

## **Beyond the (log)book: Comparing accelerometer non-wear detection techniques in toddlers**

Elyse Letts<sup>1</sup>, Sarah M da Silva<sup>1</sup>, Natascja Di Cristofaro<sup>1</sup>, Sara King-Dowling<sup>2†</sup>, Joyce Obeid<sup>1,3†\*</sup>

1. Child Health & Exercise Medicine Program, Department of Pediatrics, McMaster University, Hamilton, Canada

2. Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, USA

3. Department of Kinesiology, McMaster University, Hamilton, Canada

† Co-senior authors

\*Corresponding author: [obeidj@mcmaster.ca](mailto:obeidj@mcmaster.ca)

### **ORCIDs**

Elyse Letts: <https://orcid.org/0000-0002-2543-8346>

Sarah M da Silva: <https://orcid.org/0009-0001-9512-9776>

Sara King-Dowling: <https://orcid.org/0000-0003-3670-8425>

Joyce Obeid: <https://orcid.org/0000-0002-6209-1505>

**Data availability statement.** Data may be available upon request to the corresponding author.

**Funding statement.** This work was funded by the Canadian Institutes of Health Research.

**Conflict of interest disclosure.** The authors report no conflicts of interest.

**Ethics approval statement.** This study received ethics clearance from the Hamilton integrated Research Ethics Board (HiREB #3674).

**Patient consent statement.** Informed written consent was obtained from a parent of each participant.

**Acknowledgements.** We would like to thank the participants and their families for their time and effort, as well as the funders who made this research possible. We also acknowledge the efforts of Ms. Camilla Wegrzynowska who contributed to the manual cleaning of accelerometer data.

## Abstract

**Background:** Accelerometers are increasingly used to measure physical activity and sedentary time in toddlers. Data cleaning or wear time validation can impact outcomes of interest, particularly in young children who spend less time awake. However, no study has systematically compared wear time validation strategies in toddlers. As such, the objective of this study is to compare different fully-automated methods of distinguishing wear and non-wear time (counts and raw data algorithms) to the semi-automated (counts with logbooks) criterion method in toddlers. **Methods:** We recruited 109 toddlers (age 12-35 mos) as part of the iPLAY study to wear an ActiGraph w-GT3X-BT accelerometer on the right hip for ~7 consecutive days (removed for sleep and water activities). Parents completed a logbook to indicate monitor removal and nap times. We tested 15 nonwear detection methods grouped into 4 main categories: semi-automated logbook, consecutive 0 counts, modified consecutive 0 counts (Troiano and Choi), and raw data methods (van Hees and Ahmadi). Using logbooks as the criterion standard (all wear and wake-time only wear), we calculated the accuracy and F1 scores and compared overall wear time with a two one-sided test of equivalence. **Results:** Participant daily wear time ranged from 556 to 684 minutes/day depending on method. Accuracy and F1 score ranged from 86 to 95%. Five methods were considered equivalent to the AllWear nonwear criterion (true wear time including sleep-time wear), with only one equivalent to the AwakeWear criterion. Mean absolute differences were lower for the AllWear criterion but ranged 49 to 192 minutes/day. **Conclusions:** The 5min0count, 10min\_0count, 30min\_0count, Troiano60s, and Ahmadi methods provide high accuracy and equivalency when compared to semi-automated cleaning using logbooks. This paper provides insights and quantitative results that can help researchers decide which method may be most appropriate given their population of interest, sample size, and study protocol.

**Keywords:** toddler, early years, accelerometer, actigraphy, nonwear, raw data

## Introduction

Accelerometers have been increasingly used to measure physical activity and sedentary time in toddlers.<sup>1-3</sup> A key element of analysis is wear time validation, which encompasses determining accelerometer wear and nonwear periods (e.g., removed for water activities, discomfort, etc.<sup>4</sup>) and differentiating wake-time vs. sleep-time wear.<sup>5</sup> Existing evidence from school-age children and adults suggests that nonwear detection methods can significantly impact variables of interest,<sup>5-7</sup> however, their impact in toddlers remains unclear.<sup>3,8</sup>

In school-age children, wear time may vary 17%, or up to 87 minutes per day, depending on method<sup>6</sup> which could result in large overestimations of sedentary time.<sup>4</sup> This is mirrored in adult studies, where nonwear methods can change daily wear time by up to 150 minutes.<sup>9</sup> Further, in young children who nap regularly, determining nap wear time, so as not to overestimate sedentary time, is of greater concern.<sup>10,11</sup> Given these large differences, it is important to clearly report the wear time validation methods used in toddlers.<sup>5,6</sup>

A recent review<sup>8</sup> of accelerometry in the early years found that 1/3 of the 627 papers did not report nonwear detection methods, suggesting comprehensive/transparent reporting of methods remains a limitation. Of studies that did report methods, 20 minutes of consecutive 0 counts was most common.<sup>8</sup> Other methods included logbooks (participant/proxy self-report), 1-120 minutes of consecutive 0 counts, and raw data variability.<sup>8</sup>

To date, a generally accepted criterion measure for determining nonwear time is logbooks. These logs allow the participant and/or caregiver proxy to note when the accelerometer was worn/removed each day. Despite being used as a criterion, they face inherent bias because they are self-reported. They also pose a significant participant burden and may not be completed accurately, leaving researchers to make a “best guess” based on observed count patterns. To

address these limitations, logbooks are often used in conjunction with automated methods where data are manually checked against the logbooks to confirm the on/off times. Fully-automated methods of nonwear detection are growing, especially in large population-based studies as they can be used independently of logbooks, therefore reducing analysis burden.

