Comparative Evaluation of Ecological Momentary Assessment, Global Physical Activity Questionnaire, and Bouchard's Physical Activity Record for Measuring Physical Activity: A Multilevel Modeling Approach

Jung Min Noh, Song Hyun Im, JooYong Park, Jaemyung Kim, Mi Young Lee, JiYeob Choi

Submitted to: Journal of Medical Internet Research
on: April 24, 2024

Table of Contents

Comparative Evaluation of Ecological Momentary Assessment, Global Physical Activity Questionnaire, and Bouchard's Physical Activity Record for Measuring Physical Activity: A Multilevel Modeling Approach

Jung Min Noh[1]; Song Hyun Im[2]; JooYong Park[1, 3]; Jaemyung Kim[4]; Mi Young Lee[5]; JiYeob Choi[1]

[1]Seoul National University Seoul KR
[2]Nexon Korea Seoul KR
[3]Eulji University Seongnam KR
[4]Oregon State University Corvallis US
[5]Kookmin University Seoul KR

Corresponding Author:
JiYeob Choi
Seoul National University
103 Daehak-ro, Jongno-gu
Seoul
KR

Abstract

Background: There is growing interest in the real-time assessment of physical activity and physiological variables. Acceleration, particularly those collected through wearable sensors, has been increasingly adopted as an objective measure of physical activity. However, sensor-based measures often pose challenges for large-scale studies due to their associated costs, inability to capture contextual information, and restricted user populations. Smartphone-delivered Ecological Momentary Assessment (EMA) offers an unobtrusive and undemanding means to measure physical activity to address these limitations.

Objective: To evaluate the usability of EMA by comparing its measurement outcomes with two self-report assessments of physical activity: Global Physical Activity Questionnaire (GPAQ) and a modified version of Bouchard's Physical Activity Record (BAR).

Methods: 235 participants (137 females, 98 males, 94 repeated) participated in one or more 7-day study. Waist-worn sensors provided by Actigraph™ captured accelerometer data while participants completed three self-report measures of physical activity. The multilevel modeling method was used with EMA, GPAQ, and BAR as separate measures, with eight sub-domains of physiological activity (overall physical activity; overall excluding occupational; move; moderate and vigorous exercise; moderate and vigorous occupational; sedentary) to model accelerometer data.

Results: Among the three measurement outcomes, EMA (? = .185, p = .005) and BAR (? = .270, p < .001) exhibited higher overall performance over GPAQ (? = .140, p = .019). EMA also showed a more balanced performance, compared to other measurement tools, in modeling various physical activity domains, including occupational, leisure, and sedentary behaviour.

Conclusions: Multilevel modeling on three self-report assessments of physical activity indicates that smartphone-delivered EMA is a valid and efficient method for assessing physical activity.telemedicine; smartphone; wearable electronic devices; physical activity

(JMIR Preprints 24/04/2024:59878)
DOI: https://doi.org/10.2196/preprints.59878

Preprint Settings

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Noh, J M[1]; Im, S H[2;] Park, J[1,3]; Kim, J[4]; Lee, M Y[5]; Choi, J Y[1,6,7].

[1] Department of Biomedical Sciences, Seoul National University, South Korea; [2] Intelligence Laboratories, Nexon Korea, South Korea; [3] Department of Big Data Medical Convergence, Eulji University, South Korea; [4] School of Exercise, Sport, and Health Sciences, Oregon State University, USA; [5] College of Physical Education and Sport Science, Kookmin University, South Korea; [6] Institute of Health Policy and Management, Seoul National University Medical Research Centre, South Korea; [7] Cancer Research Institute, Seoul National University, South Korea

Comparative Evaluation of Ecological Momentary Assessment, Global Physical Activity Questionnaire, and Bouchard's Physical Activity Record for Measuring Physical Activity: A Multilevel Modeling Approach

Abstract

Background: There is growing interest in the real-time assessment of physical activity and physiological variables. Acceleration, particularly those collected through wearable sensors, has been increasingly adopted as an objective measure of physical activity. However, sensor-based measures often pose challenges for large-scale studies due to their associated costs, inability to capture contextual information, and restricted user populations. Smartphone-delivered Ecological Momentary Assessment (EMA) offers an unobtrusive and undemanding means to measure physical activity to address these limitations.

Objective: To evaluate the usability of EMA by comparing its measurement outcomes with two self-report assessments of physical activity: Global Physical Activity Questionnaire (GPAQ) and a modified version of Bouchard's Physical Activity Record (BAR).

Methods: 235 participants (137 females, 98 males, 94 repeated) participated in one or more 7-day study. Waist-worn sensors provided by Actigraph™ captured accelerometer data while participants completed three self-report measures of physical activity. The multilevel modeling method was used with EMA, GPAQ, and BAR as separate measures, with eight sub-domains of physiological activity (overall physical activity; overall excluding occupational; move; moderate and vigorous exercise; moderate and vigorous occupational; sedentary) to model accelerometer data.

Results: Among the three measurement outcomes, EMA ($\beta = .185$, $p = .005$) and BAR ($\beta = .270$, $p < .001$) exhibited higher overall performance over GPAQ ($\beta = .140$, $p = .019$). EMA also showed a more balanced performance, compared to other measurement tools, in modeling various physical activity domains, including occupational, leisure, and sedentary behaviour.

Conclusions: Multilevel modeling on three self-report assessments of physical activity indicates that smartphone-delivered EMA is a valid and efficient method for assessing physical activity.

Keywords: telemedicine; smartphone; wearable electronic devices; physical activity

Introduction

A physically active lifestyle has long been recognised as both a prerequisite for and predictor of maintaining good health. Decades of epidemiologic research have identified the preventive effects of physical activity against various physiological and mental health issues, including heart and other cardiovascular diseases [1-3], depression and suicidal thoughts [4-7], and cancer [2, 8-10].

The recent development of digital and wearable technologies has made it possible to continuously track physical activities in real life through sensors embedded in digital devices. This expansion provides researchers with a broader range of choices, as both research-grade and consumer-grade wearables, with varying costs and capacities to measure health conditions, are now available in the market. While the potential benefits of these devices in research are substantial, it is essential to acknowledge several limitations: Firstly, sensor-based measures pose challenges for large-scale epidemiological studies due to their associated costs and administrative difficulties in managing devices. Secondly, this approach is inadequate for capturing contextual information associated with the activity. Lastly, there are limitations to the populations and circumstances capable of using health-tracking devices, creating potential risks for digital inequality [11-15]. Consequently, the advancement of sensor technologies does not diminish the importance of report-based assessments.

The Global Physical Activity Questionnaire (GPAQ) is an instrument designed to collect self-reports on physical activity in the domains of occupational activity, travel, recreational activities (exercise), and SB [16]. Comprising 16 questions, it is well-recognised for obtaining information on physical activity [17]. However, the retrospective approach adopted by the GPAQ exposes the system to an enhanced risk of memory bias and a lack of temporal specificity associated with the activity [18-19]. This limitation applies similarly to other measures that also rely on retrospective reports.

