

IDENTIFICATION AND DEVELOPMENT OF PHYSICAL ACTIVITY ASSESSMENT METHODS IN
TODDLERS

By

Cailyn A. Van Camp

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Kinesiology – Doctor of Philosophy

2024

PUBLIC ABSTRACT

Engaging in physical activity (PA) is shown to have numerous health benefits for toddlers. It is unknown how many ways are available to measure the amount of PA toddlers engage in. The first aim of this dissertation was to review the available research to determine what methods exist for toddler activity measurement. Only 16 studies were found, highlighting that there is not a lot of research on this topic. Most of the methods were device-based, while only a few used surveys. Three additional studies have applied developed methods to their own sample to determine if the methods are valid in various other populations. No one has created a measurement tool for toddlers using direct observation. These tools can be used to capture a complete picture of toddler PA behavior instead of just looking at movement amount. Therefore, the second aim of this dissertation was to create a direct observation tool that can be used in toddlers. The Observational System for Recording Physical Activity in Children – Toddlers (OSRAC-T) was created and determined to be reliable. This version of the OSRAC system included an added category to assess toddler support (e.g., weight bearing or not weight bearing). We also added additional codes, such as tantrum, to account for toddler-specific behavior. When using the OSRAC-T to assess behavior, most of toddlers' time was observed inactive and either sitting, squatting or standing. Toddlers spent a small amount of time in vigorous activity. We determined that the tool was reliable when comparing two independent people's codes and when comparing live coding to video coding. Despite being reliable, using direct observation can be difficult and time consuming for researchers. Because of this, some people prefer to use device-based measurements of physical activity. Therefore, the final aim of this dissertation was to apply already developed device-based measurement

methods in an independent sample of toddlers. We compared accuracy depending on where the device was placed and how the device was set up to collect data. Various calculations were performed to determine accuracy of each method to estimate time spent at different intensity levels. Two methods were determined to be the same as direct observation in assessing sedentary time. Accuracy did not change based on where the device was worn or how the axes data was collected using. The difference between these applied metrics and DO could be due to how our data was collected and assessed. This dissertation aimed to provide a complete overview of how PA has been assessed in toddlers and then to not only assess those current methods, but to create a method using direct observation. Being able to accurately assess PA in toddlers will lead to a better childcare center design that could improve PA levels in this group. We could also use this information to better create and evaluate programs that would increase PA in this population.

ABSTRACT

Engaging in an adequate amount of physical activity (PA) during the toddler years (12-36 months) has not only been associated with physical health, but with positive cognitive and mental health outcomes as well. The first aim of this dissertation was to systematically review the literature on toddler PA assessment methods. We found 16 articles on this topic that met inclusion criteria, which highlights the notion that there are limited assessment methods available for use in toddlers. Most identified methods were developed using accelerometry, while a few survey-based methods have been created. Few studies have cross-validated methods developed for use in preschoolers in toddlers. No direct observation (DO) methods have been developed for use in toddlers. These DO tools can provide more information on the activity and social contexts of toddler PA behavior. The review also shed light on the lack of standardization across measurement protocols to facilitate comparability across studies. The second aim of this dissertation was to develop an observational tool (Observational System for Recording Physical Activity in Children – Toddler (OSRAC-T)) to assess PA in toddlers and determine if the tool was reliable. We created this tool by adapting the Observational System for Recording Physical Activity in Children – Preschool (OSRAC-P). This tool allows for the assessment of PA level, type, and context, in addition to activity initiation and prompts, and group context. In collaboration with experts, a new category to assess support (i.e., weight-bearing v. non-weightbearing) and new codes (e.g., tantrum) were added to the original code to better apply to toddlers. Using the OSRAC-T, toddlers spent the majority of the observed time sedentary (70%) and only a small amount of time was spent in moderate-to-vigorous PA (10%). Sit/squat and stand accounted for most activity types (68.6%). We concluded that the tool was

a reliable observation system ($k = 0.46 - 0.69$). Additionally, comparable observations were made during live and video coded sessions ($k = 0.47 - 0.83$). This instrument can be used to inform toddler childcare providers or intervention design. The use of observation can be challenging for researchers due to the extensive time commitment needed and the potential for high levels of subjectivity in assessment, which leads many researchers to opt for accelerometer-based assessment methods. The final aim of this dissertation was to cross-validate various cut-points for assessing sedentary time (ST) and PA in toddlers using hip- and wrist-placed accelerometers and vertical axis and vector magnitude data. The same sample and methods were used for studies 2 and 3. Mean absolute difference (MAD), Pearson's r correlation coefficient, and equivalence testing were calculated for all classification methods, for each activity intensity. Percentage of time classified as sedentary, light, or MVPA varied greatly between cut-points. Accelerometer placement or data axes used did not influence accuracy. No set of cut-points, applied to any intensity, were determined to be equivalent to direct observation. The variability in these estimation methods may be due to the varying epochs at which cut-points were calibrated or the way activity intensities were classified using direct observation. This dissertation addressed recent calls to advance physical activity assessment methods in young children, including toddlers. Further development in toddler activity assessment will lead to better intervention design to promote activity engagement.

To the girls who were told you weren't smart enough. I was once in your shoes. But now I have a PhD and they don't. Never give up.

ACKNOWLEDGEMENTS

I would like to start by thanking my advisor, Dr. Karin Pfeiffer, who has guided and supported my work over the last four years. I am truly grateful that I was sent and accepted to work in your lab and to have the opportunity to collaborate and work with you throughout my doctoral education. I would not be where I am today without that support. I would also like to thank those who I have had the opportunity to work with both at MSU and beyond (the CMAH lab). I owe a world of thanks to Dr. Kimberly Clevenger who has been so supportive with not only this dissertation, but numerous other projects I have been a part of during my time at MSU. I would also like to thank my other committee members Dr. Kerry McIver-Cordan and Dr. Janet Hauck for your assistance and direction on this dissertation. I am forever grateful for everyone that has contributed to and pushed this project to fruition.

I next want to thank my friends in the Kinesiology department that have helped to get me here. I want to specifically thank Amy Boettcher and Ashley McPeck for just being you. I literally would not have survived grad school without you both. I want to thank my undergraduate assistants that assisted me with my work in the last four years. I appreciate every one of you more than you know. I want to thank the parents and the children in this study for volunteering to be a part of this research. Thanks to the MSU College of Education and the Midwest Chapter of the American College of Sports Medicine for assisting with funding for these projects.

Lastly, but most importantly, I would like to thank my family and friends for their various levels of support throughout this degree. I know I was difficult to deal with at times, but you never gave up on me. This degree is almost as much yours as it is mine (Tyler!!).

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
REFERENCES	10
CHAPTER 2: REVIEW OF LITERATURE	15
REFERENCES	42
CHAPTER 3: A SYSTEMATIC REVIEW OF PHYSICAL ACTIVITY ASSESSMENT METHODS DEVELOPED FOR/IN TODDLERS	51
REFERENCES	67
CHAPTER 4: DEVELOPMENT AND TESTING OF THE OBSERVATIONAL SYSTEM FOR RECORDING PHYSICAL ACTIVITY IN CHILDREN – TODDLERS.....	72
REFERENCES	91
CHAPTER 5: ACCURACY OF HIP AND WRIST COLLECTED ACCELEROMETER DATA COMPARED TO DIRECT OBSERVATION IN TODDLERS.....	94
REFERENCES	118
CHAPTER 6: OVERALL DISCUSSION AND CONCLUSIONS	121
REFERENCES	132
APPENDIX A. NOVEL PHYSICAL ACTIVITY ASSESSMENT METHOD STUDY CHARACTERISTICS	136
APPENDIX B. CATEGORIES AND CODES OUTLINED IN THE OSRAC-T	140

CHAPTER 1: INTRODUCTION

The health benefits of habitual physical activity (PA) in early years of life (0 – 4 years) are well-established (Timmons et al., 2012). Engaging in an adequate amount of PA at this age has not only been associated with physical health, but with positive cognitive and mental health outcomes as well (Lee et al., 2017). For example, previous findings suggest that increases in moderate-to-vigorous physical activity (MVPA) are associated with decreases in body mass index (BMI) z-score and increases in social cognition (Cliff et al., 2009; Lee et al., 2017). Because of these recognized associations, governing bodies have recently published guidelines for movement in children younger than 5 years of age (Willumsen & Bull, 2020). It is recommended by The World Health Organization that toddlers (1 – 2 years old) engage in 180-minutes of PA per day at any intensity (light, moderate, or vigorous). However, the United States advisory committee determined there was not enough evidence to suggest guidelines for children younger than three years old (U.S. Department of Health and Human Service, 2018). Regardless, to identify the optimal amount and type of PA and inform future recommendations, it is imperative that we can accurately assess PA levels in toddlers.

PA can be assessed in many ways, such as survey-based instruments (questionnaires or recall diaries), direct observation (DO), or accelerometry (Troost, 2007). Each method has strengths and limitations. Several surveys have been developed for use in the toddler population; however, the dependency on parent- or teacher report and recall can introduce bias (Bingham et al., 2016; Bonn et al., 2012). Other measures may be more appropriate in younger children (Troost, 2007).

Direct observation is considered the criterion measure of PA in younger populations (Troost, 2007). Direct observation requires a trained researcher to observe and code participant

behavior using a specific tool. This technique allows for additional environmental and social contexts of activity to be accounted for in PA assessment. Several observational tools exist for use in children including the Children's Physical Activity Form (CPAF; O'hara et al., 1989), the Children's Activity Rating Scale (CARS; Puhl et al., 1990), the Behaviors of Eating and Activity for Children's Health Evaluation System (BEACHES; McKenzie, Sallis, Nader, et al., 1991), the System for Observing Fitness Instruction Time (SOFIT; McKenzie, Sallis, & Nader, 1991), the System for Observing Play and Leisure Activity in Youth (SOPLAY; (McKenzie et al., 2002) and the Observational System for Recording Physical Activity in Children – Preschool version (OSRAC-P; Brown et al., 2006).

Very few studies have used direct observation (DO) to assess PA in toddlers (Dinkel et al., 2019; Fees et al., 2015; Gubbels et al., 2011; Van Cauwenberghe et al., 2011). The only observational tool that has been utilized in this age group is the OSRAC-P, which is the most frequently used observation tool in children (Clevenger et al., 2020). The OSRAC-P can be used to describe PA level (e.g., sedentary, sedentary + limbs, slow and easy), type (e.g., sit, squat, walk), location (indoor, outdoor), group composition (e.g., solitary, 1-1 peer, group), initiator of activity (adult, child, cannot tell), indoor/outdoor context (e.g., art, self-care, ball, sandbox), and prompts both indoors and outdoors. Studies of toddlers have used or adapted parts of the OSRAC-P, such as assessing PA intensity and location (Dinkel et al., 2019; Fees et al., 2015; Gubbels et al., 2011; Van Cauwenberghe et al., 2011); three assessed activity type and prompts (Dinkel et al., 2019; Fees et al., 2015; Van Cauwenberghe et al., 2011), and two assessed activity contexts, group composition, and initiators (Dinkel et al., 2019; Fees et al., 2015).

All studies mentioned above report that toddlers engage in elevated levels of sedentary time both indoors and outdoors; however, more time indoors is spent sedentary compared to outdoors (Dinkel et al., 2019; Fees et al., 2015; Gubbels et al., 2011; Van Cauwenberghe et al., 2011). Similarly, more time outdoors is spent in MVPA compared to indoors. All studies reported the most common types of activity to be sitting/squatting, standing, and walking. Most activities are child initiated and done without prompts from staff or peers. Differences were reported in activity contexts and group composition indoors when compared to outdoors. It is well known in preschoolers that these contexts are associated with higher or lower levels of PA (Brown et al., 2009; Clevenger et al., 2020, 2021, 2022; Pate et al., 2008). Similar research in toddlers is much more limited.

Associations have been reported between PA levels of toddlers and group composition (Gubbels et al., 2011), location (Dinkel et al., 2019), and activity type (Fees et al., 2015). Gubbels et al. (2011) concluded that larger groups were associated with lower levels of PA. Dinkel et al. (2019) reported that in fine motor areas, toddlers were more sedentary, while in open spaces they were more active. Lastly, Fees et al. (2015) found that toddlers that engaged in onlooking had decreased engagement in MVPA. No study has been conducted to assess the associations between activity level and other factors such as location, activity type, and contexts both indoors and outdoors. It should also be noted that onlooking (which is negatively associated with PA levels) is not an activity included in the OSRAC-P; it was added, in addition to several other categories, so that the tool was better suited for toddlers (Fees et al., 2015). No study has fully assessed all environmental contexts of PA and their impact on toddler PA level.

Two of these four studies concluded that the OSRAC-P would need to be adapted to be used within a toddler population (Dinkel et al., 2019; Fees et al., 2015). In these studies, additional location categories were created (Dinkel et al., 2019) or activity codes and contexts were added (Fees et al., 2015). As noted, one of these activity types was onlooking. Onlooking is a behavior described as a child observing other children playing, but not engaging themselves (Hawley & Little, 1999). Fees et al. (2015) reported this behavior was more common among toddlers than gross motor behavior. In these adaptations, one study (Dinkel et al., 2019) assessed toddler PA outdoors, while the other assessed indoor activity (Fees et al., 2015). No study has assessed the behavior of the same sample in both settings. Also, although these studies made changes to the OSRAC-P, their purpose was not to formally create an adapted observational tool for future use by other researchers.

The OSRAC-P has been adapted into four additional versions, the Observational System for Recording Physical Activity in Children: Home (OSRAC-H; (McIver et al., 2009), the Observational System for Recording Physical Activity in Children: Youth Sport (OSRAC-YS; (Cohen et al., 2014), the Observational System for Recording Physical Activity in Children: Elementary School (OSRAC-E; (McIver et al., 2016), and the Observational System for Recording Physical Activity in Children: Developmental Disabilities (OSRAC-DD; (Schenkelberg et al., 2021). All adaptations included an instrument development phase and an instrument evaluation phase. All tools were concluded to be reliable and valid for PA in their target population, indicating that adaptations are both possible and successful. A tool developed specifically for toddlers would allow for a better understanding of PA behavior and would help to inform future intervention methods to increase PA in this population.

DO has several limitations, including its inability to capture the sporadic nature of children's activity. Additionally, due to the nature of DO being so time intensive, it can become burdensome for researchers and cannot be used for 24-hour monitoring. Accelerometry can be used to better capture this sporadic movement. Accelerometry is a non-invasive technique that requires participants to wear a small device that assesses one or more planes of movement. The device records the magnitude and direction of acceleration, which can be translated into PA intensity levels (e.g., moderate; Cliff et al., 2009). A common way of translating acceleration into PA levels is through use of cut points. The development of cut points in preschoolers (3 – 5 years old) is extensive (Pate et al., 2006; Pfeiffer et al., 2006; Sirard et al., 2005). However, some researchers propose that these cut points cannot be translated to toddlers due to differences in activity patterns (Cardon et al., 2011). When applying preschool cut points to toddlers, some instances result in an overestimation of sedentary and activity levels, while others result in an underestimation (Troost et al., 2012; Van Cauwenberghe et al., 2011). To overcome these discrepancies, cut points have been developed and validated in toddlers.

At least seven sets of cut points have been calibrated and validated in toddlers (Costa et al., 2014; Hager et al., 2016; Johansson, Ekelund, et al., 2015; Kelly & Villalpando, 2016; Oftedal et al., 2014; Pulakka et al., 2013; Trost et al., 2012). All studies recruited healthy participants without any physical impairments that would impact activity levels. Oftedal et al. (2014) recruited an additional sample of toddlers with varying classifications of cerebral palsy and calibrated cut points for each sample. All studies were conducted using different devices (uniaxial v. triaxial) and device settings, such as wear locations (e.g., hip, wrist, ankle, back) and epoch lengths (e.g., 5s, 15s, 30s). Across studies, final cut points ranged from ≤ 5 counts per 5s

to < 208 counts per 15s for sedentary time to < 35 counts per 15s to ≥ 1101 counts per 30s for MVPA. These ranges are quite large, making it difficult to know which cut points to apply. To our knowledge, no advanced methodologies, such as machine learning techniques, have been developed within this population.