One automated algorithm for nonwear detection is consecutive 0 counts. The basis of this method is that it is unlikely that an individual will remain completely motionless (0 counts) for a given amount of time (e.g., 60 minutes).<sup>5</sup> Variations of this method change the total 0-count time (e.g., 20 minutes for younger children who are less likely to remain motionless) and/or allow for short-duration spikes of activity within the timeframe. Two further extensions of this method are integrated within ActiLife software (Troiano<sup>12</sup> and Choi<sup>13</sup>), making them accessible and thus, commonly<sup>7</sup> used. Finally, another growing category of automatic nonwear methods uses raw accelerations.<sup>14</sup> While these methods may face challenges given the file size, data volume, and technical ability required to process the data, they have been consistently growing in use, but have limited testing in toddlers.

As outlined, there are a wide range of methods to determine nonwear time, all of which have been tested in adults<sup>5,7</sup> and school-age children.<sup>4,6</sup> However, to our knowledge, no study to date has compared nonwear detection methods in toddlers while accounting for potential nap-based wear,<sup>15</sup> leaving researchers to use methods developed for older children or adapt adult methods. As such, this paper aims to answer the following question: What is the accuracy of existing automated accelerometer non-wear detection methods when compared to semi-automated logbook-based cleaning in toddlers?

## **Methods**

### *Participants*

We recruited healthy toddlers aged 12.0-35.9 months with no known physical disability, motor delay, or medical condition that affected movement, as part of the investigating the validity and reliability of accelerometer-based measures of Physical Activity and sedentary time in toddlers (iPLAY) study. Participants were recruited through local childcare centres, websites/social media, and flyers throughout Hamilton, Ontario, Canada. A parent of each participant provided written informed consent. This study received ethics approval from the Hamilton Integrated Research Ethics Board (HiREB #3674).

### *Study procedures*

Participants were fitted with an ActiGraph wGT3X-BT (dynamic range  $\pm 8$  g) by a member of the research team. The accelerometer was worn on the right hip (elastic belt) for approximately 7 consecutive days and removed only for night-time sleep and water activities. Parents were given the option to remove or keep the accelerometer on during daytime naps. All on/off times and nap times were recorded in a logbook (supplementary material 1 has a sample logbook). Accelerometers were initialized in ActiLife and recorded tri-axially at 30Hz with idle sleep mode activated. Upon accelerometer return, the raw .gt3x files, 1-second, and 60-second epoch AGD files were downloaded.

### *Nonwear methods*

We tested 15 different nonwear methods, broadly grouped into 4 categories: logbooks (criterion), consecutive 0 counts, modified consecutive 0 counts, and raw data algorithms. Table 1 provides further details on each method tested.

Logbook method (criterion): The logbook method is a semi-automated protocol in which AGD files are compared to written logbooks. Using ActiLife software and 1-second AGD files, we identified potential nonwear using vector magnitude counts with a minimum nonwear length of

10 minutes, 0 spike tolerance, and ignored wear periods under 1 minute. The resulting wear/nonwear periods were then manually compared to logbooks. All nonwear periods noted in the logbook were removed using a 15-minute error window (e.g., participant notes device removal at 10:00 am, acceptable off-times could be from 9:45 to 10:15 am). If a nonwear period noted in the logbook was not detected by the software, a log diary was added to exclude it (e.g., participant removes the accelerometer for water activity, but a parent holds it, the software detects wear). If any on/off times were blank in the logbook, we used the time provided by ActiLife provided that it (1) was similar to other days in that week and (2) had a clear distinction between wear/nonwear. If the time provided by ActiLife did not meet these criteria, we used the first/last reported time (e.g., on time after nap) as the on/off time, respectively. Because parents were given the option to wear or remove the device during naps, we have elected to use two versions of logbook-based cleaning as criterion measures: 1) follows the logbook and leaves nap times in if accelerometer was worn during the nap (referred to as Logbook\_AllWear hereafter); and 2) removes all nap times and only captures the time when the child is awake and wearing the accelerometer (referred to as Logbook\_AwakeWear hereafter).

Consecutive 0 counts methods: The consecutive 0 count methods set a minimum required time of 0 counts to be considered nonwear. A count of 0 can be seen in both sedentary activity (e.g., no movement) and when the accelerometer is not worn (e.g., placed on a table). The theory is that it is unlikely for participants to be motionless for extended periods, suggesting the device is not being worn. Common minimum consecutive 0 counts range from 20 minutes (more common in children<sup>8</sup>), 60 minutes (more common in adults<sup>17</sup>), up to 120 minutes or more.<sup>8,17</sup> Based on the findings from the 2022 review in the early years,<sup>8</sup> we tested 5, 10, 20, 30, 60, and 90 minutes. All

consecutive 0 count methods were implemented in Python, where .gt3x files were read with pygt3x(v0.5.2), and counts obtained using agcounts(v0.2.3).