In contrast, the Bouchard Physical Activity Record (BAR) offers a means for promptly gathering reports on physical activity [20]. It enables respondents to record their physical activity at 15-minute intervals, rating their activity level on a scale from 1 (sedentary behaviour; SB) to 9 (high-intensity exercise). However, BAR is constrained by its reliance on traditional pen-and-paper recording methods and its structure as a log-based system. Consequently, respondents do not encounter customised sets of questions when submitting reports.

Smartphone-delivered Ecological Momentary Assessment (EMA) addresses these limitations by providing real-time and flexible assessment of ongoing activities through interaction with mobile devices. Previous studies have demonstrated the strengths of EMA over retrospective reporting methods in assessing various clinical conditions across populations, including occupational stress in patients with cardiovascular disease [21], anxiety and depression [22-24], and general health-related quality of life [25]. The relative strengths of mobile-based questionnaires also include much easier storage, retrieval, and utilisation of data over time. Yet, there has been inconsistent findings in the agreement between momentary and recall-based assessment methods [26-27], highlighting the need to examine different self-report instruments comprehensively alongside sensor-based measures. For these reasons, we aimed to evaluate EMA methods by comparing the measurement outcomes of EMA with two well-recognised self-report assessment tools to measure physical activity: the GPAQ and a modified version of the BAR.

Methods

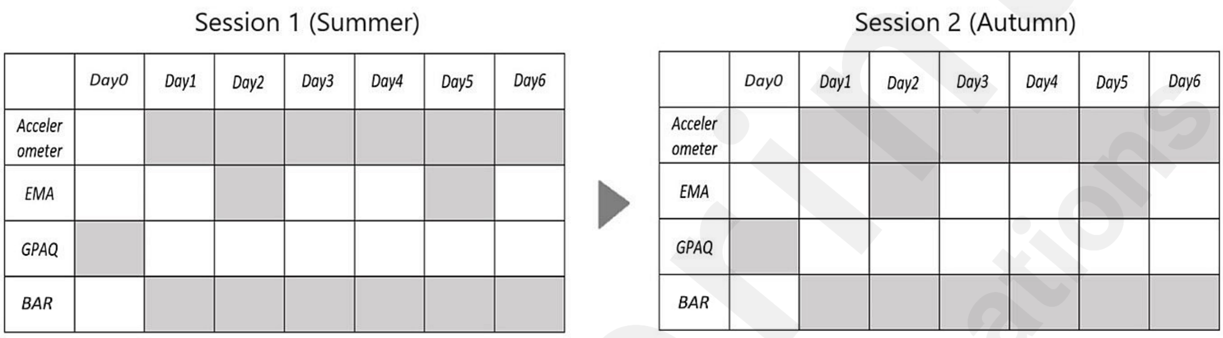
Recruitment

Participants were recruited between May and November 2015 from a public health centre in Seoul, South Korea. Eligibility for study participation included: adults (18 years of age or older) who voluntarily visited a centre for non-orthopaedic or neuromuscular causes; who was capable of undertaking daily physical activities; had no plan for hospitalisation during the study; and who had access to their own mobile phones. 241 participants provided an informed consent to participate in the study.

Design and Procedures

A 7-day study was conducted during the summer and autumn sessions of 2015. Participants were given the option to participate in one or both sessions. On Day 0, participants' body composition and handgrip strength were measured. Participants were asked to wear the ActiGraph GT3X+ (ActiGraph LLC, Pensacola, FL) over the right waist within an anterior axillary line for 7 consecutive days (Days 1-6). They were instructed to always wear the device while awake, except for water activities such as swimming and showering. Three self-report measures of physical activity (i.e. EMA, GPAQ, BAR) were completed according to the following timeline: EMA was completed on one weekday and one weekend day; GPAQ was completed on the first day before participants were provided with the wrist sensor; BAR was completed daily between Days 1 and 6 (Fig 1).

Fig 1. Design of a 7-day study.



Days that participants provided data through each instrument are coloured in grey. Accelerometer data were provided from days 1 to 6. Participants were asked to wear a wrist sensor every day, except during swimming and showering. EMA was performed on one weekday and one weekend day. GPAQ was completed on Day 0, and BAR data were collected between Days 1 and 6.

Materials

Ecological Momentary Assessment (EMA). EMA was generated through Survey Monkey (https://www.surveymonkey.com). The assessment comprised five questions, with the specific questions displayed flexibly based on the respondent's answer to the preceding question (Fig 2). Physical activity questionnaires were sent to participants via SMS messages with a 2-hour interval between 8 am and 8 pm on one weekday and one weekend. Upon receiving an alert, participants reported their primary physical activity during the preceding 30 minutes. Participants categorised their activities as either sedentary (sitting, lying down), transport, occupational (moderate or vigorous), or exercise (light, moderate, or vigorous).

Fig 2. Flowchart of the EMA questionnaire.

Global Physical Activity Questionnaire (GPAQ). GPAQ [16, 28] is a self-report assessment tool that collects data on three physical activity domains: occupational (moderate & vigorous), recreational (moderate & vigorous), and transport, as well as the information on sedentary behaviour. The questionnaire consists of 16 questions that examine the amount of time (in minutes) spent on each activity domain during a typical week. To calculate the overall metabolic equivalents (METs), the time spent on moderate and transportation physical activities was multiplied by 4 METs, while the time spent on vigorous physical activity was multiplied by 8 METs. Sedentary time was assessed by asking about the duration spent sitting or reclining on a typical day. To estimate the average daily time spent on moderate and vigorous physical activity, the time spent on those intensity physical activity was divided by seven. The Korean version of GPAQ, previously validated for native Korean respondents [29], was used in this study.