Several studies have utilized different cut points to assess toddlers' PA levels across 24-hours (Armstrong et al., 2019; Bisson et al., 2019; Johansson, Hagströmer, et al., 2015; Kwon et al., 2019). Across these studies, free-living PA levels were assessed during waking hours for three to seven days. ActiGraphs were the most common accelerometer brand used and they were frequently placed at the right hip/waist. Most studies utilized 15-second epochs, and Trost (2012) cut points were most frequently applied. Overall, all studies varied in methodology, resulting in a wide range of reported activity levels. For example, MVPA levels reported ranged from 6.5 – 89.9 minutes/day across studies. A recent systematic review and meta-analysis aimed to assess accelerometer-based toddlers' PA levels (Bruijns et al., 2020). Authors reported these vast differences but concluded that when controlling for these variables (e.g., accelerometer brand, wear location, and epoch length) among studies, that toddlers are meeting the daily PA recommendations of 180-minutes of PA at any intensity. However, sedentary time still accounts for almost 50% of the day, and activity at higher intensities is minimal. Therefore, accurate measurement of PA level using accelerometry is a key first step to informing and assessing activity promotion.

Summary

Toddler PA behavior is different from that of preschool aged children. In preschoolers, we know that this behavior is heavily influenced by the environment, which has been reported

at childcare centers and can be seen both in the classroom and on the playground. Similar associations have been proposed in toddlers, but activity has not been fully described in this population. More information is needed on the best practices for device-based PA assessment in toddlers. Furthermore, there is currently no DO tool developed to specifically describe toddler PA and its context. Thus, this dissertation's overarching purpose is to identify and develop PA assessment methods in toddlers. Following a systematic review of the literature on PA assessment methods in toddlers, we will use the data to inform the creation of an observation tool for toddlers. This tool will allow us to accurately identify PA behaviors such as activity types, contexts, and group compositions. Additionally, best practices for accelerometer monitoring will be assessed. This dissertation will provide practical recommendations for measuring PA in toddlers to inform the methodology of future large-scale studies and design interventions that promote PA in this age group.

AIMS AND HYPOTHESIS

Aim 1. To systematically review the literature on toddler physical activity assessment methods.

Hypothesis 1a: Available literature would support that physical activity assessment methods specifically developed for toddlers are limited.

Hypothesis 1b: Accelerometry would be the most utilized objective assessment method and cut points will be the most common data reduction approach used.

Aim 2. To develop the Observational System for Recording Physical Activity in Children – Toddlers and examine its reliability and validity.

Hypothesis 2a: The OSRAC-T would be a reliable method for assessing physical activity in toddlers (Intraclass Correlation Coefficient (ICC) > 0.60; Interobserver Agreement (IOA) > 70%).

Aim 3. To cross-validate numerous cut-points for assessing physical activity and sedentary time in toddlers using direct observation as the criterion measure.

Hypothesis 3a. A hip-worn accelerometer would provide a more valid estimate of free-play physical activity than a wrist worn accelerometer when both are compared to direct observation.

Hypothesis 3b. Hip and wrist accelerometer data collected using an ActiGraph wGT3X-BT monitor would be moderately correlated ($r = 0.50$) to direct observation.

REFERENCES

- Ahmadi, M. N., & Trost, S. G. (2022). Device-based measurement of physical activity in pre-schoolers: Comparison of machine learning and cut point methods. *PLOS ONE*, *17*(4), e0266970. <https://doi.org/10.1371/journal.pone.0266970>
- Ahmadi, M., O'Neil, M., Fragala-Pinkham, M., Lennon, N., & Trost, S. (2018). Machine learning algorithms for activity recognition in ambulant children and adolescents with cerebral palsy. *Journal of NeuroEngineering and Rehabilitation*, *15*(1), 105. <https://doi.org/10.1186/s12984-018-0456-x>
- Altenburg, T. M., de Vries, L., op den Buijsch, R., Eyre, E., Dobell, A., Duncan, M., & Chinapaw, M. J. M. (2022). Cross-validation of cut-points in preschool children using different accelerometer placements and data axes. *Journal of Sports Sciences*, *40*(4), 379–385. <https://doi.org/10.1080/02640414.2021.1994726>
- Arem, H., Keadle, S. K., & Matthews, C. E. (2015). Invited Commentary: Meta-Physical Activity and the Search for the Truth. *American Journal of Epidemiology*, *181*(9), 656–658. <https://doi.org/10.1093/aje/kwu472>
- Bianchim, M. S., McNarry, M. A., Larun, L., Barker, A. R., Williams, C. A., & Mackintosh, K. A. (2020). Calibration and validation of accelerometry using cut-points to assess physical activity in paediatric clinical groups: A systematic review. *Preventive Medicine Reports*, *19*, 101142. <https://doi.org/10.1016/j.pmedr.2020.101142>
- Bingham, D., Collings, P., Clemes, S., Costa, S., Santorelli, G., Griffiths, P., & Barber, S. (2016). Reliability and Validity of the Early Years Physical Activity Questionnaire (EY-PAQ). *Sports*, *4*(2), 30. <https://doi.org/10.3390/sports4020030>
- Breau, B., Coyle-Asbil, H. J., & Vallis. (2022). The Use of Accelerometers in Young Children: A Methodological Scoping Review. *Journal for the Measurement of Physical Behavior*, *5*(3), 185–201.
- Browne, M. W. (2000). Cross-Validation Methods. *Journal of Mathematical Psychology*, *44*(1), 108–132. <https://doi.org/10.1006/jmps.1999.1279>
- Bruijns, B. A., Truelove, S., Johnson, A. M., Gilliland, J., & Tucker, P. (2020). 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*, *17*(1), 14. <https://doi.org/10.1186/s12966-020-0912-4>
- Butte, N. F., Wong, W. W., Lee, J. S., Adolph, A. L., Puyau, M. R., & Zakeri, I. F. (2014). Prediction of Energy Expenditure and Physical Activity in Preschoolers. *Medicine & Science in Sports & Exercise*, *46*(6), 1216–1226. <https://doi.org/10.1249/MSS.0000000000000209>

- Cain, K. L., Sallis, J. F., Conway, T. L., Van Dyck, D., & Calhoun, L. (2013). Using Accelerometers in Youth Physical Activity Studies: A Review of Methods. *Journal of Physical Activity and Health, 10*(3), 437–450. <https://doi.org/10.1123/jpah.10.3.437>
- Carson, V., Lee, E.-Y., Hesketh, K. D., Hunter, S., Kuzik, N., Predy, M., Rhodes, R. E., Rinaldi, C. M., Spence, J. C., & Hinkley, T. (2019). Physical activity and sedentary behavior across three time-points and associations with social skills in early childhood. *BMC Public Health, 19*(1), 27. <https://doi.org/10.1186/s12889-018-6381-x>
- Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). *Physical Activity, Exercise, and Physical Fitness: Definitions and Distinctions for Health-Related Research*.
- Costa, S., Barber, S. E., Cameron, N., & Clemes, S. A. (2014). Calibration and validation of the ActiGraph GT3X+ in 2–3 year olds. *Journal of Science and Medicine in Sport, 17*(6), 617–622. <https://doi.org/10.1016/j.jsams.2013.11.005>
- Evenson KR, Catellier DJ, Gill K, Ondrak KS, & McMurray RG. (2008). Calibration of two objective measures of physical activity for children. *Journal of Sports Sciences, 26*(14), 1557–1565. rzh. <https://doi.org/10.1080/02640410802334196>
- Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of Accelerometer Output for Children. *Medicine & Science in Sports & Exercise, 37*(11), S523–S530. <https://doi.org/10.1249/01.mss.0000185658.28284.ba>
- Haar, S., Fees, B., Trost, S., Crowe, L. K., & Murray, A. (2013). Design of a Garment for Data Collection of Toddler Language and Physical Activity. *Clothing and Textiles Research Journal, 31*(2), 125–140. <https://doi.org/10.1177/0887302X13478161>
- Hager, E. R., Gormley, C. E., Latta, L. W., Treuth, M. S., Caulfield, L. E., & Black, M. M. (2016). Toddler physical activity study: Laboratory and community studies to evaluate accelerometer validity and correlates. *BMC Public Health, 16*(1), 936. <https://doi.org/10.1186/s12889-016-3569-9>
- Henriksson, H., Alexandrou, C., Henriksson, P., Henström, M., Bendtsen, M., Thomas, K., Müssener, U., Nilsen, P., & Löf, M. (2020). MINISTOP 2.0: A smartphone app integrated in primary child health care to promote healthy diet and physical activity behaviours and prevent obesity in preschool-aged children: Protocol for a hybrid design effectiveness-implementation study. *BMC Public Health, 20*(1), 1–11. rzh. <https://doi.org/10.1186/s12889-020-09808-w>
- Hidding, L. M., Chinapaw, Mai. J. M., van Poppel, M. N. M., Mokkink, L. B., & Altenburg, T. M. (2018). An Updated Systematic Review of Childhood Physical Activity Questionnaires. *Sports Medicine, 48*(12), 2797–2842. rzh. <https://doi.org/10.1007/s40279-018-0987-0>

- Johansson, E., Ekelund, U., Nero, H., Marcus, C., & Hagströmer, M. (2015). Calibration and cross-validation of a wrist-worn Actigraph in young preschoolers: Calibration of Actigraph in toddlers. *Pediatric Obesity, 10*(1), 1–6. <https://doi.org/10.1111/j.2047-6310.2013.00213.x>
- Kelly, L. A., & Villalpando, J. (2016). Development of Actigraph GT1M Accelerometer Cut-Points for Young Children Aged 12-36 Months. *Journal of Athletic Enhancement, 5*(4). <https://doi.org/10.4172/2324-9080.1000233>
- Klesges, L. M., & Klesges, R. C. (1987). *The assessment of children's physical activity: A comparison of methods. 19*(5), 511–517.
- Kwon, S., Honegger, K., & Mason, M. (2019). Daily Physical Activity Among Toddlers: Hip and Wrist Accelerometer Assessments. *International Journal of Environmental Research and Public Health, 16*(21), 4244. <https://doi.org/10.3390/ijerph16214244>
- Lee, E.-Y., Hesketh, K. D., Hunter, S., Kuzik, N., Rhodes, R. E., Rinaldi, C. M., Spence, J. C., & Carson, V. (2017). Meeting new Canadian 24-Hour Movement Guidelines for the Early Years and associations with adiposity among toddlers living in Edmonton, Canada. *BMC Public Health, 17*(S5), 840. <https://doi.org/10.1186/s12889-017-4855-x>
- Lettink, A., Altenburg, T. M., Arts, J., van Hees, V. T., & Chinapaw, M. J. M. (2022). Systematic review of accelerometer-based methods for 24-h physical behavior assessment in young children (0–5 years old). *International Journal of Behavioral Nutrition & Physical Activity, 19*(1), 1–63. <https://doi.org/10.1186/s12966-022-01296-y>
- Li, S., Howard, J. T., Sosa, E. T., Cordova, A., Parra-Medina, D., & Yin, Z. (2020). Calibrating Wrist-Worn Accelerometers for Physical Activity Assessment in Preschoolers: Machine Learning Approaches. *JMIR Formative Research, 4*(8), e16727. <https://doi.org/10.2196/16727>
- Matthews, C. E., Keadle, S. K., Berrigan, D., Lyden, K., & Troiano, R. P. (2021). Influence of Accelerometer Calibration Approach on Moderate–Vigorous Physical Activity Estimates for Adults—Corrigendum. *Medicine & Science in Sports & Exercise, 53*(9), 2018–2018. <https://doi.org/10.1249/MSS.0000000000002669>
- Matthews, C. E., Keadle, S. K., Sampson, J., Lyden, K., Bowles, H. R., Moore, S. C., Libertine, A., Freedson, P. S., & Fowke, J. H. (2013). Validation of a Previous-Day Recall Measure of Active and Sedentary Behaviors. *Medicine & Science in Sports & Exercise, 45*(8), 1629–1638. <https://doi.org/10.1249/MSS.0b013e3182897690>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & for the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ, 339*(jul21 1), b2535–b2535. <https://doi.org/10.1136/bmj.b2535>

- Montoye, A. H. K., Clevenger, K. A., Mackintosh, K. A., McNarry, M. A., & Pfeiffer, K. A. (2019). Cross-Validation and Comparison of Energy Expenditure Prediction Models Using Count-Based and Raw Accelerometer Data in Youth. *Journal for the Measurement of Physical Behaviour*, 2(4), 237–246. <https://doi.org/10.1123/jmpb.2018-0011>
- Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., Currie, M., Qureshi, R., Mattis, P., Lisy, K., & Mu, P.-F. (2020). Chapter 7: Systematic reviews of etiology and risk. In *JBI Manual for Evidence Synthesis*.
- Oftedal, S., Bell, K. L., Davies, P. S. W., Ware, R. S., & Boyd, R. N. (2014). Validation of Accelerometer Cut Points in Toddlers with and without Cerebral Palsy. *Medicine & Science in Sports & Exercise*, 46(9), 1808–1815. <https://doi.org/10.1249/MSS.0000000000000299>
- Pate, R. R., Almeida, M. J., McIver, K. L., Pfeiffer, K. A., & Dowda, M. (2006). Validation and Calibration of an Accelerometer in Preschool Children*. *Obesity*, 14(11), 2000–2006. <https://doi.org/10.1038/oby.2006.234>
- Pereira, J. R., Sousa-Sá, E., Zhang, Z., Cliff, D. P., & Santos, R. (2020). Concurrent validity of the ActiGraph GT3X+ and activPAL for assessing sedentary behaviour in 2–3-year-old children under free-living conditions. *Journal of Science and Medicine in Sport*, 23(2), 151–156. <https://doi.org/10.1016/j.jsams.2019.08.009>
- Pfeiffer, K. A., Mciver, K. L., Dowda, M., Almeida, M. J. C. A., & Pate, R. R. (2006). Validation and Calibration of the Actical Accelerometer in Preschool Children. *Medicine & Science in Sports & Exercise*, 38(1), 152–157. <https://doi.org/10.1249/01.mss.0000183219.44127.e7>
- Pulakka, A., Cheung, Y., Ashorn, U., Penpraze, V., Maleta, K., Phuka, J., & Ashorn, P. (2013). Feasibility and validity of the ActiGraph GT3X accelerometer in measuring physical activity of Malawian toddlers. *Acta Paediatrica*, 102(12), 1192–1198. <https://doi.org/10.1111/apa.12412>
- Saint-Maurice, P. F., Welk, G. J., Bartee, R. T., & Heelan, K. (2017). Calibration of context-specific survey items to assess youth physical activity behaviour. *Journal of Sports Sciences*, 35(9), 866–872. <https://doi.org/10.1080/02640414.2016.1194526>
- Sarker, H., Anderson, L., Borkhoff, C., Abreo, K., Tremblay, M., Lebovic, G., Maguire, J., Parkin, P., & Birken, C. (2015). Validation of Parent-Reported Physical and Sedentary Activity by Accelerometry in Young Children. *Canadian Journal of Diabetes*, 39, S44–S44. <https://doi.org/10.1016/j.jcjd.2015.01.169>
- Sirard, J. R., & Pate, R. R. (2001). Physical Activity Assessment in Children and Adolescents.

Sports Med.

- Sirard, J. R., Trost, S. G., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2005). Calibration and Evaluation of an Objective Measure of Physical Activity in Preschool Children. *Journal of Physical Activity and Health*, 2(3), 345–357. <https://doi.org/10.1123/jpah.2.3.345>
- Timmons, B. W., LeBlanc, A. G., Carson, V., Connor Gorber, S., Dillman, C., Janssen, I., Kho, M. E., Spence, J. C., Stearns, J. A., & Tremblay, M. S. (2012). Systematic review of physical activity and health in the early years (aged 0–4 years). *Applied Physiology, Nutrition, and Metabolism*, 37(4), 773–792. <https://doi.org/10.1139/h2012-070>
- Trost, S. G. (2007). State of the Art Reviews: Measurement of Physical Activity in Children and Adolescents. *American Journal of Lifestyle Medicine*, 1(4), 299–314. <https://doi.org/10.1177/1559827607301686>
- Trost, S. G., Fees, B. S., Haar, S. J., Murray, A. D., & Crowe, L. K. (2012). Identification and Validity of Accelerometer Cut-Points for Toddlers. *Obesity*, 20(11), 2317–2319. <https://doi.org/10.1038/oby.2011.364>
- Trost, S. G., Mciver, K. L., & Pate, R. R. (2005). Conducting Accelerometer-Based Activity Assessments in Field-Based Research. *Medicine & Science in Sports & Exercise*, 37(11), S531–S543. <https://doi.org/10.1249/01.mss.0000185657.86065.98>
- Tulve, N. S., Jones, P. A., McCurdy, T., & Croghan, C. W. (2007). A Pilot Study Using an Accelerometer to Evaluate a Caregiver’s Interpretation of an Infant or Toddler’s Activity Level as Recorded in a Time Activity Diary. *Research Quarterly for Exercise & Sport*, 78(4), 375–383. trh.
- van Cauwenberghe E, Labarque V, Trost SG, de Bourdeaudhuij I, & Cardon G. (2011). Calibration and comparison of accelerometer cut points in preschool children. *International Journal of Pediatric Obesity*, 6(2–2), e582-9. rzh. <https://doi.org/10.3109/17477166.2010.526223>

CHAPTER 2: REVIEW OF LITERATURE

Early childhood (0 – 5 years old) is a period of rapid physical and cognitive development (Willumsen & Bull, 2020). Despite an increase in awareness, childhood obesity rates continue to rise during this period. Between 2017 – 2020 the Centers for Disease Control and Prevention (CDC) reported that the obesity prevalence among children and adolescents (2 – 19 years old) was 19.7%. When assessing specific age groups, in those 2 – 5 years old, the prevalence was 12.7%. Childhood obesity is associated with numerous negative health outcomes, both physical and psychological (Karnik & Kanekar, 2012). It has been linked to an increased risk of cardiovascular disease, high blood pressure, anxiety, and depression in adulthood (Karnik & Kanekar, 2012).