Modified consecutive 0 counts methods: The modified consecutive 0 counts methods follow the same principle as the consecutive 0 count method, but add further “rules” around spike tolerances (non-0-count epochs surrounded by 0 counts), size of spikes, etc. Two common methods are available in ActiLife, and are developed by Troiano et al.<sup>12</sup> (referred to as Troiano method hereafter) and Choi et al.<sup>13</sup> (referred to as Choi method hereafter). The Troiano method was applied to the NHANES primary accelerometer analyses.<sup>18</sup> Both methods were developed on a 60-second epoch count but are often applied to other epochs. We implemented both methods in ActiLife software using the count data in the original 60-second epoch and a 1-second epoch to allow for direct comparison to the other methods used in this paper.

Raw data methods: Raw data methods (using g’s, not counts) generally look at the variance of a signal within a given window and consider it nonwear if below a given threshold for that time window. These methods also often integrate additional “rules” related to short wear times surrounded by nonwear. The three methods tested here are those developed by van Hees et al.<sup>19</sup> and by Ahmadi et al.<sup>14</sup> (referred to as Ahmadi method hereafter). The van Hees method (used in GGIR<sup>20</sup>) uses a 3 mg standard deviation or 50 mg range threshold for 2 of the 3 axes using a 60-minute interval (sliding 15-minute window). It then checks the length of wear periods relative to bordering nonwear periods (see table 1 for details). We tested a previously published Python implementation of this method<sup>7</sup> which computes only the thresholds part of the method (not the bordering nonwear periods; referred to as van Hees method hereafter). This method was validated against the original published R code. We also tested an adapted version of the code which incorporates the bordering nonwear period times criteria (see table 1 for details; referred

to as van Hees 30/80 method hereafter). For both versions we read in the raw data from .gt3x files using pygt3x. The Ahmadi method uses a 13 mg vector magnitude threshold for a minimum of 30 minutes to determine nonwear, with bordering wear time rules (see table 1). We adapted the published R code to Python for use in this study and tested it against the R code.

**Table 1.** Details of nonwear detection methods compared in this study

<b>Method</b>	<b>Details on calculation</b>	<b>Software for implementation</b>
Logbook_ AwakeWear	<ul style="list-style-type: none"> <li>• 10 consecutive minutes of 0 counts to flag potential non-wear periods</li> <li>• Parent-reported on/off times of accelerometer excluded</li> <li>• Parent-reported daily nap times excluded</li> <li>• See supplementary material 1 for example logbook</li> </ul>	ActiLife
Logbook_ AllWear	<ul style="list-style-type: none"> <li>• 10 consecutive minutes of 0 counts to flag potential non-wear periods</li> <li>• Only parent-reported on/off times of accelerometer</li> <li>• See supplementary material 1 for example logbook</li> </ul>	ActiLife
5min_0count	<ul style="list-style-type: none"> <li>• Counts, 1-second epoch</li> <li>• 5 consecutive minutes of 0 counts</li> </ul>	Python
10min_0count	<ul style="list-style-type: none"> <li>• Counts, 1-second epoch</li> <li>• 10 consecutive minutes of 0 counts</li> </ul>	Python
20min_0count	<ul style="list-style-type: none"> <li>• Counts, 1-second epoch</li> <li>• 20 consecutive minutes of 0 counts</li> </ul>	Python
30min_0count	<ul style="list-style-type: none"> <li>• Counts, 1-second epoch</li> <li>• 30 consecutive minutes of 0 counts</li> </ul>	Python
60min_0count	<ul style="list-style-type: none"> <li>• Counts, 1-second epoch</li> <li>• 60 consecutive minutes of 0 counts</li> </ul>	Python
90min_0count	<ul style="list-style-type: none"> <li>• Counts, 1-second epoch</li> <li>• 90 consecutive minutes of 0 counts</li> </ul>	Python
Troiano60sec	<ul style="list-style-type: none"> <li>• Counts, 60-second epoch</li> <li>• 60 consecutive minutes of 0 counts                             <ul style="list-style-type: none"> <li>○ Including up to 2 consecutive minutes of non-0 counts up to and including 100 counts/min</li> <li>○ Any minute of &gt;100 counts/min is wear</li> </ul> </li> </ul>	ActiLife
Troiano1sec	<ul style="list-style-type: none"> <li>• Counts, 1-second epoch</li> <li>• 60 consecutive minutes of 0 counts</li> </ul>	ActiLife