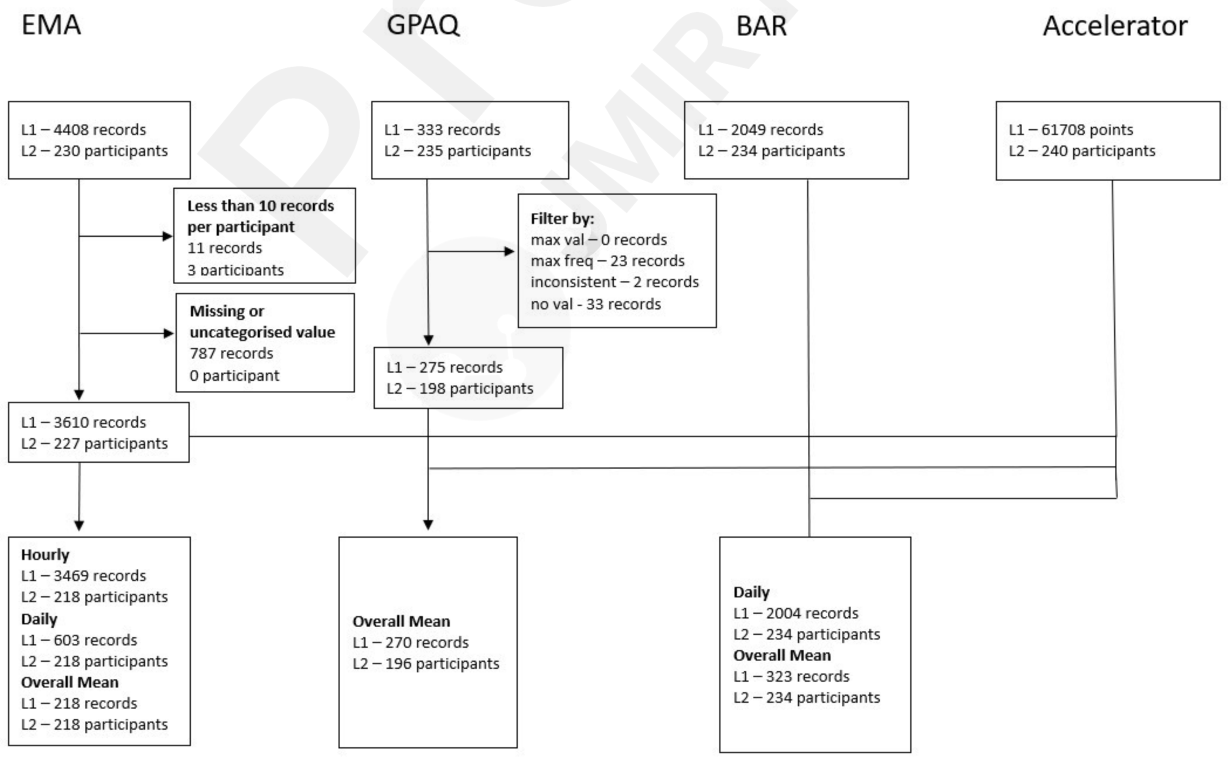
Bouchard's Three-day Physical Activity Record (BAR). BAR [20] is a commonly utilised self-reporting method where participants record their physical activity for each 15-minute interval over a span of three days. Activities are rated on a scale from 1 to 9 (1 indicating sedentary activity and 9 indicating intense work or high-intensity exercise) to generate a total physical activity score. In this study, the questionnaire items in BAR were adapted to gather participants' estimates of total MVPA, moderate and vigorous occupational activity, transport (movement), moderate and vigorous exercise (leisure-related activity), and sedentary behaviour. Participants were asked to complete BAR for 6 consecutive days.

Accelerometer. Accelerometer data was collected using Actigraph GT3X+, a triaxial accelerometer which is a valid method to assess daily free-living physical activity [30]. The device is characterised by its compact size (4.6 cm x 3.3 cm x 1.5 cm) and lightweight design (19 grams). The accelerometer collects the acceleration signal at a 30 Hz sampling rate. The acceleration signal is summed over a 60-second time interval (epoch) and stored as activity counts. In this study, we utilised the estimation of moderate-to-vigorous physical activity (MVPA) time, sedentary time, number of steps, and METs. The physical activity intensity was classified using a Freedson (1998) algorithm [31], MVPA: > 2689 counts per minute (CPM); the sedentary time was defined as < 100 CPM (Troiano 2008). The METs variable was estimated using a Freedson (1998) equation, 1.439008 + (0.000795 x counts/min).

Statistical Analysis

The data cleaning process is illustrated in Fig 3. Of the 4408 EMA records (*n* = 230), 11 records (*n* = 3) were eliminated as participants provided less than 10 reports (i.e. 10/12, less than 83%). Additionally, 787 records were removed due to missing or uncategorised values, resulting in 3610 records (*n* = 227). On the GPAQ, of the 333 records (*n* = 235), 58 records were excluded based on the GPAQ Analysis Guideline [16], leaving 275 records (*n* = 198). EMA data, collected repeatedly within participants every two hours, was analysed in three variations based on time intervals: hourly, daily, and overall mean scores per participant. Similarly, BAR data was examined in two variations: daily and mean scores per participant. Filtered EMA, GPAQ, and BAR data were then separately integrated with accelerometer data, resulting in three datasets each containing 3469 (*n* = 218, 89 repeated), 270 (*n* = 196, 77 repeated), and 2004 (*n* = 234, 94 repeated) records for the final analysis.

Fig 3. Flow chart of data cleansing and integration.



Max val = Maximum value; Max freq = Maximum frequency; No val = No value

To address repeated measures within participants, the multilevel analysis method was employed with step counts and a total MVPA from accelerometer data as the dependent measure. Multilevel modeling is suitable for addressing the hierarchical data structure inherent in the mixed-design study, which includes both within- and between-participant measures [32]. In this study, both BAR and EMA data were structured into three levels. Hourly observations formed the lowest level of the hierarchy (level 1), nested within days (level 2), which were in turn nested within individuals (level 3). Data driven from three self-report instruments (EMA, GPAQ, and BAR), each with eight physical activity domains (total MVPA; total MVPA excluding occupational; movement (transport); moderate and vigorous exercise; moderate and vigorous occupational; sedentary), were separately used to assess the association with objectively obtained accelerometer data (stepscounts and total MVPA). Standardised coefficients were calculated for the comparisons between different instruments. Participants' age and gender were included as covariates to account for differences in outcomes attributed to demographics, and overall and repeated (those who enrolled in both summer and autumn sessions of the study) participant groups were separately examined.

Results

Participant Characteristics

A total of 235 participants (137 females, 98 males; 94 repeated) participated in the study. The mean age of participants was 52.71 (*SD* = 9.07). The characteristics of these participants are shown in Table 1. The majority of participants were married (*n* = 200, 85.11%) at the time of the study and had no smoking experience (*n* = 170, 72.34%). All participants completed at least secondary level education and a larger proportion of participants (*n* = 117, 49.79%) completed a higher level education. Of the participants who enrolled in both summer and autumn sessions of the study (*n* = 94), the mean age was 52.10 (*SD* = 7.50). The majority of participants were married (*n* = 85, 90.43%) females (*n* = 80, 85.11%), who completed a higher level education (*n* = 55, 58.51%) and had no smoking experience (*n* = 81, 86.17%).

Table 1. Characteristics of participants enrolled in the study (May-July, Sept-Nov 2015).