Obesity is, in part, due to an imbalance of calories consumed and calories expended. This can be due to an increase in consumption of calorie dense food or a decrease in physical activity (PA) levels (Centers for Disease Control and Prevention, 2022). PA is defined as any bodily movement produced by skeletal muscle that requires energy to be expended above resting values (Caspersen et al., 1985). Because PA requires an increase in energy output, it has been identified as a target strategy used for both obesity treatment and prevention. Promoting health behaviors, such as engagement in PA is particularly important during childhood because participation tracks across time, meaning more active children tend to become more active adults (Carson et al., 2019).

In 2008, the United States Department of Health and Human Services (USDHHS) released the *Physical Activity Guidelines for Americans*. These guidelines outlined the promotion of PA for positive health outcomes and included PA recommendations for adults (18 and older) and youth (6 – 17 years old). It was recommended that youth achieve 60 minutes or

more of moderate-to-vigorous PA (MVPA) daily. The 60 minutes was to be composed of aerobic, muscle strengthening, and bone strengthening activity (U.S. Department of Health and Human Service, 2008).

In 2018, the 2nd edition of the *Physical Activity Guidelines for Americans* was released (U.S. Department of Health and Human Service, 2018). This new edition additionally outlined the importance of physical activity during the preschool years (3 – 5 years old) rather than just during childhood and adolescence (6 – 17 years old). Preschool aged children are recommended to be physically active throughout the day to enhance growth and development. The guidelines suggest that although the amount of activity needed for young children is not well known, preschoolers should be encouraged to participate in both active play and structured activities for at least 180-minutes per day. This play can occur at a light, moderate, or vigorous intensity. Despite incorporating younger children, the guidelines state that evidence for children less than three years old was not reviewed.

Around the same time that the *Physical Activity Guidelines for Americans* was updated, governing bodies in Canada and Australia released 24-hour movement guidelines for children 0 – 5 years old (Okely et al., 2017; Tremblay et al., 2017). These were followed up with the World Health Organization's (WHO) *Recommendations for 24-Hour Physical Activity, Sedentary Behavior and Sleep for Children under 5 Years of Age* in 2020 (Willumsen & Bull, 2020). These guidelines highlight the importance of overall health by including recommendations for PA, sedentary time (ST), and sleep. All three sets of guidelines include specific recommendations for infants (< 1 year), toddlers (1 – 3 years), and children (3 – 4 years). For toddlers, it is suggested that they engage in at least 180-minutes of PA daily. This activity can be at any intensity (light,

moderate, or vigorous). It is also recommended that they engage in no more than 60 minutes of screen time and are not restrained for more than an hour at a time. They should also have 11 – 14 hours of good quality sleep a day. Among toddlers, meeting these guidelines has been associated with positive health outcomes such as lower BMI z-scores and higher social cognition (Cliff et al., 2017; Lee et al., 2017). Despite these benefits, reports estimate that less than 15% of toddlers are meeting all three criteria (Cliff et al., 2017; Lee et al., 2017).

When assessing if toddlers are meeting the PA guidelines, the literature presents inconclusive results. While several studies conclude that toddlers are participating in ≥ 180 -minutes of physical activity (J. Hnatiuk et al., 2012; Johansson, Hagströmer, et al., 2015; Lee et al., 2017), others conclude that they are not (Herzig et al., 2017; Vanderloo & Tucker, 2015; Wijtzes et al., 2013). However, it is difficult to compare results across studies due to the methodology applied in measuring physical activity.

Measuring Physical Activity in Toddlers

It is important that PA behaviors in toddlers are assessed accurately. PA can be assessed using a variety of methodologies, each with their own strengths and limitations (Troost, 2007). Questionnaire-based methods of PA assessment include surveys, diaries, questionnaires, or recalls. Surveys are typically low cost to develop and distribute, so they can be used for large scale population health studies. However, when using surveys to assess youth, young children are unable to accurately self-report their behavior. Proxy report by parents or teachers is not ideal because these individuals are not always observing the child's activity, leading to bias. Better methods may exist that are better able to capture PA using techniques such as device-based monitoring (e.g., pedometers and accelerometers) and direct observation (DO). Using

these methods may be more costly and time intensive for researchers, but results in more valid conclusions (Troost, 2007).

Survey Based Assessment of Physical Activity

Survey or questionnaire-based assessment of PA allows for a 24-hour snapshot of movement behavior and large-scale implementation. Surveys are typically low-cost and can capture additional information regarding PA such as type and context. In the last 20 years, many questionnaires have been developed to measure behaviors such as sleep and ST in younger populations; however, few questionnaires have been developed to assess any of these behaviors in just toddlers (Arts et al., 2022).

Several questionnaires have been developed or validated to assess PA in toddlers (Bingham et al., 2016; Bonn et al., 2012; Burdette et al., 2004; Rice et al., 2013; The TARGet Kids Collaboration et al., 2015). All surveys were developed using relatively small samples (n = 28 – 250) and according to a recent narrative review, no intercontinental surveys have been developed for those < 5 years old (Aubert et al., 2021). Although surveys include toddler-aged children, no surveys were created using only toddlers; most included children up to 5 or 6 years old (Bonn et al., 2012; Rice et al., 2013; The TARGet Kids Collaboration et al., 2015). This is problematic because toddlers exhibit unique movement patterns compared to preschoolers (Colson & Dawkins, 1997). For example, according to the National Institutes of Health, the ability to walk alone develops between 8.2 months (1st percentile) and 17.6 months (99th percentile). Only 50% of young children can walk independently at 12-months (WHO Multicentre Growth Reference Study Group & Onis, 2007). This is important as a healthy sample of preschoolers would be walking independently.

Four of the five surveys were developed to assess PA at home and rely on parent-report, while only one was developed for use at childcare centers and relies on teacher report. Burdette et al. (2004) developed a checklist and recall assessing preschoolers' outdoor PA. Bingham et al. (2016) developed the Early Years Physical Activity Questionnaire (EY-PAQ), and The TARGet Kids Collaboration et al. (2015) developed a survey based on the Canadian Health Measures Survey. Rice et al. (2013) aimed to validate modified versions of the Burdette Parent Proxy Report and the Harro Parent and Teacher Proxy Report in a younger population attending childcare centers. Measuring PA in child care settings is important because 45 percent of toddlers (1 – 2 years old) and 82 percent of preschoolers (3 – 5 years old) spend at least one day per week in a childcare setting (Corcoran & Stanley, 2017). All surveys, except one, were validated against accelerometry. Two were validated using an ActiGraph (Bingham et al., 2016; Rice et al., 2013), one used the ActiCal (The TARGet Kids Collaboration et al., 2015), and one used a RT3 Triaxial Research Tracker (Burdette et al., 2004). In all studies, the accelerometer was worn over 7 days. Various cut points were used among all studies. Two used Pate's preschool cut points (Bingham et al., 2016; Rice et al., 2013), and one each used arbitrarily developed cut points (Burdette et al., 2004), Wong and Adolph (The TARGet Kids Collaboration et al., 2015), and Costa for ST (Bingham et al., 2016). One survey was not validated, but the feasibility of the web-based questionnaire was assessed (Bonn et al., 2012).

Most surveys showed weak, but significant, correlations to accelerometry. Burdette et al. (2004) reported weak correlations between both the checklist ($r = 0.33$) and recall ($r = 0.20$) and accelerometry. Further validation found that the Burdette Proxy report was weakly correlated to total PA ($r = 0.30$) MVPA ($r = 0.34$), while the Harro Proxy report was not

associated with PA at all (Rice et al., 2013). The EY-PAQ was only valid when measuring MVPA after applying boundaries, but the correlation was also weak ($r = 0.30$). Lastly, the TARGeT Kids Collaboration et al. (2015) reported the strongest correlation ($r = 0.39$). However, they reported that significantly more total PA was reported by the accelerometer than parents (parents under reported PA by 2 hours per day). They also concluded that these findings only existed for those < 18 months of age, ruling the tool out for use in those older than 18-months. Two studies concluded that ST could not be assessed using developed questionnaires (Bingham et al., 2016; The TARGeT Kids Collaboration et al., 2015). The TARGeT Kids Collaboration et al. (2015) reported that parents under-reported sedentary time by 5-hours per day.

In two studies, reliability was also assessed (Bingham et al., 2016; Bonn et al., 2012). Bonn et al. (2012) reported that their web-based questionnaire had moderate reliability in measuring PA outdoors (ICC = 0.60) and good reliability when assessing time spent watching television (ICC = 0.85) when administered ~3 weeks apart. The EY-PAQ had much lower reliability in assessing MVPA (ICC = 0.35) and ST (ICC = 0.47) when administered ~1 week apart. However, as previously mentioned, the EY-PAQ was reported as an invalid method for measuring PA. No other studies assessed the reliability of their survey-based tool. This is problematic since a measurement tool cannot be deemed valid if it does not have reliability (American Educational Research Association, 1985; Washburn et al., 2000).

Many limitations exist when using survey-based PA assessments. None of the studies reported stronger than a weak correlation between survey-based assessment and accelerometry, and these studies all used small, homogenous samples. No studies were multi-cultural or were assessed internationally. Using accelerometry as a comparison measure

introduces additional limitations. All studies using accelerometry as the comparison reported the inability for the device to capture increased PA intensity due to activities (e.g. climbing, walking upstairs, or riding wheel toys), the inability for the device to detect posture changes (such as sitting v. standing), and the removal of the device for activities such as swimming or other water play as limitations. These factors play a significant role in the misclassification of PA.

In summary, there is a lack of survey-based assessments for measuring PA in toddlers, and those that have been developed show weak correlation to accelerometry. Survey-based assessments that have been developed have not yet been implemented in larger, more diverse samples, and most studies failed to report on reliability and feasibility of using each tool in their target population. Lastly, Arts et al. (2022) raises the question of the ability of proxy-reports to measure PA in younger populations due to their short bursts of activity and the inability for parents or caregivers to accurately report this type of movement.

Direct Observation for Assessing Physical Activity

DO is considered the criterion method of assessing PA, especially in younger populations (Troost, 2007). When using DO, an observer categorizes an individual's behavior based on codes outlined by a specific tool. These categories include not only PA intensity, but often contextual information as well. DO can account for the environment in which activity is occurring. For example, researchers can identify the objects with which children are engaging or the location where they are engaging in activity. While no DO system has been specifically developed for use in toddlers, several tools that have been developed to assess PA in older children. These include the Children's Physical Activity Form (CPAF) (O'hara et al., 1989), the Children's Activity

Rating Scale (CARS) (Puhl et al., 1990) the Behaviors of Eating and Activity for Children's Health Evaluation System (BEACHES) (McKenzie, Sallis, Nader, et al., 1991), the System for Observing Fitness Instruction Time (SOFIT) (McKenzie, Sallis, & Nader, 1991), the System for Observing Play and Leisure Activity in Youth (SOPLAY) (McKenzie et al., 2002), and the Observational System for Recording Physical Activity in Children – P (OSRAC-P) (Brown et al., 2006). Despite being observational tools used for PA assessment, all tools contain different components.

Early tools, such as CPAF and CARS, were only developed to categorize activity intensity. CARS was developed using a group of children aged 3 – 4 years, while CPAF was developed in a group of 8-10-year-olds (O'hara et al., 1989; Puhl et al., 1990). CARS includes five PA intensity categories including 1-stationary, no movement, 2-stationary, with movement, 3-translocation, slow/easy, 4-translocation, medium/moderate, and 5-translocation, fast/strenuous. The CPAF includes overlapping categories, but only four rather than five. The first two categories are the same as in CARS, but the last two include 3-slow trunk movement and 4-rapid trunk movement. The CPAF was created for use with the physical education setting, while CARS was created for use in all settings, but both tools were validated against heart rate monitoring (O'hara et al., 1989; Puhl et al., 1990). Although helpful in describing activity level, these tools lack the ability to describe other important PA variables. The remaining observational tools include coding schemes beyond activity intensity.

BEACHES, SOFIT, and SOPLAY were developed and validated by McKenzie and colleagues (1991, 1992, 2002). Because they were intended for use in different settings, the three tools are comprised of different content, but all three contain similar PA intensity scales. BEACHES was developed to assess PA, eating, and the environment simultaneously (McKenzie,

Sallis, Nader, et al., 1991). It is intended for use in the school or home environment, and it was developed in a sample of 4-8-year-olds. This tool contains ten categories including environment (e.g. alone, with parent, tv), physical location (e.g. inside/outside at home or school, cafeteria, playground), activity intensity (1-lying down, 2-sitting, 3-standing, 4-walking, 5-very active), eating behavior (no food ingestion or food ingestion), with whom child participates in activity (interactor), stimulus to increase or decrease PA (antecedents), prompted event, child response to prompt, consequences, and events receiving consequences. SOFIT was adapted for use in physical education and was intended to account for instructional factors of PA (McKenzie, Sallis, & Nader, 1991). An older population of 3rd – 5th graders were assessed for the SOFIT. In addition to intensity, SOFIT includes a category for education lesson context (general or subject matter) and for teacher involvement. The last tool, SOPLAY, was adapted in a group of 6th – 8th graders for use during recreational and leisure PA rather than just structured activity (e.g. PE). Unlike other tools, SOPLAY is intended for group level assessment rather than child-by-child (McKenzie et al., 2002). Like BEACHES, this tool contains a larger number of categories including PA level (same scale as previous two tools), time of day, temperature, area accessibility, area usability, presence of supervision, classification and presence of organized activity, and equipment availability. During development, children were observed before school, starting at 8am, during lunch, and after school, until 4:15pm.

All observational tools, except for CARS and BEACHES, were developed in youth older than 5-years and do not include information regarding environmental and social context. The Observational System for Recording Physical activity in Children – Preschool was developed to describe these additional factors (Brown et al., 2006). Experts developed the OSRAC utilizing

literature and prior observational tools for children. This tool adapted the PA categories from CARS and includes 5 activity levels (1-stationary or motionless, 2-stationary with limb or trunk movement, 3-slow/easy movement, 4-moderate movement, 5-fast movement). Instead of being limited to intensity assessment, the OSRAC-P includes categories to provide information about what activity is being performed (type), the physical environment (e.g., location & activity context), and the social environment (e.g., group composition, initiator, & prompts). The categories are partially adapted from the Code for Active Student Participation and Engagement–Revised (CASPER-II), which is an ecobehavioral tool used to assess preschool classrooms and student engagement, but not PA levels (Brown et al., 1999). This provides information regarding contextual and behavioral aspects of PA. This additional information that can be collected is outlined in Table 1.

After development, researchers implemented an evaluation phase in three preschools. The observation intervals contained a 5-second observation period followed by 25-seconds of coding and each child was observed for 5 hours. Of these intervals, thirteen percent of intervals were assessed by two independent researchers and good reliability was reported ($k = 0.79 - 0.83$; % agreement = 89 – 100%). The OSRAC-P is considered a valid and reliable tool for assessing PA. Despite the number of tools created for children, no tool has been specifically developed for toddlers.

