercise, 49*(3), 617–628. https://doi.org/10.1249/MSS.0000000000001124

Keyes, C. L. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior, 43*(2), 207–222. https://doi.org/10.2307/3090197

Kim, J., Conroy, D. E., & Smyth, J. M. (2020). Bidirectional associations of momentary affect with physical activity and sedentary behaviors in working adults. *Annals of Behavioral Medicine, 54*(4), 268–279. https://doi.org/10.1093/abm/kaz045

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin, 139*(4), 917. https://doi.org/10.1037/a0030811

Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. In M. Csikszentmihalyi (Ed.), *Flow and the foundations of positive psychology* (pp. 21–34). Springer. https://doi.org/10.1007/978-94-017-9088-8_2.

Li, Y.-M., Konstabel, K., Mõttus, R., & Lemola, S. (2022). Temporal associations between objectively measured physical activity and depressive symptoms: An experience sampling study. *Frontiers in Psychiatry, 13*, Article 946746. https://doi.org/10.1037/met0000287

Liao, Y., Chou, C.-P., Huh, J., Leventhal, A., & Dunton, G. (2017). Examining acute bi-directional relationships between affect, physical feeling states, and physical activity in free-living situations using electronic ecological momentary assessment. *Journal of Behavioral Medicine, 40*(3), 445–457. https://doi.org/10.1007/s10865-016-9808-9

Liao, Y., Shonkoff, E. T., & Dunton, G. F. (2015). The acute relationships between affect, physical feeling states, and physical activity in daily life: A review of current evidence. *Frontiers in Psychology, 6*, 1975. https://doi.org/10.3389/fpsyg.2015.01975

Liao, Y., Skelton, K., Dunton, G., & Bruening, M. (2016). A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: An adapted STROBE checklist for reporting EMA studies (CREMAS). *Journal of Medical Internet Research, 18*(6), Article e151. https://doi.org/10.2196/jmir.4954

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016). PRESS peer review of electronic search strategies: 2015 guideline statement. *Journal of Clinical Epidemiology, 75*, 40–46. https://doi.org/10.1016/j.jclinepi.2016.01.021

McNeish, D., & Kelley, K. (2019). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods, 24*(1), 20–35. https://doi.org/10.1037/met0000182

Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nyström, C., Mora-Gonzalez, J., Löf, M., Labayen, I., Ruiz, J. R., & Ortega, F. B. (2017). Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. *Sports Medicine, 47*(9), 1821–1845. https://doi.org/10.1007/s40279-017-0716-0

Migueles, J. H., Rowlands, A. V., Huber, F., Sabia, S., & van Hees, V. T. (2019). GGIR: A research community–driven open source R package for generating physical activity and sleep outcomes from multi-day raw accelerometer data. *Journal for the Measurement of Physical Behaviour, 2*(3), 188–196. https://doi.org/10.1123/jmpb.2018-0063

Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C. M., & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI Evidence Synthesis, 18*(10), 2119–2126. https://doi.org/10.11124/JBIES-20-00167

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Gorber, S. C., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity, 5*, 56. https://doi.org/10.1186/1479-5868-5-56

Reed, J., & Buck, S. (2009). The effect of regular aerobic exercise on positive-activated affect: A meta-analysis. *Psychology of Sport and Exercise, 10*(6), 581–594. https://doi.org/10.1016/j.psychsport.2009.04.009

Reed, J., & Ones, D. S. (2006). The effect of acute aerobic exercise on positive activated affect: A meta-analysis. *Psychology of Sport and Exercise, 7*(5), 477–514. https://doi.org/10.1016/j.psychsport.2009.05.009

Reichert, M., Tost, H., Reinhard, I., Schlotz, W., Zipf, A., Salize, H.-J., Meyer-Lindenberg, A., & Ebner-Priemer, U. W. (2017). Exercise versus nonexercise activity: E-diaries unravel distinct effects on mood. *Medicine & Science in Sports & Exercise, 49*(4), 763–773. https://doi.org/10.1249/MSS.0000000000001149