		Overall		Repeat	
		n	%	*n*	%
Total		235		94	
Gender	Female	161	68.51	80	85.11
	Male	74	31.49	14	14.89
Age, years	< 50	86	36.60	39	41.49
	50 - 59	98	41.70	34	36.17
	60 +	51	21.70	21	22.34
Education	<= middle school	24	10.21	4	4.26
	High school	93	39.57	34	36.17
	College +	117	49.79	55	58.51
	Missing	1	0.43	1	1.06
Marriage status	Unmarried	18	7.66	4	4.26
	Married	200	85.11	85	90.43

	Others	17	7.23	5	5.32
Income, KRW/month	< 200	38	16.17	18	19.15
	200 - 400	117	49.79	41	43.62
	400 +	77	32.77	34	36.17
	Missing	3	1.28	1	1.06
Job status	Office	95	40.43	33	35.11
	Manual	64	27.23	19	20.21
	Housewives	54	22.98	30	31.91
	Others	21	8.94	12	12.77
	Missing	1	0.42		
Smoking	Never	170	72.34	81	86.17
	Ex	33	14.04	10	10.64
	Current	27	11.49	1	1.06
	Missing	5	2.13	2	2.13
Alcohol	Never	93	39.57	34	36.17
	Ex	18	7.66	5	5.32
	Current	119	50.64	53	56.38
	Missing	5	2.13	2	2.13
BMI	< 23	96	40.85	39	41.49
	23 - 25	61	25.96	25	26.60
	25 +	66	28.09	22	23.40
	Missing	12	5.11	8	8.51

Average Time Engaged in Various Physical Activity Types

The mean durations of time (minutes per day) spent in each activity type, measured through different instruments (EMA, GPAQ, BAR, accelerometer), are summarised in Table 2. A rank-based comparison of self-report instruments indicated that the time spent on sedentary behaviour (SB) took the largest proportion within each instrument. Similarly, leisure exercise (Exer) took the smallest proportion across different self-report measures of physical activity. On EMA and BAR, transport (Move) time were reported over the time spent on occupational activities, while occupational activities were reported over transport time on GPAQ. Across all self-report methods, participants underreported the time spent on SB ($Mean_{EMA}$ = 251.31, SD_{EMA} = 68.53; $Mean_{GPAQ}$ = 457.27, SD_{GPAQ} = 257.28; $Mean_{BAR}$ = 502.0, SD_{BAR} = 141.58), compared to the SB tracked through wearable sensors (*Mean*= 1048.02, *SD* = 208.77), while overreporting the time spent on total MVPA ($Mean_{ACC}$ = 104.38, SD_{ACC} = 102.38 vs. $Mean_{EMA}$ = 194.41, SD_{EMA} = 68.14; $Mean_{GPAQ}$ = 146.17, SD_{GPAQ} = 155.21; $Mean_{BAR}$ = 146.67, SD_{BAR} = 91.16).

Table 2. Duration (minutes per day) of engaging in different physical activity types by accelerometer, EMA, GPAQ, and BAR.

	Accelerometer		EMA (Daily)		EMA (Overall Mean)		GPAQ		BAR (Daily)		BAR (Overall Mean)	
	Total	*Repeat*	*Total*	*Repeat*	*Total*	*Repeat*	*Total*	*Repeat*	*Total*	*Repeat*	*Total*	*Repeat*
	Mean (SD)	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*
n	235	94	227	89	227	89	196	77	234	94	234	94
SB	1048.02	1033.39	251.02	245.53	251.3	247.2	457.27	477.27	504.69	511.99	502.0	511.26

	(208.77)	(205.46)	(106.85)	(107.09)	1 (68.53)	1 (70.96)	(257.28)	(259.21)	(190.63)	(183.17)	(141.58)	(121.03)
MVPA	-	-	-	-	-	-	-	-	-	-	-	-
Occu	-	-	72.92 (80.44)	87.74 (84.30)	67.79 (57.17)	86.78 (54.90)	62.92 (115.05)	52.67 (73.52)	56.27 (98.59)	61.11 (87.34)	53.75 (68.49)	61.3 (50.14)
Exer	-	-	29.12 (53.04)	23.39 (49.56)	31.66 (40.29)	22.07 (26.43)	32.54 (51.06)	31.36 (49.78)	28.41 (58.1)	26.36 (55.13)	28.48 (41.35)	26.4 (35.31)
Move	-	-	93.15 (91.28)	92.80 (91.48)	94.97 (57.95)	93.40 (48.10)	37.84 (52.74)	42.52 (44.23)	65.31 (73.98)	67.82 (77.96)	64.44 (47.75)	68.06 (43.11)
Total	104.38 (102.38)	99.93 (91.41)	195.19 (107.64)	203.94 (107.47)	194.41 (68.14)	202.25 (69.90)	146.17 (155.21)	124.89 (97.70)	150.0 (130.23)	155.29 (124.42)	146.67 (91.16)	155.76 (78.45)
MET	25.79 (3.87)	25.58 (3.61)	-	-	-	-	-	-	-	-	-	-
Steps	8385.32 (5195.79)	8520.92 (5255.28)	-	-	-	-	-	-	-	-	-	-

SB = sedentary behaviour; MVPA = moderate-to-vigorous physical activity; Occu = occupational activity; Exer = leisure exercise; MET = metabolic equivalent of task.

Comparison between EMA, GPAQ, and BAR

Results of multilevel modeling with the measurement outcomes from three self-report instruments (EMA, GPAQ, and BAR) are shown in Tables 3-4. In general, EMA and BAR exhibited better overall performance in modeling the accelerometer data compared to GPAQ (e.g. EMA mean: β = .185, p = .005; BAR mean: β = .270, p < .001; GPAQ mean: β = .140, p = .019, based on overall participants with stepscounts from accelerometer as dependent variables; Table 3). By contrast, among the specified types of physical activity, GPAQ showed a more balanced performance with the lowest variations between the results of different activity domains (SD = .069), followed by EMA (SD = .133) and BAR (SD = .175; Table 3). Similar results were found with a total MVPA as dependent measures (e.g. EMA mean: β = .214, p < .001; BAR mean: β = .271, p < .001; GPAQ mean: β = .088, p = .175; Table 4).

Table 3. Main outcomes of multilevel modeling with the accelerometer data (stepscounts) as dependent measures.

	Y: Accelerometer (stepscounts)											
	EMA (Hourly)		EMA (Daily)		EMA (Overall mean)		GPAQ		BAR (Daily)		BAR (Overall mean)	
	β	p	β	p	β	p	β	p	β	p	β	p
Total Participants												
Total MVPA	.191	.001	.277	.001	.185	.005	.140	.019	.343	.001	.270	.001
Total exc occu	.168	.001	.194	.001	.099	.140	.170	.003	.451	.001	.372	.001
Mod occu	-.047	.005	-.099	.016	-.023	.755	.065	.250	-.003	.881	.015	.800
Vig occu	.000	.985	-.025	.533	-.151	.031	.028	.658	.054	.024	-.005	.916
Move	.115	.001	.143	.001	.080	.237	.120	.040	.334	.001	.255	.001
Mod exer	.169	.001	.214	.001	.176	.009	.117	.054	.235	.001	.298	.001
Vig exer	.066	.001	.114	.002	.019	.781	.098	.110	.147	.001	.098	.078
Sitting / lying	-.165	.001	-.230	.001	-.161	.016	-.044	.464	-.085	.001	-.121	.033
SD	.125		.175		.133		.069		.188		.175	