Table 1. Categories and Codes Outlined in the OSRAC-P

Category	Code	Category	Code
<i>Activity Level</i>	1-Stationary 2-Limbs 3-Slow-Easy 4-Moderate 5-Fast Can't Tell	<i>Indoor Context</i>	Art Books/Preacademic Gross Motor Group Time Large Blocks Manipulative Music Nap Self-Care Snacks Sociodramatic Teacher Arranged Time Out Transition Videos Other N/A Can't Tell
<i>Activity Type</i>	Climb Crawl Dance Jump/Skip Lie Down Pull/Push Rough & Tumble Ride Rock Roll Run Sit/Squat Stand Swim Swing Throw Walk Other Can't Tell	<i>Outdoor Context</i>	Ball Fixed Game OpenSpace Pool Portable Sandbox Snacks Sociodramatic Props Teacher Arranged Time Out Wheel Other N/A Can't Tell
<i>Location</i>	Indoor Outdoor Transition Can't Tell	<i>Prompt</i>	None Teacher Prompt-Increase PA Teacher Prompt-Decrease PA Peer Prompt-Increase PA Peer Prompt-Decrease PA
<i>Group Composition</i>	Solitary 1-1 Adult 1-1 Peer Group Adult Group Can't Tell		
<i>Initiator of Activity</i>	Adult Child Can't Tell		

Direct Observation for Assessing Physical Activity in Toddlers. Few studies have used DO to assess PA in toddlers (Dinkel et al., 2019; Fees et al., 2015; Gubbels et al., 2011; Specker et al., 1999; Van Cauwenberghe et al., 2011). All studies were conducted at childcare centers and three used the OSRAC-P as their observational tool. One study utilized the CARS system to evaluate PA levels (Specker et al., 1999). Toddlers aged 12 – 36 months were included in the studies, with one study also including infants (6 – 12 months). Samples ranged from 31 – 175 toddlers aged 20.0 – 31.2 months; however, one study did not report specific sample size or characteristics (Dinkel et al., 2019). Two studies included both indoor and outdoor observations (Gubbels et al., 2011; Van Cauwenberghe et al., 2011); one was conducted only indoor (Fees et al., 2015), and one only outdoor (Dinkel et al., 2019)

Rather than using in-person DO, all but one study videorecorded observations, and coding was done later using video software (Dinkel et al., 2019; Fees et al., 2015; Van Cauwenberghe et al., 2011). Toddlers were observed for 5 – 15 second intervals followed by 25 – 30 seconds of recording for each category simultaneously. Observational intervals of five-second were used by Fees et al (2015) and Dinkel et al (2019), while 15-second intervals were used by Gubbels et al (2011) and Van Cauwenberghe et al (2011), contributing to a total of 1,640, 6,046, 1,382 and 4,218 observations, respectively. All studies assessed PA intensity and location, three assessed activity type, and prompts, and two assessed activity contexts, group composition, and initiators.

When assessing PA intensity, all studies reported higher amounts of time spent engaging in sedentary time (1-Stationary + 2-Limbs) indoors compared to outdoors (59.4 – 74% v. 31.2 – 41.9%). In line with this, more time was spent engaging in MVPA (4-Moderate + 5-Fast)

outdoors compared to indoors (13.3 – 21.3% v. 5.5 – 6%). However, one study that included indoor and outdoor observations did not compare PA intensities between the two settings (Van Cauwenberghe et al., 2011). Another study included slow activity with MVPA in a category called “active play” which resulted in higher reports of activity (Dinkel et al., 2019). Only two studies reported average intensity from the OSRAC-P using the 1 - 5 scale discussed previously. One reported average intensities ranging from 2.36 indoor to 2.82 outdoor (Gubbels et al., 2011). The other study reported an overall intensity of 2.6 ± 0.9 during both indoor and outdoor activity (Van Cauwenberghe et al., 2011).

When assessing the activity type and context, similar conclusions were reported across studies. All three studies reported sitting/squatting (24.3 – 25.4%), standing (20.5 – 36.0%), and walking (18.2 – 33.0%) as contributing to the greatest proportion of observed activity type (Dinkel et al., 2019; Fees et al., 2015; Van Cauwenberghe et al., 2011). Most activities across intervals were child initiated (87.0 – 91.2%) (Dinkel et al., 2019; Fees et al., 2015) and done without a prompt from staff or peers (88.0 – 94.0%) (Dinkel et al., 2019; Fees et al., 2015; Gubbels et al., 2011). These trends were seen across indoor and outdoor activities. When assessing activity contexts and group composition, the most common codes were different across settings. Indoors, the most common activity contexts were manipulative experiences (16%) and self-care (15%), while outdoors they were open space (43.2%) and fixed equipment (37.6%) (Dinkel et al., 2019; Fees et al., 2015). When assessing group composition, children were most likely to perform an activity together with their peers indoors. Outdoors, one study concluded toddlers were most likely to engage in activity alone (37.7%) or with several peers (41.2%) (Gubbels et al., 2011), while a second study concluded toddlers were likely to engage in

activity with an adult present (49.0%) or alone (20.0%) (Dinkel et al., 2019). Although PA behaviors of toddlers have been explored, the relationship between these variables is more important when informing promotion of physical activity.

Personal, Social, and Environmental Factors. It has been well documented that activity can be heavily influenced by the environment (physical or social) in preschoolers (Brown et al., 2009; Clevenger et al., 2020, 2021, 2022; Pate et al., 2008). In a recent systematic review, Clevenger et al. (2020) aimed to assess preschoolers' physical activity by schoolyard location. Based on numerous studies (n=24), physical activity participation was found to vary based on outdoor location. Authors reported that the highest levels of PA were elicited in open spaces and that spaces with tracks for using wheel toys, and various pieces of fixed equipment promoted activity engagement (Clevenger et al., 2020). A study assessing PA, activity type, and indoor location, concluded that factors within the classroom also elicit differences in activity levels (Clevenger et al., 2022). For example, the loft area location and gross motor activity type were associated with the highest level of PA. However, the intensity of gross motor activity varied based on location (25.3 counts/s in the manipulative location v. 67.0 counts/s in the blocks location). A similar study conducted outdoors reported that preschoolers spent most of their time in open spaces, but that different activity types were occurring (e.g., open space, socioprops, ball or object, etc.) that all elicited various levels of activity (40.9, 41.4, 49.7 counts/s respectively) (Clevenger et al., 2021). These relationships have been less studied among toddlers.

Few studies have assessed the relationships between PA intensity and the other social or environmental categories (Dinkel et al., 2019; Fees et al., 2015; Gubbels et al., 2011).

Gubbels et al. (2011) assessed the relationships between prompts and group size and activity intensity and stratified those relationships by gender, age, and location. The authors concluded that a larger group size (staff or peers) was associated with lower activity intensities indoor and outdoor, but that only 2-year-olds had lower activity intensities when more staff members were present. Positive prompts were associated with decreases in activity intensity indoors for girls, while being associated with increases in activity intensity outdoors for boys. Dinkel et al. (2019) assessed the relationship between outdoor area and intensity. They concluded that in fine motor areas, toddlers were most sedentary and commonly participated in activities such as sitting/squatting. In open space and gross motor areas toddlers were most active and commonly participated in running, walking, and standing. However, these conclusions varied based on childcare centers. Lastly, Fees et al. (2015) assessed the relationships between activity type and intensity and concluded that toddlers who engaged in onlooking had decreased odds of participating in MVPA. Assessing these relationships can help inform future interventions aimed at increasing PA levels in toddlers.

Adaptations of the OSRAC-P for Assessing Physical Activity. Two of these studies proposed that the OSRAC-P was not suitable for use in toddlers and made adaptations to categories and codes (Dinkel et al., 2019; Fees et al., 2015). Although not toddler specific, Dinkel et al. (2019) created an additional category called the play area category that aimed to identify between play areas in the outdoor play space. The areas were separated based on the function of equipment in that area and include the following codes: open spaces, gross motor play, and fine motor play. A second study by Fees et al. (2015) added nine activity codes and 3 indoor activity context codes. Finally, despite not utilizing the entirety of the OSRAC-P, when

assessing infant/toddler behavior a weight-bearing category was added in one study. These changes/additions are explained further in Table 2.

Table 2. Categories and Codes Added to OSRAC-P for Toddler Activity Assessment

Author	Added Category or Code	Description
Fees et al. (2015)	Activity Type	
	<i>Bend</i>	Bending at waistline
	<i>Bounce</i>	Flex knees or torso while feet remain on floor
	<i>Carried/Held</i>	Carried or held by adult
	<i>Creep</i>	Push self with stomach on floor
	<i>Fall Down</i>	Falling, tripping, or stumbling
	<i>Hanging</i>	Removing body weight from floor
	<i>Hesitation</i>	Pause before engaging
	<i>Kicking</i>	1 leg swings to move object or air
	<i>Scoot</i>	Dragging body across surface while lying or sitting
	Indoor Context	
	<i>Conversing</i>	Talking to peer or adult
	<i>Tantrum</i>	Engaging in emotional outburst; disengaged
	<i>Onlooking</i>	Watching others play – does not enter play
Dinkel et al. (2019)	Outdoor Play Areas	
	<i>Open Space</i>	Green spaces and sidewalks
	<i>Gross Motor</i>	Large play structures that allow for movements such as running or walking
	<i>Fine Motor</i>	Small equipment used for fine motor skills such as grasping or pulling
Specker et al. (1999)	Weight-Bearing	
	<i>Time weight-bearing</i>	% time bearing weight on legs

Since development, the OSRAC-P has been adapted into four additional versions, the Observational System for Recording Physical Activity in Children: Home (OSRAC-H) (McIver et al., 2009), the Observational System for Recording Physical Activity in Children: Youth Sport (OSRAC-YS) (Cohen et al., 2014), the Observational System for Recording Physical Activity in Children: Elementary School (OSRAC-E) ((McIver et al., 2016), and the Observational System for

Recording Physical Activity in Children: Developmental Disabilities (OSRAC-DD) (Schenkelberg et al., 2021). Two adaptations were done using a similar preschooler population (McIver et al., 2009; Schenkelberg et al., 2021) and two were done using youth K – 5th grade (Cohen et al., 2014; McIver et al., 2016). All adaptations included an instrument development and an instrument evaluation phase.

During the development phase, researchers conducted literature reviews in addition to interviews with experts in each population. This knowledge acquisition led to the addition of new categories or new codes for pre-existing categories. For example, during the OSRAC-DD development, categories were added for Repetitive Behavior/Stereotypes and Interactions (Schenkelberg et al., 2021). Additional codes, such as therapy and therapist initiated were added to Indoor Education/Play Contexts and Activity Initiator categories, respectively. After the addition of updated content, authors stated that extensive field testing was done using each tool to ensure all behaviors in the target population were captured. However, authors did not elaborate upon the specifics of field testing, such as time spent observing or number of children observed. If needed, following field testing, the tool was updated to reflect missing content. Following the development phase, there was an evaluation phase.

The evaluation or validation phase included final samples ranging from 19 – 71 children. Of the four studies, three included 19 – 25 children (Cohen et al., 2014; McIver et al., 2009; Schenkelberg et al., 2021). All but one study used 30-second observation intervals composed of a 5-second observation followed by 25-seconds of recording. The fourth study used a 20-second observation interval with 10-seconds of observing and 10-seconds of recording (Cohen et al., 2014). When recording behavior, the highest intensity was coded and then the remaining

categories were coded for the activity being done. For example, if the focal child was sitting with no limb movement and then stood up and began running, the intensity for the interval would be coded as 5-Fast. The remaining categories would then be coded in line with that level 5 intensity. Total observation intervals averaged 6,055 across all four studies but ranged from 580 in the youth sport adaptation to 11,360 in the elementary student adaptation. An average of 38% (20 – 66%) of observational intervals were coded by at least two researchers independently to assess reliability.

Interobserver correlation coefficients (ICCs) and interobserver agreements (IOAs) were calculated to assess reliability between coders. In all adaptations, similar ICCs and IOAs were reported. Most ICCs for each category were above 0.70, indicating moderate-to-good reliability. Similar trends were seen when assessing agreement, with most categories resulting in IOAs greater than 80%, indicating again moderate-to-good reliability. All adaptations of the OSRAC-P were concluded to be reliable for PA assessment in their target populations.

DO is considered the gold standard for measuring physical activity in children. Utilizing DO is the best way to gain information regarding behaviors other than intensity, such as social and environmental contexts. Very few studies have used this measurement method to assess physical activity levels in toddlers, and those who have reported that no existing observational tool is fully able to describe their behavior. One reason may be because DO requires extensive training and is very time-consuming for the researchers (Troost, 2007). It can also lead to reactivity if in-person observation is occurring (Troost, 2007). Additionally, the sporadic nature of child's PA is difficult to capture using DO, and in respect to the OSRAC systems, only one child

can be assessed at a time. Device-based methods are better able to capture these sporadic movement patterns and can be used to assess multiple children at once.

Device-based Methods of Physical Activity Assessment

Accelerometers are small devices that measure movement, and most commercially available devices now assess in multiple planes rather than a single plane. Accelerometers can quantify both magnitude and direction of accelerations (Troost, 2007). Different techniques can then be applied to determine PA intensity. Most commonly, researchers have used cut-points to quantify PA. However, the same cut-points cannot always be applied across studies due to differences in population, monitor settings (e.g., placement), or processing (e.g., filtering).

Toddler Physical Activity: Cut-Point Development. Several sets of cut-points have been calibrated and validated in toddlers using cross-validation (Costa et al., 2014; Hager et al., 2016; Johansson, Ekelund, et al., 2015; Kelly & Villalpando, 2016; Oftedal et al., 2014; Pulakka et al., 2013; Trost et al., 2012). However, only four of these studies used an independent sample to cross-validate cut-points (Costa et al., 2014; Hager et al., 2016; Oftedal et al., 2014; Trost et al., 2012) while three used the same sample (Johansson, Ekelund, et al., 2015; Kelly & Villalpando, 2016; Pulakka et al., 2013). Independent sample cross-validation is recommended to optimize generalization (Browne, 2000). Using a separate validation sample can determine if issues in applying the model will result due to biases such as selection bias. These same assumptions of generalizability cannot be made using the same sample for both calibration and validation (Browne, 2000).

All studies were conducted in toddlers with no limitations to active play, except for one that included toddlers with cerebral palsy (Oftedal et al., 2014). Studies include toddlers aged 6

– 36 months, however one study included children up to 47 months (Costa et al., 2014).

Calibration samples ranged from 18 – 65 toddlers, while validation samples ranged from 12 – 191 toddlers.

Four studies utilized uniaxial accelerometers (Hager et al., 2016; Kelly & Villalpando, 2016; Oftedal et al., 2014; Trost et al., 2012), while the rest used triaxial accelerometers (ActiGraph GT3X or GT3X+) (Costa et al., 2014; Johansson, Hagströmer, et al., 2015; Oftedal et al., 2014; Pulakka et al., 2013). One study compared the use of uniaxial devices to that of triaxial (Oftedal et al., 2014). Authors concluded that triaxial devices should be used over uniaxial devices. Devices were commonly placed on the waist at the right hip (Costa et al., 2014; Kelly & Villalpando, 2016; Pulakka et al., 2013; Trost et al., 2012). A single study was conducted with accelerometer placement at each of the following locations: non-dominant wrist (Johansson, Ekelund, et al., 2015), non-dominant ankle (Hager et al., 2016), and lower back (Oftedal et al., 2014). Four studies applied 15-second epochs (Costa et al., 2014; Kelly & Villalpando, 2016; Pulakka et al., 2013; Trost et al., 2012), three applied 5-second epochs (Costa et al., 2014; Johansson, Ekelund, et al., 2015; Oftedal et al., 2014), and one used 30-seconds (Hager et al., 2016). Costa et al. (2014) compared 15-second to 5-second epochs and reported that using 15-seconds resulted in an underestimated of sedentary time (8.7 – 24.2%) and an overestimation of total physical activity (2.8 – 9.2%) when compared to CARS using second-by-second coding. Therefore, they concluded that 5-second epochs should be used over 15-seconds in this age group.

All cut-points were validated using DO as the criterion measure. Most calibration studies utilized a 15 – 30-minute videorecorded session that was either semi-structured or contained

active free play. The video recorded sessions were then assessed using DO. DO was done using a variation of CARS in all except two studies, which utilized CPAF (Kelly & Villalpando, 2016; Pulakka et al., 2013). Only two studies required a free-living validation component that required toddlers to wear a device across seven days (Hager et al., 2016; Pulakka et al., 2013). In one study, the same sample of toddlers was used for validation as calibration (Pulakka et al., 2013), while in the others, an additional sample of toddlers was recruited (Hager et al., 2016). One study did not report additional validity of their cut points (Kelly & Villalpando, 2016). Details regarding the identified cut points can be found in Table 3.