	<ul style="list-style-type: none"> <li>○ Including up to 2 consecutive minutes of non-0 counts up to and including 100 counts/min</li> <li>○ Any minute of &gt;100 counts/min is wear</li> </ul>	
Choi60sec	<ul style="list-style-type: none"> <li>● Counts, 60-second epoch</li> <li>● 90 consecutive minutes of 0 counts               <ul style="list-style-type: none"> <li>○ Including up to 2 consecutive minutes of non-0 counts as long as 0 counts are detected in the 30 minutes before and after the non-0es</li> <li>○ Any other non-0 is wear</li> </ul> </li> </ul>	ActiLife
Choi1sec	<ul style="list-style-type: none"> <li>● Counts, 1-second epoch</li> <li>● 90 consecutive minutes of 0 counts               <ul style="list-style-type: none"> <li>○ Including up to 2 consecutive minutes of non-0 counts as long as 0 counts are detected in the 30 minutes before and after the non-0es</li> <li>○ Any other non-0 is wear</li> </ul> </li> </ul>	ActiLife
vanHees	<ul style="list-style-type: none"> <li>● Raw data</li> <li>● 60-minute sliding intervals (15 min slide window) of either:               <ul style="list-style-type: none"> <li>○ standard deviation of raw accelerations (g) &lt;0.003g for at least 2 out of 3 axes</li> <li>○ range of raw acceleration &lt;0.05g for at least 2 out of 3 axes</li> </ul> </li> </ul>	Python
vanHees_30/80	<ul style="list-style-type: none"> <li>● Raw data</li> <li>● 60-minute sliding intervals (15 min slide window) of either:               <ul style="list-style-type: none"> <li>○ standard deviation of raw accelerations (g) &lt;0.003g for at least 2 out of 3 axes</li> <li>○ range of raw acceleration &lt;0.05g for at least 2 out of 3 axes</li> </ul> </li> <li>● If wear period is &lt;6 hours and &lt; 30% of combined bordering nonwear periods, marked nonwear</li> <li>● If wear period is &lt;3 hours and &lt; 80% of combined bordering nonwear periods, marked nonwear</li> </ul>	Python
Ahmadi	<ul style="list-style-type: none"> <li>● Raw data</li> <li>● 30-minute sliding interval (1 second slide window) where standard deviation of vector magnitude (g) per second is &lt; 0.013 g</li> <li>● If wear periods &lt; 30 minutes and &lt; 30% of combined bordering nonwear periods, marked non-wear</li> </ul>	Python

## *Statistical Analysis*

For each method, we determined second-by-second wear/nonwear classification and day-level summary measures, the daily total minutes of wear/nonwear. We calculated the relative and absolute differences between fully automatic methods to both Logbook criteria. We also conducted a Two-One Sided Test (TOST) of equivalence with an equivalence bound of 5% of the average criterion minutes/day of wear. Further, we calculated epoch-level accuracy and an F1 score. Accuracy was calculated as the number of epochs correctly identified (compared to both criterion) divided by the total number of epochs. An F1 score is the harmonized mean of precision and recall. It is calculated as  $2 * \text{True\_Positives} / (2 * \text{True\_Positives} + \text{False\_Positives} + \text{False\_Negatives})$ , giving a value between 0 and 1 where, like accuracy, higher is better.<sup>21</sup> F1 score (weighted) was chosen as a metric in addition to accuracy as it can also account for the imbalance of labels (many cases of one label but very few cases of another). It is important to note that when there is a binary outcome (such as wear/nonwear), simple F1 score and accuracy are mathematically the same, while the weighted F1 can better account for the label imbalance. Accuracy and F1 score were calculated in Python(v3.9.6) using scikitlearn(v1.2.2) and TOST equivalence testing was completed in R(v4.3.1) using package TOSTER(v0.8.0).

## **Results**

Participant (n=109) and parent demographics are presented in Table 2. Participants had 6.9 average wear days (range: 3-9). All participants had completed logbooks, although 11 participants missed some on/off times (5 participants missing 1 on/off time, 4 missing 2 times, 1 missing 4 times, and 1 missing all off times). All participants had at least one nap, and 90 (83%) participants wore the accelerometer during at least one nap (426 total naps with accelerometer).

**Table 2.** Detailed participant characteristics

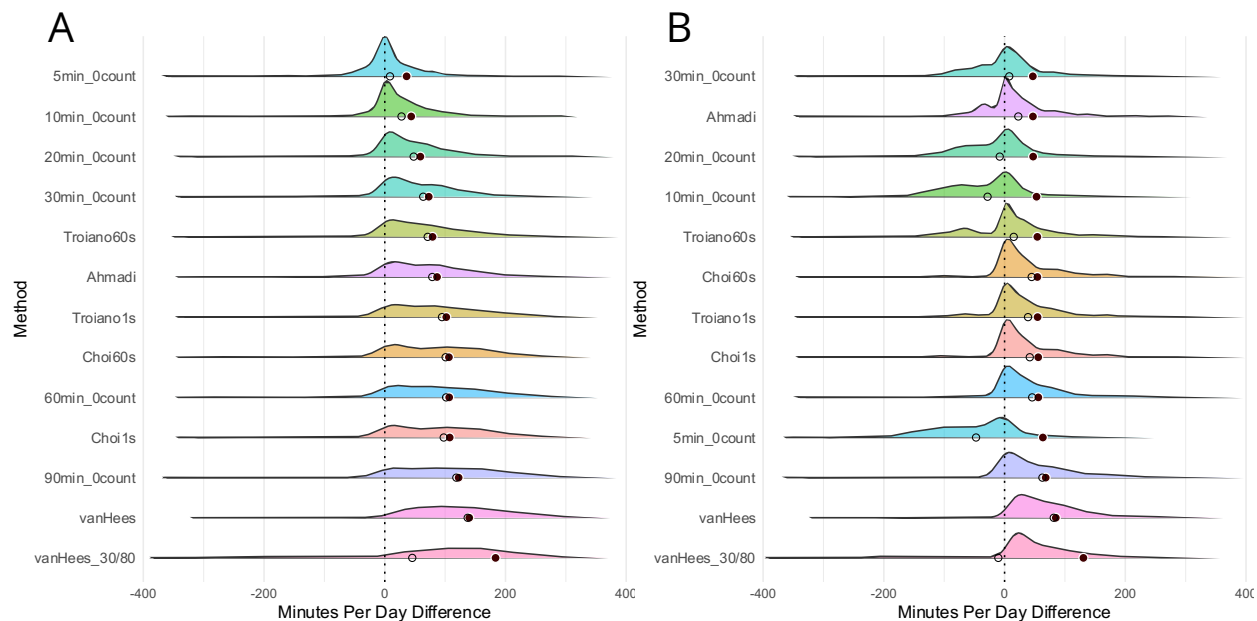
Participant characteristic	Overall n = 109
Age (months)	
Mean (SD)	21.5 (7.0)
Range	12.0-36.2
Sex	
Female	56 (51%)
Male	53 (49%)
Race/Ethnicity	
White	78 (72%)
Asian	5 (5%)
Latin American	3 (3%)
Arab	1 (1%)
Multiracial	22 (20%)
Home language	
English	98 (90%)
Other (Mandarin, Punjabi, Spanish, Portuguese, Arabic, Khmer, Italian)	8 (8%)
Not reported	3 (2%)
Marital status	
Married	86 (79%)
Living common-law or living with a partner	16 (15%)
Single, never married	4 (4%)
Separated	1 (1%)
Not reported	2 (2%)
Parent 1 highest education	
College/University	103 (94%)
High school	4 (4%)
Not reported	2 (2%)
Parent 2 highest education	
College/University	96 (88%)
High school	6 (6%)
Not reported	7 (6%)