Repeated Participants												
Total MVPA	.251	.001	.413	.001	.221	.061	.162	.021	.407	.001	.277	.004
Total exc occu	.203	.001	.311	.001	.211	.088	.149	.048	.509	.001	.347	.001
Mod occu	-.040	.102	-.111	.059	.077	.561	.108	.121	.002	.948	.040	.691
Vig occu	-.003	.904	-.010	.874	-.089	.639	-.034	.576	.002	.933	-.195	.021
Move	.151	.001	.164	.004	.032	.772	.194	.001	.380	.001	.285	.001
Mod exer	.213	.001	.315	.001	.251	.030	.011	.872	.248	.001	.256	.021
Vig exer	.053	.028	.107	.052	.100	.414	.025	.706	.175	.001	.068	.495
Sitting / lying	-.218	.001	-.363	.001	-.245	.045	-.034	.664	-.145	.001	-.218	.018
SD	.159		.257		.170		.092		.230		.221	

Total MVPA = Total moderate-to-vigorous physical activity; Total exc occu = Total physical activity excluding occupational; Mod occu = Moderate occupational activity; Vig occu = Vigorous occupational activity; Mod exer = Moderate leisure exercise; Vig exer = Vigorous leisure exercise; Sitting / lying = Sitting or lying down (sedentary behaviour).

Table 4. Main outcomes of multilevel modeling with the accelerometer data (totalmvpa) as dependent measures.

	Y: Accelerometer (totalmvpa)											
	EMA (Hourly)		EMA (Daily)		EMA (Overall mean)		GPAQ		BAR (Daily)		BAR (Overall mean)	
	β	*p*	*β*	*p*	*β*	*p*	*β*	*p*	*β*	*p*	*β*	*p*
Total Participants												
Total MVPA	.166	.001	.214	.001	.136	.057	.088	.175	.271	.001	.201	.001
Total exc occu	.156	.001	.161	.001	.094	.219	.120	.047	.376	.001	.400	.001
Mod occu	-.036	.034	-.090	.019	-.135	.073	.008	.881	-.028	.162	-.124	.084
Vig occu	-.003	.853	-.052	.160	-.149	.045	.028	.667	.064	.005	.005	.936
Move	.073	.001	.066	.059	-.027	.710	.064	.245	.234	.001	.278	.001
Mod exer	.162	.001	.213	.001	.276	.001	.055	.254	.233	.001	.314	.001
Vig exer	.077	.001	.134	.001	-.002	.980	.073	.165	.177	.001	.125	.083
Sitting / lying	-.140	.001	-.169	.001	-.106	.144	.015	.925	-.032	.151	.055	.504
SD	.110		.147		.149		.038		.147		.175	
Repeated Participants												
Total MVPA	.212	.001	.380	.001	.159	.174	.130	.072	.359	.001	.287	.001
Total exc occu	.193	.001	.309	.001	.166	.166	.139	.076	.477	.001	.406	.001
Mod occu	-.026	.300	-.078	.200	.072	.580	.089	.212	-.029	.320	-.035	.725
Vig occu	-.006	.815	-.012	.843	-.176	.186	-.037	.570	-.004	.883	-.166	.049
Move	.092	.001	.085	.119	-.059	.652	.175	.010	.302	.001	.307	.001
Mod exer	.216	.001	.315	.001	.308	.006	-.004	.935	.284	.001	.308	.002
Vig exer	.063	.010	.146	.008	.126	.284	.046	.503	.243	.001	.113	.254
Sitting / lying	-.183	.001	-.316	.001	-.176	.147	.007	.929	-.098	.002	-.202	.012

SD	.140	.235	.174	.077	.208	.236

Total MVPA = Total moderate-to-vigorous physical activity; Total exc occu = Total physical activity excluding occupational; Mod occu = Moderate occupational activity; Vig occu = Vigorous occupational activity; Mod exer = Moderate leisure exercise; Vig exer = Vigorous leisure exercise; Sitting / lying = Sitting or lying down (sedentary behaviour).

Given the capacity to capture repeated observations within each participant, EMA and BAR were considered with varying timeframes. It is noted that on both EMA and BAR, the daily-based measures exhibited much better performance than the overall mean per participants (e.g. EMA daily: $\beta = .277$, $p < .001$ vs. EMA mean: $\beta = .185$, $p = .005$; BAR daily: $\beta = .343$, $p < .001$ vs. BAR mean: $\beta = .270$, $p < .001$; Table 3). Interestingly, the performance outcomes of the EMA based on hourly timeframes (EMA hourly: $\beta = .191$, $p < .001$) were not as strong as the EMA based on daily measures. Among the overall and repeated participant groups, data driven from the repeated participant group usually showed higher performance than the data from the overall participant group (e.g. EMA overall: $\beta = .277$, $p < .001$ vs. EMA repeated: $\beta = .413$, $p < .001$, Table 3).

A rank-based comparison of different types of physical activity indicated that moderate physical activities, encompassing moderate occupational activity and moderate exercise, were generally better than vigorous activities, including vigorous occupational activity and vigorous exercise, in modeling the accelerometer data (e.g. GPAQ moderate exercise: $\beta = .117$, $p = .054$ vs. GPAQ vigorous exercise: $\beta = .098$, $p = .110$, Table 3). The pattern was found consistently across different self-report instruments.

Discussion

Principle Results

We conducted a comprehensive comparative evaluation of three self-report assessment tools for physical activity by using accelerometer objectively measured through waist-worn sensors. In comparison to widely recognised instruments, the GPAQ and a modified version of BAR, smartphone-delivered EMA method demonstrated strengths across various metrics of physical activity. Notably, compared to BAR, which exhibited strong performance in total MVPA but showed the largest discrepancy in performance across specific activity measures, or GPAQ, which demonstrated the highest consistency across different activity types but the lowest overall performance, EMA consistently performed well across sedentary, occupational, and leisure-related activities, as well as total MVPA.

Interpretation of Findings

We observed consistent patterns in participants' daily activity reports and their associations with accelerometer data collected via waist sensors. Generally, participants tended to underestimate time spent in SB compared to the SB tracked by wearable sensors, while overestimating total MVPA time. This aligns with previous research findings that reported overestimation of activity levels through self-reported measures over sensor-derived outcomes [33-35].

Moderate levels of physical activity, including moderate occupational activity and moderate leisure exercise, generally showed better modeling performance compared to vigorous activities (vigorous occupational; vigorous exercise) across different self-report instruments. While several studies have reported the relative weakness of commercialised sensors in detecting high-intensity compared to low-intensity exercises [36], the findings in this research may also be attributed to participant demographics, with a large proportion of females above 40 years of age and a low average amount of time participants spent on overall exercise (M = 30.89, SD = 2.14).