Table 3. Validated Toddler Cut points

Author (Year)	Age (mo) (average \pm SD)	Device	Device Placement	Epoch	Axis	Derived Cut points	
Trost (2012)	25.2 \pm 4.8	ActiGraph GT1M	Right Hip	15s	NR	SB	\leq 48
						MVPA	$>$ 418
Pulakka (2013)	17.0	ActiGraph GT3X	Right Hip	15s	VM	SB	$<$ 208
						MVPA	\geq 208
					V-Axis	SB	$<$ 35
						MVPA	\geq 35
Costa (2014)	33.5 \pm 6.6	ActiGraph GT3X+	Right Hip	5s	Axis 1	SB	\leq 5
						MVPA	\geq 165
Ofstedal (2014)*	24 \pm 6	ActiGraph GT3X & GT3X+	Lower Back	5s	VM	SB	\leq 40
						TPA	$>$ 40
Johansson (2015)	26 \pm 6.0	ActiGraph GT3X+	Non-dominant wrist	5s	Y-Axis	SB	\leq 89
						MVPA	\geq 440
Hager (2016)	24.5	Actical	Non-dominant ankle	30s	NR	SB	\leq 20
						MVPA	\geq 1101
Kelly (2016)	19.5 \pm 5.93	ActiGraph GT1M	Right Hip	15s	NR	SB	\leq 181
						MVPA	\geq 435

NR: Not Reported, SB: Sedentary time, MVPA: Moderate-to-Vigorous Activity, *children with cerebral palsy

Accelerometers to Assess Physical Activity in Toddlers. Many studies have assessed toddler PA levels using accelerometry (n=26; Armstrong et al., 2019; Bisson et al., 2019; Borkhoff et al., 2015; Dlugonski et al., 2017; Felzer-Kim & Hauck, 2020; Hager et al., 2016; Hauck & Felzer-Kim, 2019; Herzig et al., 2017; J. Hnatiuk et al., 2012; J. A. Hnatiuk et al., 2017; Johansson, Hagströmer, et al., 2015; Kelly et al., 2022; Konstabel et al., 2014; Kwon et al., 2019; Lee et al., 2017; McCullough et al., 2018; Oftedal et al., 2014; Orlando et al., 2019; Pereira et al., 2021; Pulakka et al., 2017; Santos et al., 2017; Tanaka & Kuroda, 2022; Taylor et al., 2018; Vanderloo & Tucker, 2015; Wijtzes et al., 2013). ActiGraphs (GT1M, GT3X, or GT3X+) were the most used accelerometer (n=19), and devices were most commonly placed at the right hip (n=18).

In one study, a wrist-worn ActiGraph wGT3X was assessed in addition to a hip-worn ActiGraph (Kwon et al., 2019). Authors concluded that the waist was a more feasible wear location in toddlers, but due to no wrist cut-points being available at the time the study was conducted, they were not able to compare PA time between locations. Authors also concluded that they observed toddlers frequently adjust wrist-worn devices or are holding on to things like railings which could result in additional noise in this data (Kwon et al., 2019). This is interesting because wrist placement has been found to increase compliance in older children (Fairclough et al., 2016; Nyberg et al., 2009) and is also good for capturing 24-hour movement behaviors (Lettink et al., 2022). Epoch lengths varied across studies, but 15-seconds was utilized most frequently (n=15). Again, this is interesting because it has been suggested that smaller epoch lengths should be used to capture sporadic movement in young children (Aibar & Chanal, 2015; Sanders et al., 2014). Trost et al. (2012) cut points were applied in many studies (n=13), followed by Hager (2016; n=2) and Johansson (2013; n=2). However, in the latter studies, cut

points were utilized in future studies by the authors who calibrated them. Several studies have applied preschool cut-points to toddler populations (Borkhoff et al., 2015; Herzig et al., 2017). More information regarding accelerometer characteristics and applied cut-points can be found in Table 4.

Table 4. Summary of Accelerometer Characteristics (N=26)

Accelerometer Characteristic	Number of Studies
Brand	
ActiGraph	19
Actical	5
MTN300VVHO	1
Sapphire Sensor	1
Epochs	
≤ 5-s	4
15-s	15
30-s	1
60-s	3
NR	3
Location	
Hip/Waist	18
Lower Back	2
Non-dominant ankle	2
Non-dominant wrist	2
Thigh	1
Multiple Locations	1
Cut-points Used	
Adolph (2012)	1
Butte (2014) et al.	1
Evenson (2008) et al.	1
Hager (2013)et al.	2
Johansson (2015) et al.	2
Kelly (2016)et al.	1
Meredith-Jones (2021)et al.	1
Oftedal (2014)et al.	1
Sirad (2005)et al.	1
Trost (2012)et al.	132
Not Reported	2

Due to the range of accelerometer characteristics across studies, it can be difficult to compare accelerometer-based studies. A recent systematic review and meta-analysis aimed to examine toddlers' PA behavior across daytime hours measured by accelerometry (Bruijns et al., 2020). Their analysis included 20 studies that assessed toddler PA or sedentary time using accelerometry. The review concluded that toddlers engaged in 72.9 – 636.5 min/day of total PA, 48.5 – 582.4 min/day of light PA, 6.5 – 89.9 min/day of MVPA, and 172.7 – 545.0 min/day of ST. They then adjusted for different accelerometer characteristics such as accelerometer placement, cut-point validity, device type, and epoch length and reported adjusted times spent in each activity intensity. Adjusted values were reported for total PA (246.19 min/day), light PA (194.10 min/day), MVPA (60.16 min/day), and ST (337.04 min/day). After adjusting, it appears that toddlers are meeting the recommended 180 min/day of PA. However, the individually reported levels should not be disregarded.

Comparable results were reported in subsequent studies (Felzer-Kim & Hauck, 2020; Kelly et al., 2022; Orlando et al., 2019; Pereira et al., 2021; Tanaka & Kuroda, 2022). Felzer-Kim & Hauck (2020) concluded that toddlers, 1 – 3 years old, spend 31.2 ± 10.3 min/day, 55.5 ± 18.9 min/day, 464.2 ± 84.0 min/day in light PA, MVPA, and ST, respectively. Kelly et al. (2022) concluded that toddlers spend $1.8 \pm 2.6\%$, $41.7 \pm 7.0\%$, and $56.2 \pm 6.6\%$ of time in light PA, MVPA, and ST, respectively. Pereira et al. (2021) reported comparable results regarding sedentary time with 48.7% of time spent sedentary. Lastly, Tanaka et al. (2022) concluded that toddlers reach the guideline of 180-minutes/day of total PA during the day at childcare. These studies further support the argument that accelerometer-based methods are difficult to

compare. Parameters such as different epochs, wear locations, data axes, or wear time can heavily influence reported PA levels in toddlers.

This review and additional literature suggest that when assessed via accelerometry, toddlers are, on average, meeting the PA guideline of 180-minutes per day of total PA. However, given the vast differences in accelerometer characteristics adopted for each study, it is hard to directly compare results. Few studies have aimed to determine the best device settings to be used in toddlers, such as which epoch length or device location is best for measurement. Additionally, in most cases, toddlers are spending 50% of their time sedentary and several studies reported low levels of PA at higher intensities (Borkhoff et al., 2015; Taylor et al., 2018; Vanderloo & Tucker, 2015). Additionally, with device-based PA assessment, we cannot determine in which activities toddlers are engaging or the environmental or social factors eliciting these higher intensities of exercise. DO is an alternative method of physical activity assessment that can identify these key factors that can inform methods to elicit these higher intensities.

Summary

Obesity rates in early life (0 – 4 years) continue to increase, despite the well-established health benefits of habitual PA, which is a mitigating factor (Timmons et al., 2012). In 2020, *Recommendations for 24-Hour Physical Activity, Sedentary Behavior and Sleep for Children under 5 Years of Age* were the first global guidelines to suggest that toddlers (1 – 2 years old) should engage in 180-minutes of PA at any intensity, daily (Willumsen & Bull, 2020). It is important to have a PA measure that can capture the sporadic nature of children’s play to ensure these guidelines are being met. DO is promising for this, but a system relevant to

toddlers needs to be developed. Additionally, accelerometry is promising, but which existing method is best needs to be identified to enhance comparability across future studies and to be used in surveillance.

REFERENCES

- Aibar, A., & Chanal, J. (2015). Physical Education: The Effect of Epoch Lengths on Children's Physical Activity in a Structured Context. *PLOS ONE*, *10*(4), e0121238. <https://doi.org/10.1371/journal.pone.0121238>
- Armstrong, B., Covington, L. B., Unick, G. J., & Black, M. M. (2019). Featured Article: Bidirectional Effects of Sleep and Sedentary Behavior Among Toddlers: A Dynamic Multilevel Modeling Approach. *Journal of Pediatric Psychology*, *44*(3), 275–285. <https://doi.org/10.1093/jpepsy/jsy089>
- Arts, J., Gubbels, J. S., Verhoeff, A. P., Chinapaw, Mai. J. M., Lettink, A., & Altenburg, T. M. (2022). A systematic review of proxy-report questionnaires assessing physical activity, sedentary behavior and/or sleep in young children (aged 0–5 years). *International Journal of Behavioral Nutrition and Physical Activity*, *19*(1), 18. <https://doi.org/10.1186/s12966-022-01251-x>
- Aubert, S., Brazo-Sayavera, J., González, S. A., Janssen, I., Manyanga, T., Oyeyemi, A. L., Picard, P., Sherar, L. B., Turner, E., & Tremblay, M. S. (2021). Global prevalence of physical activity for children and adolescents; inconsistencies, research gaps, and recommendations: A narrative review. *International Journal of Behavioral Nutrition and Physical Activity*, *18*(1), 81. <https://doi.org/10.1186/s12966-021-01155-2>
- Bingham, D., Collings, P., Clemes, S., Costa, S., Santorelli, G., Griffiths, P., & Barber, S. (2016). Reliability and Validity of the Early Years Physical Activity Questionnaire (EY-PAQ). *Sports*, *4*(2), 30. <https://doi.org/10.3390/sports4020030>
- Bisson, M., Tremblay, F., Pronovost, E., Julien, A.-S., & Marc, I. (2019). Accelerometry to measure physical activity in toddlers: Determination of wear time requirements for a reliable estimate of physical activity. *Journal of Sports Sciences*, *37*(3), 298–305. <https://doi.org/10.1080/02640414.2018.1499391>
- Bonn, S. E., Surkan, P. J., Trolle Lagerros, Y., & Bälter, K. (2012). Feasibility of A Novel Web-Based Physical Activity Questionnaire for Young Children. *Pediatric Reports*, *4*(4), e37. <https://doi.org/10.4081/pr.2012.e37>
- Borkhoff, C. M., Heale, L. D., Anderson, L. N., Tremblay, M. S., Maguire, J. L., Parkin, P. C., & Birken, C. S. (2015). Objectively measured physical activity of young Canadian children using accelerometry. *Applied Physiology, Nutrition, and Metabolism*, *40*(12), 1302–1308. <https://doi.org/10.1139/apnm-2015-0164>
- Brown, W. H., Odom, S. L., Shouming Li, & Zercher, C. (1999). Ecobehavioral Assessment in Early Childhood Programs: A Portrait of Preschool Inclusion. *The Journal of Special Education*, *33*(3), 138–153. <https://doi.org/10.1177/002246699903300302>

- Brown, W. H., Pfeiffer, K. A., McIver, K. L., Dowda, M., Addy, C. L., & Pate, R. R. (2009). Social and Environmental Factors Associated With Preschoolers' Nonsedentary Physical Activity. *Child Development, 80*(1), 45–58. <https://doi.org/10.1111/j.1467-8624.2008.01245.x>
- Brown, W. H., Pfeiffer, K. A., McIver, K. L., Dowda, M., Almeida, J. M. C. A., & Pate, R. R. (2006). Assessing Preschool Children's Physical Activity: The Observational System for Recording Physical Activity in Children-Preschool Version. *Research Quarterly for Exercise and Sport, 77*(2), 167–176. <https://doi.org/10.1080/02701367.2006.10599351>
- Browne, M. W. (2000). Cross-Validation Methods. *Journal of Mathematical Psychology, 44*(1), 108–132. <https://doi.org/10.1006/jmps.1999.1279>
- Bruijns, B. A., Truelove, S., Johnson, A. M., Gilliland, J., & Tucker, P. (2020). 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, 17*(1), 14. <https://doi.org/10.1186/s12966-020-0912-4>
- Burdette, H. L., Whitaker, R. C., & Daniels, S. R. (2004). Parental Report of Outdoor Playtime as a Measure of Physical Activity in Preschool-aged Children. *Archives of Pediatrics & Adolescent Medicine, 158*(4), 353. <https://doi.org/10.1001/archpedi.158.4.353>
- Cardon, G., Van Cauwenberghe, E., & De Bourdeaudhuij, I. (2011). What do we know about physical activity in infants and toddlers: A review of the literature and future research directions. *Science & Sports, 26*(3), 127–130. <https://doi.org/10.1016/j.scispo.2011.01.005>
- Carson, V., Lee, E.-Y., Hesketh, K. D., Hunter, S., Kuzik, N., Predy, M., Rhodes, R. E., Rinaldi, C. M., Spence, J. C., & Hinkley, T. (2019). Physical activity and sedentary behavior across three time-points and associations with social skills in early childhood. *BMC Public Health, 19*(1), 27. <https://doi.org/10.1186/s12889-018-6381-x>
- Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). *Physical Activity, Exercise, and Physical Fitness: Definitions and Distinctions for Health-Related Research*.
- Centers for Disease Control and Prevention. (2022). *Prevalence of Childhood Obesity in the United States*. U.S. Department of Health and Human Service. <https://www.cdc.gov/obesity/data/childhood.html>
- Clevenger, K. A., Erickson, K. T., Grady, S. C., & Pfeiffer, K. A. (2021). Characterizing preschooler's outdoor physical activity: The comparability of schoolyard location- and activity type-based approaches. *Early Childhood Research Quarterly, 56*, 139–148. <https://doi.org/10.1016/j.ecresq.2021.03.012>

- Clevenger, K. A., McKee, K. L., & Pfeiffer, K. A. (2022). Classroom Location, Activity Type, and Physical Activity During Preschool Children's Indoor Free-Play. *Early Childhood Education Journal*, 50(3), 425–434. <https://doi.org/10.1007/s10643-021-01164-7>
- Clevenger, K. A., Wierenga, M. J., Howe, C. A., & Pfeiffer, K. A. (2020). A Systematic Review of Child and Adolescent Physical Activity by Schoolyard Location. *Kinesiology Review*, 9(2), 147–158. <https://doi.org/10.1123/kr.2019-0009>
- Cliff, D. P., McNeill, J., Vella, S. A., Howard, S. J., Santos, R., Batterham, M., Melhuish, E., Okely, A. D., & de Rosnay, M. (2017). Adherence to 24-Hour Movement Guidelines for the Early Years and associations with social-cognitive development among Australian preschool children. *BMC Public Health*, 17(S5), 857. <https://doi.org/10.1186/s12889-017-4858-7>
- Cliff, D. P., Reilly, J. J., & Okely, A. D. (2009). Methodological considerations in using accelerometers to assess habitual physical activity in children aged 0–5 years. *Journal of Science and Medicine in Sport*, 12(5), 557–567. <https://doi.org/10.1016/j.jsams.2008.10.008>
- Cohen, A., McDonald, S., McIver, K., Pate, R., & Trost, S. (2014). Assessing Physical Activity During Youth Sport: The Observational System for Recording Activity in Children: Youth Sports. *Pediatric Exercise Science*, 26(2), 203–209. <https://doi.org/10.1123/pes.2013-0095>
- Costa, S., Barber, S. E., Cameron, N., & Clemes, S. A. (2014). Calibration and validation of the ActiGraph GT3X+ in 2–3 year olds. *Journal of Science and Medicine in Sport*, 17(6), 617–622. <https://doi.org/10.1016/j.jsams.2013.11.005>
- Dinkel, D., Snyder, K., Patterson, T., Warehime, S., Kuhn, M., & Wisneski, D. (2019). An exploration of infant and toddler unstructured outdoor play. *European Early Childhood Education Research Journal*, 27(2), 257–271. <https://doi.org/10.1080/1350293X.2019.1579550>
- Dlugonski, D., DuBose, K. D., & Rider, P. (2017). Accelerometer-Measured Patterns of Shared Physical Activity Among Mother–Young Child Dyads. *Journal of Physical Activity and Health*, 14(10), 808–814. <https://doi.org/10.1123/jpah.2017-0028>
- Fairclough, S. J., Noonan, R., Rowlands, A. V., Van Hees, V., Knowles, Z., & Boddy, L. M. (2016). Wear Compliance and Activity in Children Wearing Wrist- and Hip-Mounted Accelerometers. *Medicine & Science in Sports & Exercise*, 48(2), 245–253. <https://doi.org/10.1249/MSS.0000000000000771>