Wear time per day for methods are presented in Table 3 and ranged from 559 (Logbook\_AwakeWear) to 684 min/day (van Hees). The wear time between Logbook\_AwakeWear and Logbook\_AllWear differed by 53.8 min/day. The results of the TOST equivalency tests are also presented in Table 3 (TOST details in supplementary material 2). Methods considered equivalent to the Logbook\_AllWear criterion were the 10min\_0count, 30min\_0count, Troiano60s, Ahmadi, and van Hees\_30/80 methods. Compared to

Logbook\_AwakeWear, only 5min\_0count was considered equivalent and had the smallest mean absolute difference in minutes of wear time. The mean absolute and relative differences per method are shown in Table 3 and Figure 1 (with distribution of relative differences). Compared to Logbook\_AllWear, 30min\_0count, Ahmadi, and 20min\_0count methods had the smallest mean absolute differences.

**Table 3.** Wear time, TOST equivalency test results, and mean absolute and relative differences for each nonwear detection method. Values shown as mean±SD.

	Average Wear Time (min/day)	Compared to Logbook_AwakeWear		Compared to Logbook_AllWear			
		Equivalent?	Mean Absolute Difference (min)	Mean Relative Difference (min)	Equivalent?	Mean Absolute Difference (min)	Mean Relative Difference (min)
Logbook_AwakeWear	559.3 ± 152.8	--	--	--	No	--	--
Logbook_AllWear	613.1 ± 165.5	No	--	--	--	--	--
5min_0count	555.6 ± 117.2	<b>Yes</b>	<b>36.2±52.2</b>	<b>8.6±62.9</b>	No	63.4±62.3	-47.3±75.3
10min_0count	574.6 ± 122.3	No	43.8±60.7	27.7±69.5	No	53.0±59.7	-28.2±74.7
20min_0count	594.8 ± 128.9	No	58.9±71.7	47.9±79.5	<b>Yes</b>	<b>47.3±61.0</b>	<b>-8.0±76.8</b>
30min_0count	610.6 ± 135.2	No	72.9±82.4	63.6±89.8	<b>Yes</b>	<b>46.8±67.3</b>	<b>7.7±81.7</b>
60min_0count	648.5 ± 148.7	No	106.7±104.4	101.6±109.3	No	55.9±87.6	45.7±93.3
90min_0count	665.5 ± 150.3	No	122.4±109.6	118.5±113.7	No	68.1±96.3	62.6±100.0
Troiano1s	641.7 ± 151.0	No	101.9±101.6	94.8±108.3	No	54.7±85.1	38.9±93.4
Troiano60s	618.0 ± 134.3	No	79.1±84.6	71.1±91.5	<b>Yes</b>	<b>54.3±70.7</b>	<b>15.2±87.9</b>
Choi1s	644.6 ± 161.3	No	107.5±105.1	97.6±114.4	No	55.8±91.1	41.7±98.3
Choi60s	647.7 ± 151.5	No	106.3±104.7	100.8±110.0	No	54.3±89.8	44.9±94.9
Ahmadi	625.5 ± 144.3	No	86.6±93.9	78.6±100.7	<b>Yes</b>	<b>47.0±76.3</b>	<b>22.7±86.7</b>
vanHees	684.3 ± 145.0	No	140.0±106.0	137.4±109.4	No	84.9±92.7	81.5±95.7
vanHees_30/80	592.4 ± 276.7	No	183.5±147.8	45.5±231.3	<b>Yes</b>	<b>130.6±158.6</b>	<b>-10.4±205.2</b>



**Figure 1.** Absolute and relative differences in daily wear time for automatic methods compared to Logbook\_AwakeWear (A) and Logbook\_AllWear (B). Solid points represent the mean absolute differences, open circles represent the mean relative differences, and the half violin plots show the distribution of the by-day relative differences (can show whether a method tends to over- or under-estimate wear time).