Comparisons among the three self-report instruments for modeling accelerometer data suggested that assessments with smaller time intervals yield stronger outcomes. This was evidenced by the superior performance of BAR and EMA over GPAQ, as well as the stronger performance of daily-based measures on both BAR and EMA compared to the overall mean per participant. Notably, the most robust performance with EMA data was observed in datasets parsed on a daily basis rather than hourly intervals, indicating that increased temporal specificity in self-report measures enhances the validity of the report, albeit to a certain extent. Furthermore, it is imperative to consider that the frequency of reporting places greater time and attentional demands on respondents. Future studies should aim to identify temporal intervals that achieve the optimal balance between respondent demands and measurement performance [37].

Limitations

The recruitment for this study was conducted at a single public health centre, where participants volunteered to take part in the study. While employing this recruitment strategy was necessary to validate the Physical Activity Questionnaire (PAQ) [38] used in previous studies, a significant proportion of the participants were married females aged 50 or older, indicating a potential skew in the overall participant demographic. This suggests that the tendencies observed in participants' engagement in different types of activity may not be fully generalisable [33, 39]. Additionally, due to technological limitations, we did not prompt participants to report their physical activity through EMA at random intervals throughout the day. As a result, there is a possibility that the data may be skewed towards periods of more representative physical activity. Future studies should consider adopting methodologies that allow for more flexible timing of EMA prompts to address this limitation.

Conclusions

Overall, our comparative modeling underscores smartphone-delivered EMA as a valid and consistent method for assessing physical activity across various activity domains. While there has been increasing interest in utilising accelerometers for passive physical activity assessment, challenges associated with sensor-based measures, including their limited suitability for large-scale studies and constrained user populations, highlight the continued necessity of employing self-report instruments for physical activity measurement. Smartphone-delivered EMA provides an additional advantage over other self-report tools by enabling real-time capture of physical activities, while allowing the storage, retrieval, and utilisation of data over time more effective. In summary, the robustness of EMA in capturing diverse physical activities, coupled with the unobtrusive nature of its data collection methodology, addresses the limitations of both traditional and emerging assessment methods, positioning EMA as a promising tool for researchers and practitioners seeking accurate and reliable assessments of physical activity.

Acknowledgements

This study was supported by the National Research Foundation of Korea Grant funded by the Korean government (NRF-2018R1A2A3075397, NRF2022R1A2B5B01002471).

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Franklin, B. A., Eijsvogels, T. M. H., Pandey, A., Quindry, J., & Toth, P. P. (2022). Physical

activity, cardiorespiratory fitness, and cardiovascular health: A clinical practice statement of the American Society for Preventive Cardiology Part II: Physical activity, cardiorespiratory fitness, minimum and goal intensities for exercise training, prescriptive methods, and special patient populations. American journal of preventive cardiology, 12, 100425. https://doi.org/10.1016/j.ajpc.2022.100425

2. Garcia, L., Pearce, M., Abbas, A., Mok, A., Strain, T., Ali, S., Crippa, A., Dempsey, P. C., Golubic, R., Kelly, P., Laird, Y., McNamara, E., Moore, S., de Sa, T. H., Smith, A. D., Wijndaele, K., Woodcock, J., & Brage, S. (2023). Non-occupational physical activity and risk of cardiovascular disease, cancer and mortality outcomes: a dose-response meta-analysis of large prospective studies. British journal of sports medicine, 57(15), 979–989. https://doi.org/10.1136/bjsports-2022-105669
3. Kraus, W. E., Powell, K. E., Haskell, W. L., Janz, K. F., Campbell, W. W., Jakicic, J. M., Troiano, R. P., Sprow, K., Torres, A., Piercy, K. L., & 2018 PHYSICAL ACTIVITY GUIDELINES ADVISORY COMMITTEE* (2019). Physical Activity, All-Cause and Cardiovascular Mortality, and Cardiovascular Disease. Medicine and science in sports and exercise, 51(6), 1270–1281. https://doi.org/10.1249/MSS.0000000000001939
4. Craft, L. L., & Perna, F. M. (2004). The Benefits of Exercise for the Clinically Depressed. Primary care companion to the Journal of clinical psychiatry, 6(3), 104–111. https://doi.org/10.4088/pcc.v06n0301
5. Koo, K. M., & Kim, K. (2020). Effects of Physical Activity on the Stress and Suicidal Ideation in Korean Adult Women with Depressive Disorder. International journal of environmental research and public health, 17(10), 3502. https://doi.org/10.3390/ijerph17103502
6. Koo, K., & Kim, K. (2021). Effects of Different Types of Physical Activity on Health-Related Quality-of-Life in Korean Women with Depressive Disorder. International journal of environmental research and public health, 18(9), 4639. https://doi.org/10.3390/ijerph18094639
7. Pearce, M., Garcia, L., Abbas, A., Strain, T., Schuch, F. B., Golubic, R., Kelly, P., Khan, S., Utukuri, M., Laird, Y., Mok, A., Smith, A., Tainio, M., Brage, S., & Woodcock, J. (2022). Association Between Physical Activity and Risk of Depression: A Systematic Review and Meta-analysis. JAMA psychiatry, 79(6), 550–559. https://doi.org/10.1001/jamapsychiatry.2022.0609
8. Choi, J., Park, J. Y., Kim, J. E., Lee, M., Lee, K., Lee, J. K., Kang, D., Shin, A., & Choi, J. Y. (2023). Impact of pre- and post-diagnosis physical activity on the mortality of patients with cancer: Results from the Health Examinees-G study in Korea. Cancer medicine, 12(15), 16591–16603. https://doi.org/10.1002/cam4.6253
9. Choi, J., Park, J., Kim, J. E., Lee, M., Kang, D., Shin, A., & Choi, J. Y. (2023). Associations between Physical Activity and Incidence of Cancer among Overweight Adults in Korea: Results from the Health Examinees-G Study. Cancer prevention research (Philadelphia, Pa.), 16(7), 405–418. https://doi.org/10.1158/1940-6207.CAPR-22-0466
10. McTiernan, A., Friedenreich, C. M., Katzmarzyk, P. T., Powell, K. E., Macko, R., Buchner, D., Pescatello, L. S., Bloodgood, B., Tennant, B., Vaux-Bjerke, A., George, S. M., Troiano, R. P., Piercy, K. L., & 2018 PHYSICAL ACTIVITY GUIDELINES ADVISORY COMMITTEE* (2019). Physical Activity in Cancer Prevention and Survival: A Systematic Review. Medicine and science in sports and exercise, 51(6), 1252–1261. https://doi.org/10.1249/MSS.0000000000001937
11. Ali, A., & Clarke, D. F. (2023). Digital measures in epilepsy in low-resourced environments. Expert review of pharmacoeconomics & outcomes research, 1–8. Advance online publication. https://doi.org/10.1080/14737167.2023.2270163
12. Cho, P. J., Yi, J., Ho, E., Shandhi, M. M. H., Dinh, Y., Patil, A., Martin, L., Singh, G., Bent, B., Ginsburg, G., Smuck, M., Woods, C., Shaw, R., & Dunn, J. (2022). Demographic Imbalances Resulting From the Bring-Your-Own-Device Study Design. JMIR mHealth and uHealth, 10(4), e29510. https://doi.org/10.2196/29510