- Fees, B. S., Fischer, E., Haar, S., & Crowe, L. K. (2015). Toddler Activity Intensity During Indoor Free-Play: Stand and Watch. *Journal of Nutrition Education and Behavior, 47*(2), 170–175. <https://doi.org/10.1016/j.jneb.2014.08.015>
- Felzer-Kim, I. T., & Hauck, J. L. (2020). Sleep duration associates with moderate-to-vigorous intensity physical activity and body fat in 1- to 3-year-old children. *Infant Behavior and Development, 58*, 101392. <https://doi.org/10.1016/j.infbeh.2019.101392>
- Gubbels, J. S., Kremers, S. P. J., van Kann, D. H. H., Stafleu, A., Candel, M. J. J. M., Dagnelie, P. C., Thijs, C., & de Vries, N. K. (2011). Interaction between physical environment, social environment, and child characteristics in determining physical activity at child care. *Health Psychology, 30*(1), 84–90. <https://doi.org/10.1037/a0021586>
- Hager, E. R., Gormley, C. E., Latta, L. W., Treuth, M. S., Caulfield, L. E., & Black, M. M. (2016). Toddler physical activity study: Laboratory and community studies to evaluate accelerometer validity and correlates. *BMC Public Health, 16*(1), 936. <https://doi.org/10.1186/s12889-016-3569-9>
- Hauck, J. L., & Felzer-Kim, I. T. (2019). Time Spent in Sedentary Activity Is Related to Gross Motor Ability During the Second Year of Life. *Perceptual and Motor Skills, 126*(5), 753–763. <https://doi.org/10.1177/0031512519858261>
- Hawley, P. H., & Little, T. D. (1999). On winning some and losing some: A social relations approach to social dominance in toddlers. *Merrill-Palmer Quarterly, 45*(2), 185–214.
- Herzig, D., Eser, P., Radtke, T., Wenger, A., Rusterholz, T., Wilhelm, M., Achermann, P., Arhab, A., Jenni, O. G., Kakebeeke, T. H., Leeger-Aschmann, C. S., Messerli-Bürgy, N., Meyer, A. H., Munsch, S., Puder, J. J., Schmutz, E. A., Stülb, K., Zysset, A. E., & Kriemler, S. (2017). Relation of Heart Rate and its Variability during Sleep with Age, Physical Activity, and Body Composition in Young Children. *Frontiers in Physiology, 8*. <https://doi.org/10.3389/fphys.2017.00109>
- Hnatiuk, J. A., Ridgers, N. D., Salmon, J., & Hesketh, K. D. (2017). Maternal correlates of young children's physical activity across periods of the day. *Journal of Science and Medicine in Sport, 20*(2), 178–183. <https://doi.org/10.1016/j.jsams.2016.06.014>
- Hnatiuk, J., Ridgers, N. D., Salmon, J., Campbell, K., McCallum, Z., & Hesketh, K. (2012). Physical Activity Levels and Patterns of 19-Month-Old Children. *Medicine & Science in Sports & Exercise, 44*(9), 1715–1720. <https://doi.org/10.1249/MSS.0b013e31825825c4>
- Johansson, E., Ekelund, U., Nero, H., Marcus, C., & Hagströmer, M. (2015). Calibration and cross-validation of a wrist-worn ActiGraph in young preschoolers: Calibration of ActiGraph in toddlers. *Pediatric Obesity, 10*(1), 1–6. <https://doi.org/10.1111/j.2047-6310.2013.00213.x>

- Johansson, E., Hagströmer, M., Svensson, V., Ek, A., Forssén, M., Nero, H., & Marcus, C. (2015). Objectively measured physical activity in two-year-old children – levels, patterns and correlates. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1), 3. <https://doi.org/10.1186/s12966-015-0161-0>
- Karnik, S., & Kanekar, A. (2012). Childhood Obesity: A Global Public Health Crisis. *International Journal of Preventive Medicine*, 3(1).
- Kelly, L. A., Knox, A., Gonzalez, C., Lennartz, P., Hildebrand, J., Carney, B., Wendt, S., Haas, R., & Hill, M. D. (2022). Objectively Measured Physical Activity and Sedentary Time of Suburban Toddlers Aged 12–36 Months. *International Journal of Environmental Research and Public Health*, 19(11), 6707. <https://doi.org/10.3390/ijerph19116707>
- Kelly, L. A., & Villalpando, J. (2016). Development of ActiGraph GT1M Accelerometer Cut-Points for Young Children Aged 12-36 Months. *Journal of Athletic Enhancement*, 5(4). <https://doi.org/10.4172/2324-9080.1000233>
- Konstabel, K., Veidebaum, T., Verbestel, V., Moreno, L. A., Bammann, K., Tornaritis, M., Eiben, G., Molnár, D., Siani, A., Sprengeler, O., Wirsik, N., Ahrens, W., & Pitsiladis, Y. (2014). Objectively measured physical activity in European children: The IDEFICS study. *International Journal of Obesity*, 38(S2), S135–S143. <https://doi.org/10.1038/ijo.2014.144>
- Kwon, S., Honegger, K., & Mason, M. (2019). Daily Physical Activity Among Toddlers: Hip and Wrist Accelerometer Assessments. *International Journal of Environmental Research and Public Health*, 16(21), 4244. <https://doi.org/10.3390/ijerph16214244>
- Lee, E.-Y., Hesketh, K. D., Hunter, S., Kuzik, N., Rhodes, R. E., Rinaldi, C. M., Spence, J. C., & Carson, V. (2017). Meeting new Canadian 24-Hour Movement Guidelines for the Early Years and associations with adiposity among toddlers living in Edmonton, Canada. *BMC Public Health*, 17(S5), 840. <https://doi.org/10.1186/s12889-017-4855-x>
- McCullough, A. K., Duch, H., & Garber, C. E. (2018). Interactive Dyadic Physical Activity and Spatial Proximity Patterns in 2-Year-Olds and Their Parents. *Children*, 5(12), 167. <https://doi.org/10.3390/children5120167>
- McIver, K. L., Brown, W. H., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2009). ASSESSING CHILDREN'S PHYSICAL ACTIVITY IN THEIR HOMES: THE OBSERVATIONAL SYSTEM FOR RECORDING PHYSICAL ACTIVITY IN CHILDREN-HOME. *Journal of Applied Behavior Analysis*, 42(1), 1–16. <https://doi.org/10.1901/jaba.2009.42-1>
- McIver, K. L., Brown, W. H., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2016). Development and Testing of the Observational System for Recording Physical Activity in Children:

- Elementary School. *Research Quarterly for Exercise and Sport*, 87(1), 101–109.
<https://doi.org/10.1080/02701367.2015.1125994>
- McKenzie, T. L., Marshall, S. J., Sallis, J. F., & Conway, T. L. (2002). *System for Observing Play and Leisure Activity in Youth* [dataset]. American Psychological Association.
<https://doi.org/10.1037/t72617-000>
- McKenzie, T. L., Sallis, J. F., & Nader, P. R. (1991). *SOFIT: System for Observing Fitness Instruction Time*. 11, 195–205.
- McKenzie, T. L., Sallis, J. F., Nader, P. R., Patterson, T. L., Elder, J. P., Berry, C. C., Rupp, J. W., Atkins, C. J., Buono, M. J., & Nelson, J. A. (1991). BEACHES: AN OBSERVATIONAL SYSTEM FOR ASSESSING CHILDREN'S EATING AND PHYSICAL ACTIVITY BEHAVIORS AND ASSOCIATED EVENTS. *Journal of Applied Behavior Analysis*, 24(1), 141–151.
<https://doi.org/10.1901/jaba.1991.24-141>
- Nyberg, G. A., Nordenfelt, A. M., Ekelund, U., & Marcus, C. (2009). Physical Activity Patterns Measured by Accelerometry in 6- to 10-yr-Old Children. *Medicine & Science in Sports & Exercise*, 41(10), 1842–1848. <https://doi.org/10.1249/MSS.0b013e3181a48ee6>
- Oftedal, S., Bell, K. L., Davies, P. S. W., Ware, R. S., & Boyd, R. N. (2014). Validation of Accelerometer Cut Points in Toddlers with and without Cerebral Palsy. *Medicine & Science in Sports & Exercise*, 46(9), 1808–1815.
<https://doi.org/10.1249/MSS.0000000000000299>
- O'hara, N. M., Baranowski, T., Wilson, B. S., Parcel, G. S., & Simons-Morton, B. G. (1989). Validity of the Observation of Children's Physical Activity. *Research Quarterly for Exercise and Sport*, 60(1), 42–47. <https://doi.org/10.1080/02701367.1989.10607412>
- Okely, A. D., Ghersi, D., Hesketh, K. D., Santos, R., Loughran, S. P., Cliff, D. P., Shilton, T., Grant, D., Jones, R. A., Stanley, R. M., Sherring, J., Hinkley, T., Trost, S. G., McHugh, C., Eckermann, S., Thorpe, K., Waters, K., Olds, T. S., Mackey, T., ... Tremblay, M. S. (2017). A collaborative approach to adopting/adapting guidelines - The Australian 24-Hour Movement Guidelines for the early years (Birth to 5 years): An integration of physical activity, sedentary behavior, and sleep. *BMC Public Health*, 17(S5), 869.
<https://doi.org/10.1186/s12889-017-4867-6>
- Orlando, J. M., Pierce, S., Mohan, M., Skorup, J., Paremski, A., Bochnak, M., & Prosser, L. A. (2019). Physical activity in non-ambulatory toddlers with cerebral palsy. *Research in Developmental Disabilities*, 90, 51–58. <https://doi.org/10.1016/j.ridd.2019.04.002>
- Pate, R. R., Almeida, M. J., Mclver, K. L., Pfeiffer, K. A., & Dowda, M. (2006). Validation and Calibration of an Accelerometer in Preschool Children*. *Obesity*, 14(11), 2000–2006.
<https://doi.org/10.1038/oby.2006.234>

- Pate, R. R., McIver, K., Dowda, M., Brown, W. H., & Addy, C. (2008). Directly Observed Physical Activity Levels in Preschool Children. *Journal of School Health, 78*(8), 438–444. <https://doi.org/10.1111/j.1746-1561.2008.00327.x>
- Pereira, J., Santos, R., Sousa-Sá, E., Zhang, Z., Burley, J., Veldman, S. L. C., & Cliff, D. P. (2021). Longitudinal differences in levels and bouts of sedentary time by different day types among Australian toddlers and pre-schoolers. *Journal of Sports Sciences, 39*(24), 2804–2811. <https://doi.org/10.1080/02640414.2021.1964747>
- Pfeiffer, K. A., Mciver, K. L., Dowda, M., Almeida, M. J. C. A., & Pate, R. R. (2006). Validation and Calibration of the Actical Accelerometer in Preschool Children. *Medicine & Science in Sports & Exercise, 38*(1), 152–157. <https://doi.org/10.1249/01.mss.0000183219.44127.e7>
- Puhl, J., Greaves, K., Hoyt, M., & Baranowski, T. (1990). Children's Activity Rating Scale (CARS): Description and Calibration. *Research Quarterly for Exercise and Sport, 61*(1), 26–36. <https://doi.org/10.1080/02701367.1990.10607475>
- Pulakka, A., Cheung, Y., Ashorn, U., Penpraze, V., Maleta, K., Phuka, J., & Ashorn, P. (2013). Feasibility and validity of the ActiGraph GT3X accelerometer in measuring physical activity of Malawian toddlers. *Acta Paediatrica, 102*(12), 1192–1198. <https://doi.org/10.1111/apa.12412>
- Pulakka, A., Cheung, Y. B., Maleta, K., Dewey, K. G., Kumwenda, C., Bendabenda, J., Ashorn, U., & Ashorn, P. (2017). Effect of 12-month intervention with lipid-based nutrient supplement on the physical activity of Malawian toddlers: A randomised, controlled trial. *British Journal of Nutrition, 117*(4), 511–518. <https://doi.org/10.1017/S0007114517000290>
- Rice, K. R., Joschtel, B., & Trost, S. G. (2013). Validity of Family Child Care Providers' Proxy Reports on Children's Physical Activity. *Childhood Obesity, 9*(5), 393–398. <https://doi.org/10.1089/chi.2013.0035>
- Sanders, T., Cliff, D. P., & Lonsdale, C. (2014). Measuring Adolescent Boys' Physical Activity: Bout Length and the Influence of Accelerometer Epoch Length. *PLoS ONE, 9*(3), e92040. <https://doi.org/10.1371/journal.pone.0092040>
- Santos, R., Zhang, Z., Pereira, J. R., Sousa-Sá, E., Cliff, D. P., & Okely, A. D. (2017). Compliance with the Australian 24-hour movement guidelines for the early years: Associations with weight status. *BMC Public Health, 17*(S5), 867. <https://doi.org/10.1186/s12889-017-4857-8>

- Schenkelberg, M. A., Brown, W. H., McIver, K. L., & Pate, R. R. (2021). An observation system to assess physical activity of children with developmental disabilities and delays in preschool. *Disability and Health Journal*, *14*(2), 101008. <https://doi.org/10.1016/j.dhjo.2020.101008>
- Sirard, J. R., Trost, S. G., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2005). Calibration and Evaluation of an Objective Measure of Physical Activity in Preschool Children. *Journal of Physical Activity and Health*, *2*(3), 345–357. <https://doi.org/10.1123/jpah.2.3.345>
- Specker, B. L., Mulligan, L., & Ho, M. (1999). Longitudinal Study of Calcium Intake, Physical Activity, and Bone Mineral Content in Infants 6–18 Months of Age. *Journal of Bone and Mineral Research*, *14*(4), 569–576. <https://doi.org/10.1359/jbmr.1999.14.4.569>
- Tanaka, S., & Kuroda, M. (2022). On-site Physical Activity Analysis for Toddler in Unconstrained Environment. *2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech)*, 150–153. <https://doi.org/10.1109/LifeTech53646.2022.9754847>
- Taylor, R. W., Haszard, J. J., Meredith-Jones, K. A., Galland, B. C., Heath, A.-L. M., Lawrence, J., Gray, A. R., Sayers, R., Hanna, M., & Taylor, B. J. (2018). 24-h movement behaviors from infancy to preschool: Cross-sectional and longitudinal relationships with body composition and bone health. *International Journal of Behavioral Nutrition and Physical Activity*, *15*(1), 118. <https://doi.org/10.1186/s12966-018-0753-6>
- The TARGet Kids Collaboration, Sarker, H., Anderson, L. N., Borkhoff, C. M., Abreo, K., Tremblay, M. S., Lebovic, G., Maguire, J. L., Parkin, P. C., & Birken, C. S. (2015). Validation of parent-reported physical activity and sedentary time by accelerometry in young children. *BMC Research Notes*, *8*(1), 735. <https://doi.org/10.1186/s13104-015-1648-0>
- Timmons, B. W., LeBlanc, A. G., Carson, V., Connor Gorber, S., Dillman, C., Janssen, I., Kho, M. E., Spence, J. C., Stearns, J. A., & Tremblay, M. S. (2012). Systematic review of physical activity and health in the early years (aged 0–4 years). *Applied Physiology, Nutrition, and Metabolism*, *37*(4), 773–792. <https://doi.org/10.1139/h2012-070>
- Tremblay, M. S., Chaput, J.-P., Adamo, K. B., Aubert, S., Barnes, J. D., Choquette, L., Duggan, M., Faulkner, G., Goldfield, G. S., Gray, C. E., Gruber, R., Janson, K., Janssen, I., Janssen, X., Jaramillo Garcia, A., Kuzik, N., LeBlanc, C., MacLean, J., Okely, A. D., ... Carson, V. (2017). Canadian 24-Hour Movement Guidelines for the Early Years (0–4 years): An Integration of Physical Activity, Sedentary Behaviour, and Sleep. *BMC Public Health*, *17*(S5), 874. <https://doi.org/10.1186/s12889-017-4859-6>
- Trost, S. G. (2007). State of the Art Reviews: Measurement of Physical Activity in Children and Adolescents. *American Journal of Lifestyle Medicine*, *1*(4), 299–314. <https://doi.org/10.1177/1559827607301686>