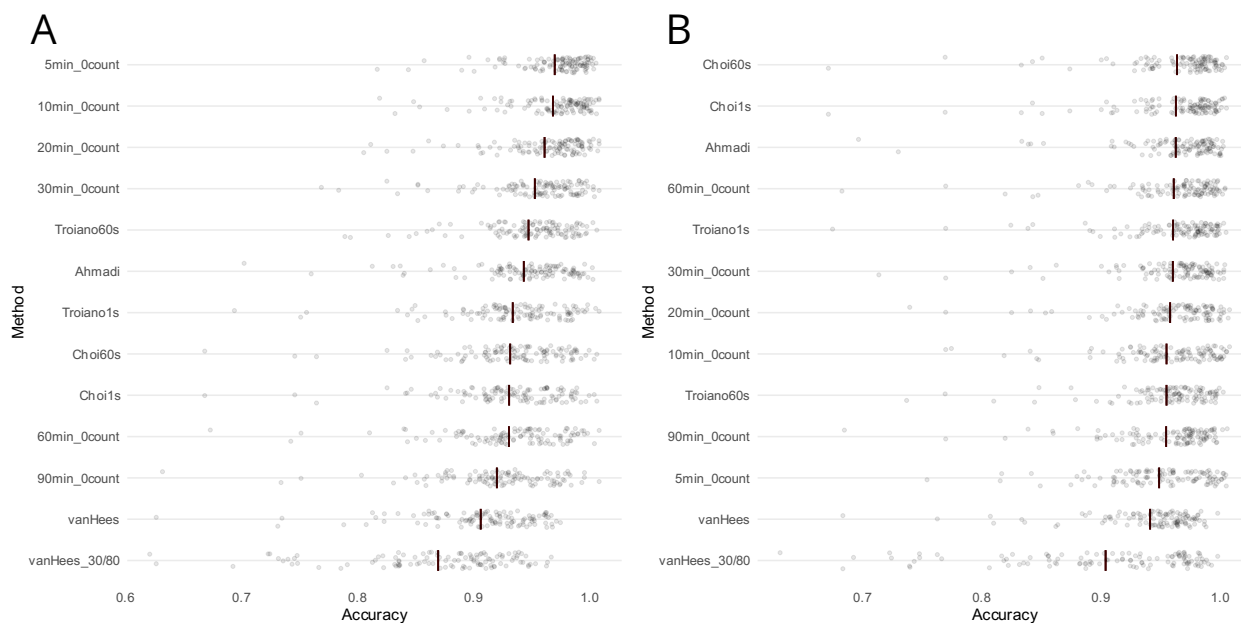
The accuracy and F1 scores of methods compared to the logbooks are shown in Table 4.

Compared to the Logbook\_AllWear, Choi60s had the highest accuracy and F1 scores (95.1%) and vanHees\_30/80 had the lowest (89.1%). Compared to Logbook\_AwakeWear, 5min\_0count (96.2%) and 10min\_0count (96.1%) had the highest accuracy and F1 scores with van Hees\_30/80 also having the lowest (86.1% accuracy, 86.2% F1 score). Figure 2 shows the average accuracy and accuracy per participant.

**Table 4.** Accuracy and F1 scores of fully automatic methods compared to both Logbook criterion methods

	Compared to Logbook_AwakeWear		Compared to Logbook_AllWear	
	Accuracy	F1 Score	Accuracy	F1 Score
5min_0count	0.962	0.962	0.943	0.943
10min_0count	0.961	0.961	0.949	0.949
20min_0count	0.953	0.954	0.952	0.952
30min_0count	0.945	0.945	0.955	0.955

60min_0count	0.922	0.923	0.956	0.956
90min_0count	0.911	0.912	0.949	0.949
Troiano1s	0.925	0.926	0.955	0.955
Troiano60s	0.939	0.940	0.949	0.949
Choi1s	0.922	0.923	0.957	0.958
Choi60s	0.923	0.924	0.958	0.959
Ahmadi	0.935	0.936	0.957	0.957
vanHees	0.898	0.899	0.936	0.936
vanHees_30/80	0.861	0.862	0.898	0.898



**Figure 2.** Accuracy for each method compared to Logbook\_AwakeWear (A) and Logbook\_AllWear (B). The solid line represents the mean accuracy, and the grey points are the accuracy for each participant.

## Discussion

Our findings demonstrate that the choice of nonwear detection method when using accelerometers in toddlers can alter wear-time outcomes, with methods differing by up to 129 min/day. Compared to Logbook\_AwakeWear, most methods overestimated awake wear time and had slightly lower accuracy, while compared to Logbook\_AllWear, methods more equally over- and under-estimated wear time and had higher accuracy. This is somewhat expected given that

existing methods were designed to detect device removal, and excluding wear-time naps will decrease the total wear time – thus increasing the mean absolute difference and shifting the other methods to a relative overestimate. Overall, accuracy and F1 scores were relatively high (>85%) and varied very little for any given method, showing that the classes (wear/nonwear) were relatively balanced. When compared to Logbook\_AwakeWear, only one method (5min\_0count) was considered equivalent. Compared to Logbook\_AllWear, more methods were equivalent, including mid-length consecutive zero counts methods and raw data methods with short wear periods. These novel results in toddlers generally support the findings from similar investigations in older children and adults, highlighting that method selection can have a meaningful impact on total wear time and provides valuable insights into wear-time considerations for toddlers.