13. Hadjiat Y. (2023). Healthcare inequity and digital health-A bridge for the divide, or further erosion of the chasm?. PLOS digital health, 2(6), e0000268. https://doi.org/10.1371/journal.pdig.0000268
14. Yao, R., Zhang, W., Evans, R., Cao, G., Rui, T., & Shen, L. (2022). Inequities in Health Care Services Caused by the Adoption of Digital Health Technologies: Scoping Review. Journal of medical Internet research, 24(3), e34144. https://doi.org/10.2196/34144
15. Zinzuwadia, A. & Singh, J. P. (2022). Wearable devices – addressing bias and inequity. The Lancet Digital Health, 12(4), e856-857. https://doi.org/10.1016/S2589-7500(22)00194-7
16. World Health Organization (WHO) (2010). Global Physical Activity Questionnaire (GPAQ) Analysis Guide. World Health Organization.
17. Bull, F. C., Maslin, T. S., & Armstrong, T. (2009). Global physical activity questionnaire (GPAQ): nine country reliability and validity study. Journal of physical activity & health, 6(6), 790–804. https://doi.org/10.1123/jpah.6.6.790
18. Lourenco, M. P. C. G., Simoes, J., Vlaeyen, J. W. S., & Cima, R. F. F. (2022). The Daily Experience of Subjective Tinnitus: Ecological Momentary Assessment Versus End-of-Day Diary. Ear and hearing, 43(1), 45–52. https://doi.org/10.1097/AUD.0000000000001054
19. 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
20. Bouchard, C., Tremblay, A., Leblanc, C., Lortie, G., Savard, R., & Thériault, G. (1983). A method to assess energy expenditure in children and adults. The American journal of clinical nutrition, 37(3), 461–467. https://doi.org/10.1093/ajcn/37.3.461
21. Kamarck, T. W., Muldoon, M. F., Shiffman, S. S., & Sutton-Tyrrell, K. (2007). Experiences of demand and control during daily life are predictors of carotid atherosclerotic progression among healthy men. Health psychology : official journal of the Division of Health Psychology, American Psychological Association, 26(3), 324–332. https://doi.org/10.1037/0278-6133.26.3.324
22. Anestis, M. D., Selby, E. A., Crosby, R. D., Wonderlich, S. A., Engel, S. G., & Joiner, T. E. (2010). A comparison of retrospective self-report versus ecological momentary assessment measures of affective lability in the examination of its relationship with bulimic symptomatology. *Behaviour research and therapy*, *48*(7), 607–613. https://doi.org/10.1016/j.brat.2010.03.012
23. Christensen, K., Victor, S. E., Littlefield, A. K., & Mitchell, S. M. (2023). A comparison of retrospectively reported and ecological momentary assessment-reported perceived social support in predicting ecological momentary assessment-reported non-suicidal self-injury. *Suicide & life-threatening behavior*, 10.1111/sltb.13031. Advance online publication. https://doi.org/10.1111/sltb.13031
24. Moore, R. C., Depp, C. A., Wetherell, J. L., & Lenze, E. J. (2016). Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. Journal of psychiatric research, 75, 116–123. https://doi.org/10.1016/j.jpsychires.2016.01.011
25. Oreel, T. H., Delespaul, P., Hartog, I. D., Henriques, J. P. S., Netjes, J. E., Vonk, A. B. A., Lemkes, J., Scherer-Rath, M., van Laarhoven, H. W. M., Sprangers, M. A. G., & Nieuwkerk, P. T. (2020). Ecological momentary assessment versus retrospective assessment for measuring change in health-related quality of life following cardiac intervention. Journal of patient-reported outcomes, 4(1), 98. https://doi.org/10.1186/s41687-020-00261-2
26. Solhan, M. B., Trull, T. J., Jahng, S., & Wood, P. K. (2009). Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. Psychological assessment, 21(3), 425–436. https://doi.org/10.1037/a0016869
27. Edmondson, D., Shaffer, J. A., Chaplin, W. F., Burg, M. M., Stone, A. A., & Schwartz, J. E. (2013). Trait anxiety and trait anger measured by ecological momentary assessment and their

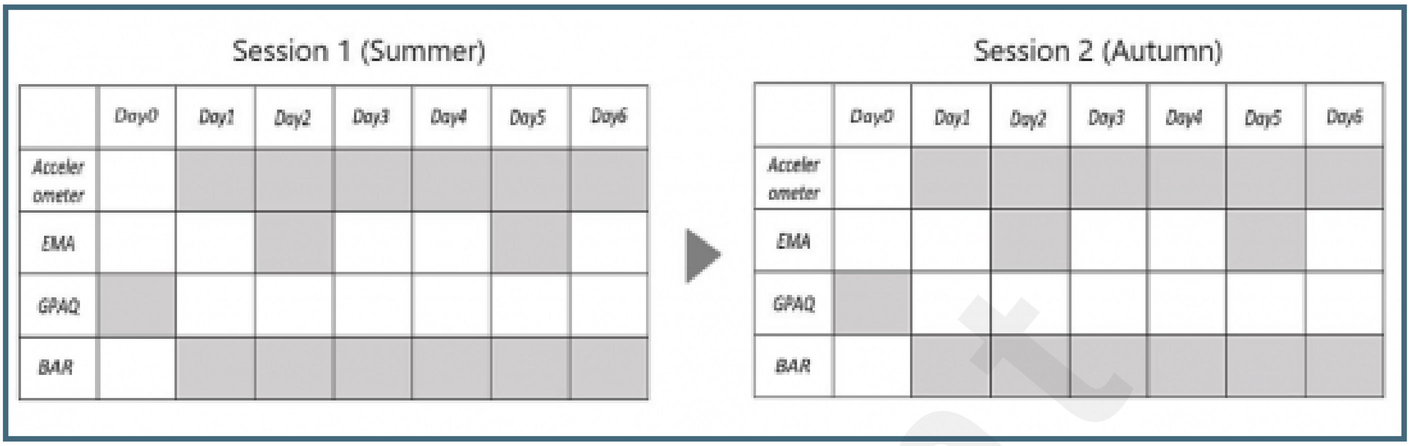
correspondence with traditional trait questionnaires. *Journal of research in personality*, *47*(6), 10.1016/j.jrp.2013.08.005. https://doi.org/10.1016/j.jrp.2013.08.005