- Trost, S. G., Fees, B. S., Haar, S. J., Murray, A. D., & Crowe, L. K. (2012). Identification and Validity of Accelerometer Cut-Points for Toddlers. *Obesity, 20*(11), 2317–2319. <https://doi.org/10.1038/oby.2011.364>
- U.S. Department of Health and Human Service. (2008). *2008 Physical Activity Guidelines for Americans*. U.S. Department of Health and Human Service. <https://health.gov/sites/default/files/2019-09/paguide.pdf>
- U.S. Department of Health and Human Service. (2018). *Physical Activity Guidelines for Americans, 2nd edition*. U.S. Department of Health and Human Service. https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf
- Van Cauwenberghe, E., Gubbels, J., De Bourdeaudhuij, I., & Cardon, G. (2011). Feasibility and validity of accelerometer measurements to assess physical activity in toddlers. *International Journal of Behavioral Nutrition and Physical Activity, 8*(1), 67. <https://doi.org/10.1186/1479-5868-8-67>
- Vanderloo, L. M., & Tucker, P. (2015). An objective assessment of toddlers' physical activity and sedentary levels: A cross-sectional study. *BMC Public Health, 15*(1), 969. <https://doi.org/10.1186/s12889-015-2335-8>
- Washburn, R. A., Heath, G. W., & Jackson, A. W. (2000). Reliability and Validity Issues concerning Large-Scale Surveillance of Physical Activity. *Research Quarterly for Exercise and Sport, 71*(sup2), 104–113. <https://doi.org/10.1080/02701367.2000.11082793>
- WHO Multicentre Growth Reference Study Group, & Onis, M. (2007). WHO Motor Development Study: Windows of achievement for six gross motor development milestones: Windows of achievement for motor milestones. *Acta Paediatrica, 95*, 86–95. <https://doi.org/10.1111/j.1651-2227.2006.tb02379.x>
- Wijtzes, A. I., Kooijman, M. N., Kiefte-de Jong, J. C., de Vries, S. I., Henrichs, J., Jansen, W., Jaddoe, V. W. V., Hofman, A., Moll, H. A., & Raat, H. (2013). Correlates of Physical Activity in 2-Year-Old Toddlers: The Generation R Study. *The Journal of Pediatrics, 163*(3), 791-799.e2. <https://doi.org/10.1016/j.jpeds.2013.02.029>
- Willumsen, J., & Bull, F. (2020). Development of WHO Guidelines on Physical Activity, Sedentary Behavior, and Sleep for Children Less Than 5 Years of Age. *Journal of Physical Activity and Health, 17*(1), 96–100. <https://doi.org/10.1123/jpah.2019-0457>

CHAPTER 3: A SYSTEMATIC REVIEW OF PHYSICAL ACTIVITY ASSESSMENT METHODS DEVELOPED FOR/IN TODDLERS

Abstract

Adequate levels of physical activity (PA) have been associated with positive physical and mental health outcomes in toddlers. Many methods of assessing PA in preschoolers have been developed; however, fewer have been reported in toddlers, and no literature that has evaluated (not just applied) potential methods exists in this age group. **Objective:** The purpose of this review is to identify existing methods of assessing PA levels that have been developed for use with toddlers. **Methods:** This systematic review was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) framework. The databases PubMed, Web of Science, SCOPUS, and EBSCO were searched for articles published prior to October 2023. Articles were included if they 1) Included human subjects, 2) 50% of age-range included 12 – 36 months, 3) Included development, validation, or cross-validation, 4) Were published by October 2023, 5) Were available in English and full-text. Information was extracted using a standardized form. Quality and risk of bias was assessed using the Checklist for Analytical Cross-Section Studies: Critical Appraisal tools for use in JBI Systematic Reviews. **Results:** Sixteen studies were included in the final review. Participants were healthy in all but one study and only 30% of studies included commonly marginalized groups. 40% of participants were considered overweight or obese. Half of the developed methods were the calibration of cut-points (N=8) and almost all studies used direct observation as the criterion (N=9). **Conclusion:** There are a limited number of methods that have been specifically developed for assessing PA in toddlers. Limitations include that no meta-analyses were conducted in this review so no synthesis of quantitative results can be done. Articles not published in English or without full-text

availability were not included. Future research should aim to further develop or validate existing tools. More advanced methods of device-based assessments are warranted.

Introduction

Physical activity (PA) is defined as any bodily movement produced by the contraction of skeletal muscle that results in an increase in caloric requirement over resting energy expenditure (Caspersen et al., 1985). In toddlers, engaging in an adequate amount of physical activity has not only been associated with physical health, but with positive cognitive and socioemotional health outcomes as well (Lee et al., 2017). Literature also suggests that PA levels established in the early years track, not only through childhood, but also into adulthood (Carson et al., 2019). Because of these associations it is essential that we can accurately assess PA within this population.

The assessment of PA in youth can be challenging due to their quick and sporadic periods of movement. PA can be assessed using a multitude of different methods (Sirard & Pate, 2001; Trost, 2007). Each method has its own strengths and weaknesses, and the preferred method depends on the study being conducted. Common assessment methods include direct observation (DO), accelerometry, and questionnaires. A large number of methods have been developed for use in preschoolers, children, and adolescents (Cain et al., 2013; Hidding et al., 2018; Lettink et al., 2022). However, a limited number of methods have been developed to assess toddler PA (Lettink et al., 2022; Timmons et al., 2012). Previously conducted reviews addressing accelerometry conclude that many of the methods applied in toddlers were developed in preschool-aged children (Breau et al., 2022; Bruijns et al., 2020).

Recent reviews highlight the variability in the methods used to assess PA in young children (Breau et al., 2022; Bruijns et al., 2020). There are differences in not only the overall method chosen (e.g., accelerometry v. DO), but also in data collection or analysis decisions such

as DO tool (e.g., Children's Physical Activity Form v. Children's Activity Rating Scale), setting (e.g., structured v. unstructured play) or accelerometer parameters (e.g., wear location, epoch length, data reduction methods). When identifying literature using accelerometry to assess PA in youth 0 – 5 years of age, a recent systematic review noted a limitation of their study was that they could not draw conclusions regarding the amount of PA in which youth were engaging due to the differences in accelerometer parameters (Bruijns et al., 2020). Another study determined that there were significant differences in accelerations when comparing a hip-worn accelerometer to a wrist-worn monitor, highlighting the importance of wear location (Kwon et al., 2019). Lastly, when comparing studies that aimed to measure 24-hour activity in toddlers, depending on the cut-point applied, total PA levels ranged from 72.9 – 636.5 min/day. Moderate-to-vigorous physical activity (MVPA) ranged from 6.5 – 89.9 min/day (Bruijns et al., 2020). These variances in methodology can make comparability across studies difficult and make conclusions difficult to elucidate.

A better understanding of the available methods developed to measure PA in toddlers is a first step towards identifying best practice measures and therefore minimizing methodological differences between studies. To our knowledge, no review has evaluated what methods have specifically been developed for (not just applied) in this age group. Therefore, the purpose of this paper is to systematically review methods of assessing PA levels that have been developed in a toddler population. It was hypothesized that available literature will support that PA assessment methods specifically developed for toddlers are limited and that accelerometry, in addition to cut-points, will be the most utilized assessment/data reduction method used.

Methods

This systematic review was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (Moher et al., 2009).

Literature search and search strategy

This systematic review is registered in the international prospective register of systematic reviews (CRD42023448641). A systematic search of the electronic databases PubMed, Web of Science, Scopus, and SPORTdiscus was conducted in October 2023. The search strategy was reviewed and agreed upon by all review team members. The electronic databases were searched using a combination of the following terms: (“physical activity” OR “activity intensity” OR “activity level” OR “sedentary”) AND (Calibrat* OR develop* OR predict* OR valid* OR cross-validat*) AND (toddler or preschool*) AND (subjective OR self-report OR questionnaire OR objective OR acceleromet* OR monitor OR sensor OR pedomet* OR device OR direct observation). Articles were filtered for human studies when applicable, and no date restriction was included. A secondary search was conducted by examining the reference sections of retrieved papers and relevant reviews to identify additional studies not identified in the initial search.

Selection of Studies

Search results were exported to Microsoft Excel. Duplicates were removed, and remaining studies were screened by title, then abstract, then by full text by two reviewers independently (CV & CB). Titles and abstracts selected by at least one reviewer progressed, but discrepancies in full-text selections were discussed by CV & CB. Studies were included if they met the following inclusion and exclusion criteria:

Inclusion Criteria

- (1) Includes human subjects
- (2) At least 50% of the sample age range includes ages 12 – 36 months (1 – 3 years)
- (3) Includes the development, validation or cross-validation of PA assessment in toddlers
- (4) Was published by October 2023
- (5) Article available in English
- (6) Full text of article available

Exclusion Criteria

- (1) Usage of previously developed methods to assess PA levels (no validation)
- (2) Greater than 50% of age inclusion outside of 12 – 36 months (e.g., 2 – 4 years)

Any discrepancies in inclusion at the full text stage were resolved by discussion between the two reviewers. Studies excluded during the full text screening phase were recorded with their reason for exclusion.

Data Extraction

Data were extracted from each article by one reviewer (CV) once full-text articles were selected for inclusion in this review. A standardized data-extraction form was used. The reviewer was not blinded to authors or journals when extracting data. Extracted data included general information about the article (e.g., authors, year, location, population, and study design), sample characteristics (e.g., age, sex, sample size, and health status), direct or indirect measure(s), and specifics regarding measurement. For example, if accelerometers were used, data such as what monitor, wear location, and epoch used were extracted.

Quality and Risk of Bias Assessment

Article quality and risk of bias was assessed using the Checklist for Analytical Cross-Section Studies: Critical Appraisal tools for use in JBI Systematic Reviews (Moola et al., 2020). This tool provides pre-established questions relevant to the methodological quality of a study. This tool was chosen due to its specific questions regarding the quality of measurement methods. Two reviewers independently answered the eight questions and scored each article included in the review. After completing the tool, reviewers rated each article as “include,” “exclude,” or “seek more information.” The two reviewers discussed any discrepancies regarding quality.

Results

As seen in the PRISMA flowchart (Figure 1.), 10,474 articles were identified from the initial search. After removing duplicates, 3,557 articles remained. Another 3,203 articles were removed during title screening, and 254 additional articles were removed after abstract screening. One-hundred full-texts were assessed and articles were removed due to age range (N=51), methodology not including validation (N=30), and format (i.e., theses and dissertations; N=3). This resulted in 16 articles being included in the final selection (Altenberg et al., 2021; Bingham et al., 2016; Costa et al., 2014; Haar et al., 2013; Hager et al., 2016; Henriksson et al., 2014; Johansson et al., 2015; Kelly & Villalpando, 2016; Klesges & Klesges, 1985; Oftedal et al., 2014; Pereira et al., 2020; Pulakka et al., 2013; Sarker et al., 2015; Trost et al., 2012; Tulve et al., 2007; Van cauwenberghe et al., 2011). No additional articles were found during the review of the included studies' citations or studies included in other reviews. Detailed information, such a

sample characteristics, study design, criterion measures, developed methods, and risk of bias assessments can be found in Appendix A.

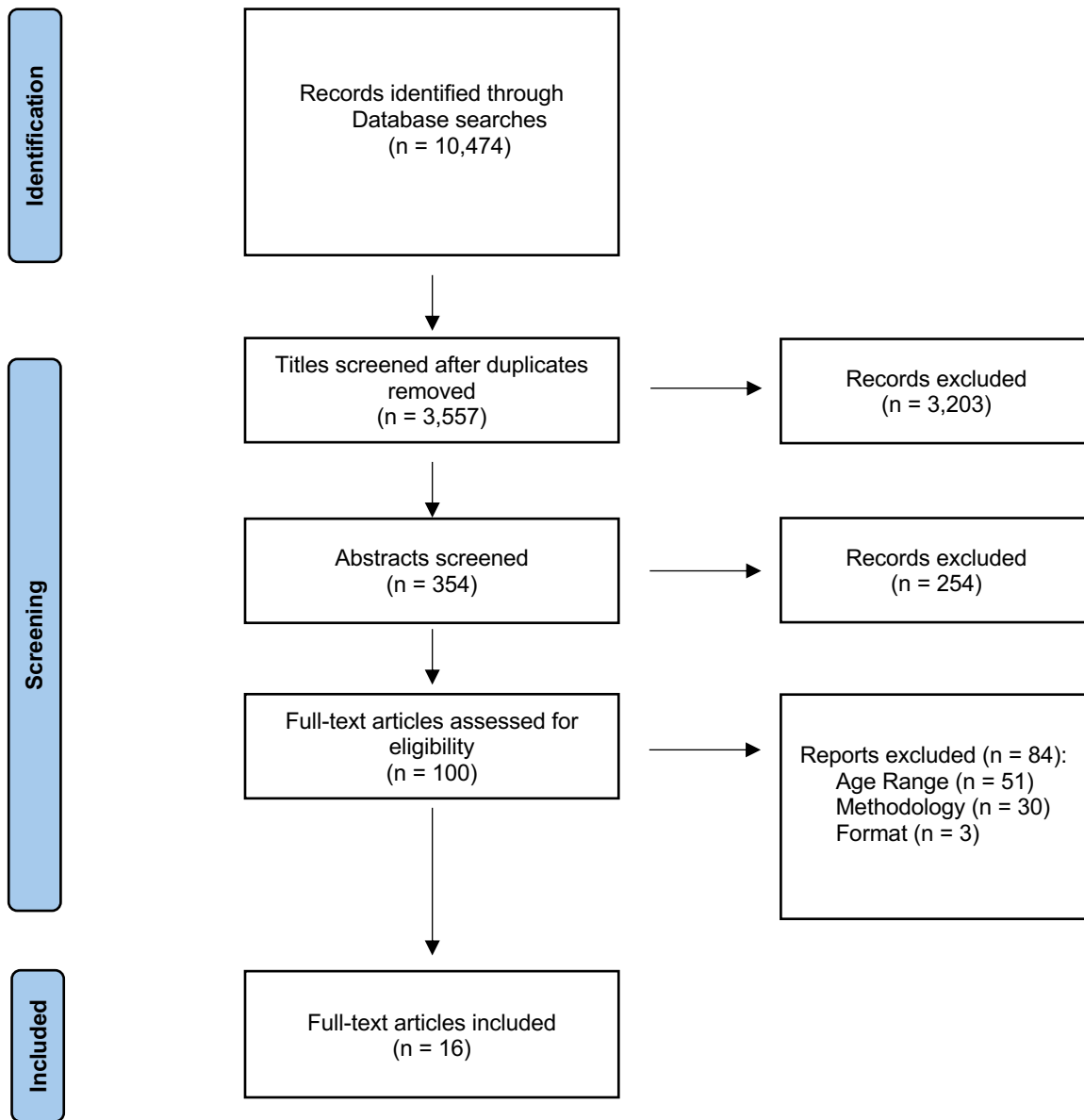


Figure 1. Flow diagram of the search for physical activity assessment methods in toddlers

Most assessment methods were device-based (n=10), while few were survey (n=4) or cross-validation of preschool cut-points (n=3). Most methods were developed and/or tested in

North America (n=7), followed by Europe (n=6), Australia (n=2), and Africa (n=1). In ten studies race/ethnicity was not reported, and in those that reported race, only ~30% of samples included commonly underrepresented groups. Only one study included toddlers with a chronic health condition (cerebral palsy), but it also included a healthy comparison group. Fourteen studies included only healthy toddlers. Of these studies, seven reported that up to 40% of their population was considered overweight or obese according to Body Mass Index (BMI) z-scores.

Device-based PA Assessment

Ten studies included the use of accelerometry. Eight of these studies developed toddler specific cut-points to assess PA, while the remaining two studies correlated counts/minute to time spent in PA according to DO. Half of the studies utilized an independent sample to cross-validate developed cut-points. The remaining studies used the same sample (n=4) or assessed concurrent validity (n=1). Sample sizes ranged from 23-103 in the original samples to 12-277 in the cross-validation samples. All but one study used a form of DO as their criterion measure of PA. The remaining study used doubly-labelled water. A version of the ActiGraph (GT1M or GT3X+) was the most common accelerometer type (n=7). The hip was the most common wear location (n= 5), and five (n =3), fifteen (n = 5), and thirty (n = 1) second epochs were used. Six studies included unstructured free play sessions, and four had structured sessions. In the structured sessions, toddlers engaged in sedentary (e.g., sitting, drawing, standing), light (e.g., walking, rolling) and moderate-to-vigorous (e.g., running, skipping, jumping) activities. All studies concluded that accelerometers/cut-points could be used distinguish sedentary time from MVPA.