In comparison to Logbook\_AllWear, the Ahmadi raw data method was often highly ranked, having the lowest mean absolute difference and one of the highest accuracies (96%). Also highly accurate are the Choi methods, but interestingly they have larger mean absolute differences. For Logbook\_AwakeWear, top-ranked were shorter 0count methods (5, 10, 20, and 30 minutes), with only the 5min\_0count method considered equivalent. This suggests that sustained motionless time during napping is shorter than 10 mins. Both van Hees methods have the largest mean absolute differences and lowest accuracy and F1 scores when compared against both criterion methods. This aligns with previous findings in school-age children, which found that it had the largest mean bias and lowest agreement.<sup>14</sup> Of note, the van Hees method was developed in adults and uses a relatively large window size (60 minutes) and low g threshold, which are not likely appropriate when applied to pediatric movement profiles.

Daily average awake wear is lower in toddlers than seen in older age groups,<sup>6</sup> given their developmental increased sleep needs.<sup>22</sup> The average awake-time wear reported in our study

(559.3 min/day) is similar to other toddler studies with 547.2<sup>23</sup> and 527.1<sup>1</sup> min/day of wear time. As napping is almost ubiquitous in this age group, this highlights the importance of decision-making around naps. Logbook\_AllWear looked at the “true” nonwear time, so included naps if the accelerometer was worn. However, as naps are difficult to detect when children are wearing the monitor,<sup>10</sup> and there are currently no algorithms created to distinguish between naps and sedentary activities using hip-worn accelerometers,<sup>15</sup> these nap times are likely to get misclassified as sedentary time. If sedentary time is the outcome of interest, researchers should consider asking parents to remove the device during naps, however, this could lead to a reduction in overall compliance as parents/caregivers may forget to put the device back on. Another solution could be our approach for Logbook\_AwakeWear – have parents report whether the device was worn or not, and manually remove the worn nap times before further analysis. While time-consuming, this likely allows for the most accurate analysis. Alternatively, our results show that the 5min\_0count method was equivalent to the Logbook\_AwakeWear criterion, so could be considered when manual cleaning is not feasible. Ultimately, the chosen nonwear method may depend on the objective/context of a specific analysis. If the analysis is looking at total physical activity and sedentary time of toddlers with a small number of participants, the semi-automatic logbook method may be ideal, allowing for a more individualized nonwear detection. However, if it’s a larger study, or has mostly incomplete logbooks, a less subjective and less time/resource-intensive automated method may be preferred, including the 5min\_0count, 30min\_0count, and Ahmadi methods, depending on whether you are looking to match AwakeWear or AllWear.

Our study is not without limitations. First, as previously mentioned, logbooks were completed to different levels of accuracy/detail, so criterion measures may not exactly reflect the ground truth of accelerometer nonwear. Second, we also had a wide age range of toddlers (12-36



months), and the accuracy of accelerometer processing methods may vary based on individual-level characteristics (e.g. motor development, nap schedule). Finally, while we had a good representation of age and sex, participants were mostly White, English-speaking, dual-parent, highly educated families. As such, these results may not be generalizable to more diverse and lower-income families, who may experience more challenges with accelerometer wear and logbook completion.<sup>24</sup> Logbooks provide a rich source of data when completed fully but add to participant burden which may be limiting for participants. An important avenue for future work should be to improve nap time vs. sedentary time detection in toddlers, as this is currently limited,<sup>10,15</sup> and understanding which techniques work best across the wide developmental spectrum of toddlerhood.

## **Conclusion**

This study presents a novel analysis of nonwear detection methods for toddler accelerometry. The toddler age poses specific issues relating to nonwear detection and data cleaning (e.g., accounting for naps, reduced wake-time). While many methods were considered equivalent to the AllWear nonwear criterion, fewer were equivalent to AwakeWear. This paper provides important insights and quantitative results that can help researchers decide which accelerometry processing method may be best given their population of interest, sample size, and study protocol.