28. Armstrong, T. & Bull, F. (2006). Development of the World Health Organization Global Physical Activity Questionnaire (GPAQ). *Journal of Public Health,* 14, 66–70. https://doi.org/10.1007/s10389-006-0024-x
29. Justin Y. Jeon. (2013). Development of the Korean Version of Global Physical Activity Questionnaire and Assessment of Reliability and Validity (RP00003336). Korea Disease Control and Prevention Agency.
30. Aadland, E., & Ylvisåker, E. (2015). Reliability of the Actigraph GT3X+ Accelerometer in Adults under Free-Living Conditions. PloS one, 10(8), e0134606. https://doi.org/10.1371/journal.pone.0134606
31. Freedson, P. S., Melanson, E., & Sirard, J. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. Medicine and science in sports and exercise, 30(5), 777–781. https://doi.org/10.1097/00005768-199805000-00021
32. Kim, J., Marcusson-Clavertz, D., Togo, F., & Park, H. (2018). A Practical Guide to Analyzing Time-Varying Associations between Physical Activity and Affect Using Multilevel Modeling. Computational and mathematical methods in medicine, 2018, 8652034. https://doi.org/10.1155/2018/8652034
33. Dyrstad, S. M., Hansen, B. H., Holme, I. M., & Anderssen, S. A. (2014). Comparison of self-reported versus accelerometer-measured physical activity. Medicine and science in sports and exercise, 46(1), 99–106. https://doi.org/10.1249/MSS.0b013e3182a0595f
34. Sanda, B., Vistad, I., Haakstad, L. A. H., Berntsen, S., Sagedal, L. R., Lohne-Seiler, H., & Torstveit, M. K. (2017). Reliability and concurrent validity of the International Physical Activity Questionnaire short form among pregnant women. BMC sports science, medicine & rehabilitation, 9, 7. https://doi.org/10.1186/s13102-017-0070-4
35. Van Holle, V., De Bourdeaudhuij, I., Deforche, B., Van Cauwenberg, J., & Van Dyck, D. (2015). Assessment of physical activity in older Belgian adults: validity and reliability of an adapted interview version of the long International Physical Activity Questionnaire (IPAQ-L). BMC public health, 15, 433. https://doi.org/10.1186/s12889-015-1785-3
36. Szymczak, H., Keller, L., Debbeler, L. J., Kollmann, J., Lages, N. C., Gollwitzer, P. M., Schupp, H. T., & Renner, B. (2020). An Increase in Vigorous but Not Moderate Physical Activity Makes People Feel They Have Changed Their Behavior. Frontiers in psychology, 11, 1530. https://doi.org/10.3389/fpsyg.2020.01530
37. Degroote, L., DeSmet, A., De Bourdeaudhuij, I., Van Dyck, D., & Crombez, G. (2020). Content validity and methodological considerations in ecological momentary assessment studies on physical activity and sedentary behaviour: a systematic review. The international journal of behavioral nutrition and physical activity, 17(1), 35. https://doi.org/10.1186/s12966-020-00932-9
38. Health Examinees Study Group (2015). The Health Examinees (HEXA) study: rationale, study design and baseline characteristics. Asian Pacific journal of cancer prevention : APJCP, 16(4), 1591–1597. https://doi.org/10.7314/apjcp.2015.16.4.1591
39. Brown, W. J., & Bauman, A. E. (2000). Comparison of estimates of population levels of physical activity using two measures. Australian and New Zealand journal of public health, 24(5), 520–525. https://doi.org/10.1111/j.1467-842x.2000.tb00503.x

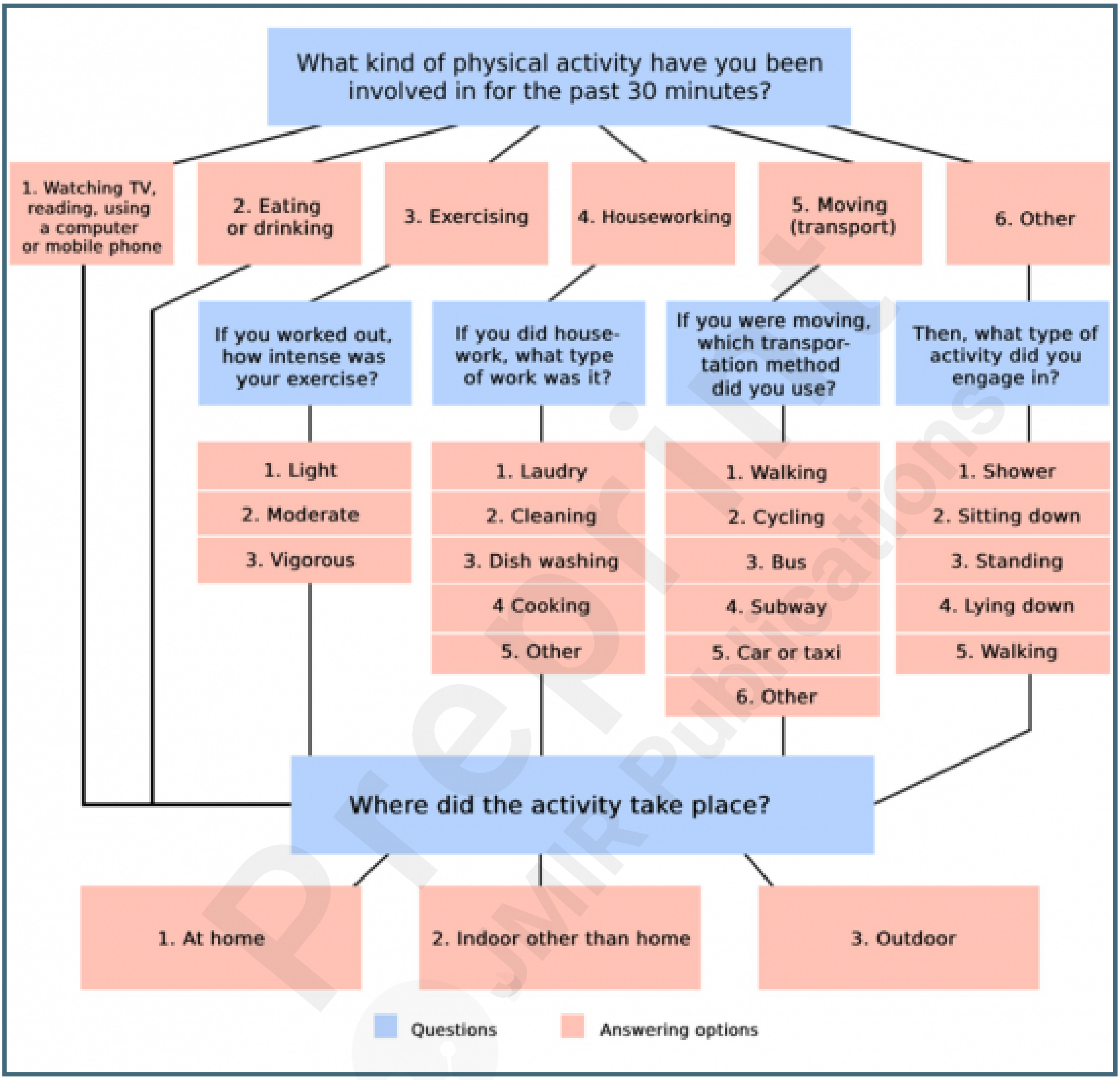
Supplementary Files

Figures

Design of a 7-day study.



Flowchart of the EMA questionnaire.



Flow chart of data cleansing and integration.