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin, 128*(6), 934–960. https://doi.org/10.1037/0033-2909.128.6.934

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press. https://doi.org/10.1037/0003-066X.55.1.68

Schneider, S., Junghaenel, D. U., Smyth, J. M., Wen, C. K. F., & Stone, A. A. (2024). Just-in-time adaptive ecological momentary assessment (JITA-EMA). *Behavior Research Methods, 56*(2), 765–783. https://doi.org/10.3758/s13428-023-02083-8

Seiferth, C., Fiedler, J., Färber, T., Pape, M., Schroeder, S., Herpertz, S., Steins-Loeber, S., & Wolstein, J. (2024). Bi-directional associations of core affect and physical activity in adults with higher body weight: An ecological momentary assessment study. *Journal of Health Psychology*. https://doi.org/10.1177/13591053241228202. Advance online publication.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Skender, S., Ose, J., Chang-Claude, J., Paskow, M., Brühmann, B., Siegel, E. M., Steindorf, K., & Ulrich, C. M. (2016). Accelerometry and physical activity questionnaires: A systematic review. *BMC Public Health, 16*, 515. https://doi.org/10.1186/s12889-016-3172-0

Stevens, C. J., Baldwin, A. S., Bryan, A. D., Conner, M., Rhodes, R. E., & Williams, D. M. (2020). Affective determinants of physical activity: A conceptual framework and narrative review. *Frontiers in Psychology, 11*, Article 568331. https://doi.org/10.3389/fpsyg.2020.568331

Sudeck, G., Jeckel, S., & Schubert, T. (2018). Individual differences in the competence for physical-activity-related affect regulation moderate the activity–affect association in real-life situations. *Journal of Sport & Exercise Psychology, 40*(4), 196–205. https://doi.org/10.1123/jsep.2018-0017

Timm, I., Giurgiu, M., Ebner-Priemer, U., & Reichert, M. (2024). The within-subject association of physical behavior and affective well-being in everyday life: A systematic literature review. *Sports Medicine, 54*(6), 1667–1705. https://doi.org/10.1007/s40279-024-02016-1

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., & Weeks, L. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. https://doi.org/10.7326/M18-0850

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine & Science in Sports & Exercise, 40*(1), 181–188. https://doi.org/10.1249/MSS.0b013e31815a51b3

Troiano, R. P., McClain, J. J., Brychta, R. J., & Chen, K. Y. (2014). Evolution of accelerometer methods for physical activity research. *British Journal of Sports Medicine, 48*(13), 1019–1023. https://doi.org/10.1136/bjsports-2014-093546

Vancampfort, D., Firth, J., Schuch, F. B., Rosenbaum, S., Mugisha, J., Hallgren, M., Probst, M., Ward, P. B., Gaughran, F., De Hert, M., … Stubbs, B. (2017). Sedentary behavior and physical activity levels in people with schizophrenia, bipolar disorder and major depressive disorder: A global systematic review and meta-analysis. *World Psychiatry, 16*(3), 308–315. https://doi.org/10.1002/wps.20458

Von Haaren, B., Loeffler, S. N., Haertel, S., Anastasopoulou, P., Stumpp, J., Hey, S., & Boes, K. (2013). Characteristics of the activity-affect association in inactive people: An ambulatory assessment study in daily life. *Frontiers in Psychology, 4*, 163. https://doi.org/10.3389/fpsyg.2013.00163

Wilhelm, P., & Schoebi, D. (2007). Assessing mood in daily life: Structural validity, sensitivity to change, and reliability of a short-scale to measure three basic dimensions of mood. *European Journal of Psychological Assessment, 23*(4), 258–267. https://doi.org/10.1027/1015-5759.23.4.258

Wrzus, C., & Neubauer, A. B. (2023). Ecological momentary assessment: A meta-analysis on designs, samples, and compliance across research fields. *Assessment, 30*(3), 825–846. https://doi.org/10.1177/10731911211067538