## References

1. Bisson M, Tremblay F, Pronovost E, Julien AS, Marc I. Accelerometry to measure physical activity in toddlers: Determination of wear time requirements for a reliable estimate of physical activity. *Journal of Sports Sciences*. 2019;37(3):298-305. doi:10.1080/02640414.2018.1499391
2. Bruijns BA, Truelove S, Johnson AM, Gilliland J, Tucker P. Infants' and toddlers' physical activity and sedentary time as measured by accelerometry: A systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity*. 2020;17(1):1-14. doi:10.1186/S12966-020-0912-4/FIGURES/2
3. Hnatiuk J, Salmon J, Campbell KJ, Ridgers ND, Hesketh KD. Early childhood predictors of toddlers' physical activity: longitudinal findings from the Melbourne InFANT Program. *The International Journal of Behavioral Nutrition and Physical Activity*. 2013;10:123. doi:10.1186/1479-5868-10-123
4. Aadland E, Andersen LB, Anderssen SA, Resaland GK. A comparison of 10 accelerometer non-wear time criteria and logbooks in children. *BMC Public Health*. 2018;18(1):323. doi:10.1186/s12889-018-5212-4
5. Evenson KR, Terry Jr. JW. Assessment of Differing Definitions of Accelerometer Nonwear Time. *Research Quarterly for Exercise and Sport*. 2009;80(2):355-362. doi:10.1080/02701367.2009.10599570
6. Vanhelst J, Vidal F, Drumez E, et al. Comparison and validation of accelerometer wear time and non-wear time algorithms for assessing physical activity levels in children and adolescents. *BMC Medical Research Methodology*. 2019;19(1):72. doi:10.1186/s12874-019-0712-1
7. Syed S, Morseth B, Hopstock LA, Horsch A. Evaluating the performance of raw and epoch non-wear algorithms using multiple accelerometers and electrocardiogram recordings. *Sci Rep*. 2020;10(1):5866. doi:10.1038/s41598-020-62821-2
8. Breau B, Coyle-Asbil HJ, Vallis LA. The Use of Accelerometers in Young Children: A Methodological Scoping Review. *Journal for the Measurement of Physical Behaviour*. 2022;5(3):185-201. doi:10.1123/jmpb.2021-0049
9. King W, Li J, Leishear K, Mitchell J, Belle S. Determining activity monitor wear time: an influential decision rule. *J Phys Act Health*. 2011;8(4):566-580.
10. Galland B, Meredith-Jones K, Gray A, et al. Criteria for nap identification in infants and young children using 24-h actigraphy and agreement with parental diary. *Sleep Medicine*. 2016;19:85-92. doi:10.1016/j.sleep.2015.10.013
11. Kuzik N, Spence JC, Carson V. Machine learning sleep duration classification in Preschoolers using waist-worn ActiGraphs. *Sleep Medicine*. 2021;78:141-148. doi:10.1016/j.sleep.2020.12.019

12. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical Activity in the United States Measured by Accelerometer. *Medicine & Science in Sports & Exercise*. 2008;40(1):181. doi:10.1249/mss.0b013e31815a51b3
13. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of Accelerometer Wear and Nonwear Time Classification Algorithm. *Med Sci Sports Exerc*. 2011;43(2):357-364. doi:10.1249/MSS.0b013e3181ed61a3
14. Ahmadi MN, Nathan N, Sutherland R, Wolfenden L, Trost SG. Non-wear or sleep? Evaluation of five non-wear detection algorithms for raw accelerometer data. *Journal of Sports Sciences*. 2020;38(4):399-404. doi:10.1080/02640414.2019.1703301
15. Coyle-Asbil HJ, Breau B, Ma DWL, Haines J, Buchholz AC, Vallis LA. Compliance with the 24-hour movement behavior guidelines and the impact of sleep methods among toddler, preschool, and school-aged children enrolled in the Guelph Family Health Study. *Journal of Science and Medicine in Sport*. 2024;27(9):631-639. doi:10.1016/j.jsams.2024.05.014
16. Obeid J, Nguyen T, Cellucci T, Larché MJ, Timmons BW. Effects of acute exercise on circulating endothelial and progenitor cells in children and adolescents with juvenile idiopathic arthritis and healthy controls: a pilot study. *Pediatr Rheumatol*. 2015;13(1):41. doi:10.1186/s12969-015-0038-4
17. Migueles JH, Cadenas-Sanchez C, Ekelund U, et al. Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sports medicine (Auckland, NZ)*. 2017;47(9):1821-1845. doi:10.1007/S40279-017-0716-0
18. Matthews CE, Chen KY, Freedson PS, et al. Amount of Time Spent in Sedentary Behaviors in the United States, 2003–2004. *Am J Epidemiol*. 2008;167(7):875-881. doi:10.1093/aje/kwm390
19. Hees VT van, Gorzelniak L, León ECD, et al. Separating Movement and Gravity Components in an Acceleration Signal and Implications for the Assessment of Human Daily Physical Activity. *PLOS ONE*. 2013;8(4):e61691. doi:10.1371/journal.pone.0061691
20. van Hees V, Migueles JH. GGIR (3.1-2). Published online 2024.
21. sklearn.metrics.f1\_score. scikit-learn. Accessed July 3, 2023. [https://scikit-learn/stable/modules/generated/sklearn.metrics.f1\\_score.html](https://scikit-learn/stable/modules/generated/sklearn.metrics.f1_score.html)
22. Hirshkowitz M, Whiton K, Albert SM, et al. National Sleep Foundation’s updated sleep duration recommendations: final report. *Sleep Health: Journal of the National Sleep Foundation*. 2015;1(4):233-243. doi:10.1016/j.sleh.2015.10.004
23. Kelly LA, Knox A, Gonzalez C, et al. Objectively Measured Physical Activity and Sedentary Time of Suburban Toddlers Aged 12–36 Months. *International Journal of Environmental Research and Public Health*. 2022;19(11):6707. doi:10.3390/ijerph19116707

24. Crumbley C, Cepni AB, Taylor A, et al. Exploring Factors Associated With Accelerometer Validity Among Ethnically Diverse Toddlers. Published online September 27, 2023. doi:10.1123/pes.2022-0114