Survey-based PA Assessment

Four studies assessed the validity of a survey or questionnaire-based measure of PA. Activity diaries (n = 2), the Early Years – Physical Activity Questionnaire (EY-PAQ; n= 1), and an adapted Canadian Health Survey (n = 1) were all used. Sample sizes ranged from 9-196 participants. In three studies an accelerometer was used for the comparison measure of PA (ActiCal n = 2). Only one study included test-retest reliability. One study included both heart rate monitoring (ActiHeart) and doubly-labelled water as comparison measures. All questionnaires were administered over 7 days, except one activity diary was kept during a free-play session. Three studies compared cut-points (toddler developed n=1) to the surveys while the others used counts/min or metabolic equivalent of task (MET) values. All PA measures had relatively low correlation with the criterion measure ($r = 0.33 - 0.42$). All four studies concluded that they could not use their method to assess sedentary time, but that sedentary time could be split from MVPA. However, authors concluded these measures were comparable to other assessment methods.

Cross-validation of cut-point methods

Finally, only three studies aimed to cross-validate previously developed cut-points (preschool or toddler developed) in toddlers. Sample sizes in these studies ranged from 31-63 participants. Two of the three studies used DO as a criterion, while the third study used an accelerometer (activPAL). All three studies included a free play session that was 19-60 minutes in duration. An ActiGraph was used in all three studies, and Pate (2006), Sirard (2005), and Van Cauwenberghe cut-points, created for preschoolers, were most commonly applied (n = 3, 3, & 2 respectively). Compared to direct observations, both studies concluded that preschool-

developed cut-points were better for distinguishing between sedentary and non-sedentary time, rather than activity intensities (i.e., sedentary, light, and moderate-to-vigorous). In all studies sedentary time classification was fair, but underestimated (AUC = 0.56 – 0.72), while light (AUC = 0.47 – 0.62) and moderate-to-vigorous (AUC = 0.50 – 0.66) PA was poor and overestimated. When comparing applied sedentary cut-points to the activPAL, authors concluded that no estimates were within 10% equivalencies. However, two sets of cut-points (48 counts/15-seconds and 5 counts/5s) resulted in the least amount of bias. These two cut-points were the only toddler-developed cut-points applied.

Risk of Bias

No studies were excluded after the JBI Critical Appraisal Tool for Cross-Sectional Studies was completed. On average, articles included 77.7% of the criteria outlined in the checklist. Studies were determined to be included if they scored yes on five or more out of seven questions (70%). Almost half of studies did not specify inclusion criteria for participants (n=7) or discuss any potential confounding factors and how they were handled (n=8). All studies reported use of a valid and reliable methods to measure both exposures and outcomes (n=16).

Discussion

The assessment of PA in toddlers presents unique challenges due to their rapid and sporadic movements and the need to consider their developmental stage and environmental context. Toddlers have not yet achieved many motor milestones that most preschool aged children have (Colson & Dworkin, 1997). Motor skills such as running, climbing, and jumping are not achieved until around 24 months, and there is large variability between individuals' skill. Fine motor skills such a writing, drawing, and block-building are not accomplished until as late

as 36-months (Colson & Dworkin, 1997). Because of this rapid period of skill development, assessing toddler physical activity may present with unique challenges. In this review, we identified methods that have been developed, validated, or cross-validated to assess PA specifically in toddlers, aiming to provide insights into the current state of PA assessment in this age group.

One noteworthy finding from our review is the limited number of methods specifically designed for assessing PA in toddlers. A recent systematic review aiming to identify only accelerometer-based measurements of PA for youth (0 – 5 years) concluded that there were 40 different approaches for preschoolers (3 – 5 years; (Lettink et al., 2022). When looking at subjective methods, a recent review reported that 67 different questionnaires have been tested in children and adolescents, some questionnaires being tested multiple times (Hidding et al., 2018). These reviews and the present review highlight that the literature on validated methods tailored to toddlers is sparse. This gap in validated tools may hinder accurate assessment of PA levels in this population and limit our ability to understand and promote PA in toddlers effectively.

Among the methods identified, accelerometer-based cut-points were the most reported. These studies all reported sensitivity and specificity values over 60% and similar AUC values, with most being 0.85 and above. These are similar to values reported for classification accuracy in popular cut-points developed for preschool (Pate et al., 2006; Pfeiffer et al., 2006; Sirard et al., 2005) and elementary-aged children (Evenson KR et al., 2008) cut-points. However, only half of the created methods were validated in an independent sample, limiting their validity and reliability (Browne, 2000). Without application in a larger sample or a more diverse

sample, we cannot determine that these cut-points could be applied across toddler populations.

It is also notable to mention that no accelerometer-based methods to assess PA levels in toddlers have been calibrated using advanced machine-learning approaches. Methods have been developed in both preschoolers (Butte et al., 2014; Li et al., 2020) and children and adolescents (M. Ahmadi et al., 2018; Montoye et al., 2019). This is significant because in preschoolers, compared to cut-point methods, machine learning approaches have shown to be advantageous (M. N. Ahmadi & Trost, 2022). More advanced approaches should be considered for this population.

Recommendations for calibrating device-based methods of activity assessment in youth include integrating energy expenditure into calibration methods, using advanced approaches (i.e., machine learning), including a control group, and conducting independent sample cross-validation (Arem et al., 2015; Bianchim et al., 2020). Additionally, in adults, calibration methods that included activities of daily-living, rather than only ambulatory activities (i.e., walking, running, jumping), resulted in better estimation of activity levels (Arem et al., 2015; Matthews et al., 2021). Most studies identified in the review utilized free play sessions, that allowed toddlers free-will in activity. This means activities of daily living and ambulatory activities had the ability to be included in validation. However, no studies utilized energy expenditure methods, such as indirect calorimetry, in their calibration. This is important because there is currently no evidence that the DO activity assessment methods used as the criterion are valid representations of energy expenditure in toddlers. Therefore, we do not know if what is

considered light activity for a preschooler is also light activity for a toddler. Future studies should include this metric to further validate developed metrics.

As mentioned, more questionnaire-based methods have been developed for children older than six than for those 0 – 5 years (Hidding et al., 2018). In a larger systematic review, questionnaire quality was assessed (Hidding et al., 2018). Authors concluded that no convincing evidence for both validity and reliability was found. Although that was not an aim of the current review, three of the studies involving a questionnaire did not assess reliability of the measure in addition to validity. Reported validity was weak to moderate for all questionnaires for all PA levels (total PA, sedentary time, and moderate-to-vigorous PA). Consequently, there is a strong need for development of high-quality questionnaires in this age group.

All survey-based studies used a device-based approach as a comparison measure. Although device-based metrics are not commonly used as a criterion in a laboratory-based setting, they are frequently used in free-living settings where other criterion metrics are not plausible (Troost et al., 2005). A variety of accelerometers have been found to be valid in youth when compared against indirect calorimetry and doubly-labelled water methods (Freedson et al., 2005; Troost et al., 2005). Therefore, in survey work, they are frequently used as a comparison to measure concurrent validity.

Previous literature suggests that when developing surveys, test-retest reliability be assessed (Arem et al., 2015; Saint-Maurice et al., 2017). This requires researchers to administer a second survey to participants to determine differences in responses. This can also account for differences in behavior across time due to things such as seasons (Arem et al., 2015). It is also recommended that, if possible, surveys and device-based metrics be used in conjunction with

one another and error correction approaches can be applied (Arem et al., 2015; Matthews et al., 2013).

It is common for researchers to take already developed methods and apply them to their data. A recent review reported that in studies assessing PA using cut-point methods in toddlers, the Trost (2012) cut-points were the most applied (Bruijns et al., 2020). These were developed and calibrated in toddlers; however, other cut-points, such as Pate (2006), Sirard (2005), and Evenson (2008) have also been used. When cross-validated, these cut-points were found to have poor to fair classification accuracy (Altenburg et al., 2022; van Cauwenberghe E et al., 2011). To maximize accuracy, only toddler calibrated cut-points should be applied in this age group. Furthermore, more validation work is needed on these developed methods.

Strengths and Limitations

A strength of this review is that study methodology was assessed by two raters using the JBI Critical Appraisal Checklist for Analytical Cross-Sectional Study. This checklist included specific questions regarding the validity and reliability of measurement methods, which would cause low-quality articles to be excluded. This review did not include any publication date parameters, meaning this review includes all available studies. This review is not without limitations. No meta-analysis was conducted; therefore, no statistical conclusions regarding these methods can be made. Additionally, articles were only included if they were published in the English language. This could have resulted in articles being excluded that would have been included otherwise. Lastly, due to the lack of identified literature on this topic, the review includes a limited number of studies, makes conclusions difficult to deduce. We also did not

include in our search terms the ability to find methods such as calorimetry or indirect calorimetry, so articles utilizing these methods may exist.

Conclusion

To our knowledge, this is the first review to identify methods of PA assessment in a toddler population. Overall, our findings underscore the need for further research and development of validated methods for assessing PA in toddlers. Future studies should focus on the creation and validation of age-appropriate tools that capture the unique PA behaviors and patterns of toddlers, not just their activity level. Additionally, efforts should be made to standardize measurement protocols and establish best practices for PA assessment in this population to facilitate comparability across studies and enable meaningful interpretation of findings. By addressing these challenges, researchers can advance our understanding of PA in toddlers and develop effective interventions to promote healthy active behaviors from an early age.

REFERENCES

- Ahmadi, M. N., & Trost, S. G. (2022). Device-based measurement of physical activity in pre-schoolers: Comparison of machine learning and cut point methods. *PLOS ONE*, *17*(4), e0266970. <https://doi.org/10.1371/journal.pone.0266970>
- Ahmadi, M., O'Neil, M., Fragala-Pinkham, M., Lennon, N., & Trost, S. (2018). Machine learning algorithms for activity recognition in ambulant children and adolescents with cerebral palsy. *Journal of NeuroEngineering and Rehabilitation*, *15*(1), 105. <https://doi.org/10.1186/s12984-018-0456-x>
- Altenburg, T. M., de Vries, L., op den Buijsch, R., Eyre, E., Dobell, A., Duncan, M., & Chinapaw, M. J. M. (2022). Cross-validation of cut-points in preschool children using different accelerometer placements and data axes. *Journal of Sports Sciences*, *40*(4), 379–385. <https://doi.org/10.1080/02640414.2021.1994726>
- Arem, H., Keadle, S. K., & Matthews, C. E. (2015). Invited Commentary: Meta-Physical Activity and the Search for the Truth. *American Journal of Epidemiology*, *181*(9), 656–658. <https://doi.org/10.1093/aje/kwu472>
- Bianchim, M. S., McNarry, M. A., Larun, L., Barker, A. R., Williams, C. A., & Mackintosh, K. A. (2020). Calibration and validation of accelerometry using cut-points to assess physical activity in paediatric clinical groups: A systematic review. *Preventive Medicine Reports*, *19*, 101142. <https://doi.org/10.1016/j.pmedr.2020.101142>
- Bingham, D., Collings, P., Clemes, S., Costa, S., Santorelli, G., Griffiths, P., & Barber, S. (2016). Reliability and Validity of the Early Years Physical Activity Questionnaire (EY-PAQ). *Sports*, *4*(2), 30. <https://doi.org/10.3390/sports4020030>
- Breau, B., Coyle-Asbil, H. J., & Vallis. (2022). The Use of Accelerometers in Young Children: A Methodological Scoping Review. *Journal for the Measurement of Physical Behavior*, *5*(3), 185–201.
- Browne, M. W. (2000). Cross-Validation Methods. *Journal of Mathematical Psychology*, *44*(1), 108–132. <https://doi.org/10.1006/jmps.1999.1279>
- Bruijns, B. A., Truelove, S., Johnson, A. M., Gilliland, J., & Tucker, P. (2020). 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*, *17*(1), 14. <https://doi.org/10.1186/s12966-020-0912-4>
- Butte, N. F., Wong, W. W., Lee, J. S., Adolph, A. L., Puyau, M. R., & Zakeri, I. F. (2014). Prediction of Energy Expenditure and Physical Activity in Preschoolers. *Medicine & Science in Sports & Exercise*, *46*(6), 1216–1226. <https://doi.org/10.1249/MSS.0000000000000209>

- Cain, K. L., Sallis, J. F., Conway, T. L., Van Dyck, D., & Calhoun, L. (2013). Using Accelerometers in Youth Physical Activity Studies: A Review of Methods. *Journal of Physical Activity and Health, 10*(3), 437–450. <https://doi.org/10.1123/jpah.10.3.437>
- Carson, V., Lee, E.-Y., Hesketh, K. D., Hunter, S., Kuzik, N., Predy, M., Rhodes, R. E., Rinaldi, C. M., Spence, J. C., & Hinkley, T. (2019). Physical activity and sedentary behavior across three time-points and associations with social skills in early childhood. *BMC Public Health, 19*(1), 27. <https://doi.org/10.1186/s12889-018-6381-x>
- Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). *Physical Activity, Exercise, and Physical Fitness: Definitions and Distinctions for Health-Related Research*.
- Colson, E., & Dworkin, P. (1997). *Toddler Development, 18*(8), 255–259.
- Costa, S., Barber, S. E., Cameron, N., & Clemes, S. A. (2014). Calibration and validation of the ActiGraph GT3X+ in 2–3 year olds. *Journal of Science and Medicine in Sport, 17*(6), 617–622. <https://doi.org/10.1016/j.jsams.2013.11.005>
- Evenson KR, Catellier DJ, Gill K, Ondrak KS, & McMurray RG. (2008). Calibration of two objective measures of physical activity for children. *Journal of Sports Sciences, 26*(14), 1557–1565. [rzh. https://doi.org/10.1080/02640410802334196](https://doi.org/10.1080/02640410802334196)
- Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of Accelerometer Output for Children. *Medicine & Science in Sports & Exercise, 37*(11), S523–S530. <https://doi.org/10.1249/01.mss.0000185658.28284.ba>
- Haar, S., Fees, B., Trost, S., Crowe, L. K., & Murray, A. (2013). Design of a Garment for Data Collection of Toddler Language and Physical Activity. *Clothing and Textiles Research Journal, 31*(2), 125–140. <https://doi.org/10.1177/0887302X13478161>
- Hager, E. R., Gormley, C. E., Latta, L. W., Treuth, M. S., Caulfield, L. E., & Black, M. M. (2016). Toddler physical activity study: Laboratory and community studies to evaluate accelerometer validity and correlates. *BMC Public Health, 16*(1), 936. <https://doi.org/10.1186/s12889-016-3569-9>
- Henriksson, H., Forsum, E., & Löf, M. (2014). Evaluation of Actiheart and a 7 d activity diary for estimating free-living total and activity energy expenditure using criterion methods in 1·5- and 3-year-old children. *British Journal of Nutrition, 111*(10), 1830–1840. <https://doi.org/10.1017/S0007114513004406>
- Hidding, L. M., Chinapaw, Mai. J. M., van Poppel, M. N. M., Mokkink, L. B., & Altenburg, T. M. (2018). An Updated Systematic Review of Childhood Physical Activity Questionnaires. *Sports Medicine, 48*(12), 2797–2842. [rzh. https://doi.org/10.1007/s40279-018-0987-0](https://doi.org/10.1007/s40279-018-0987-0)

- Johansson, E., Ekelund, U., Nero, H., Marcus, C., & Hagströmer, M. (2015). Calibration and cross-validation of a wrist-worn Actigraph in young preschoolers: Calibration of Actigraph in toddlers. *Pediatric Obesity, 10*(1), 1–6. <https://doi.org/10.1111/j.2047-6310.2013.00213.x>
- Kelly, L. A., & Villalpando, J. (2016). Development of Actigraph GT1M Accelerometer Cut-Points for Young Children Aged 12-36 Months. *Journal of Athletic Enhancement, 5*(4). <https://doi.org/10.4172/2324-9080.1000233>
- Klesges, L. M., & Klesges, R. C. (1987). *The assessment of children's physical activity: A comparison of methods. 19*(5), 511–517.
- Kwon, S., Honegger, K., & Mason, M. (2019). Daily Physical Activity Among Toddlers: Hip and Wrist Accelerometer Assessments. *International Journal of Environmental Research and Public Health, 16*(21), 4244. <https://doi.org/10.3390/ijerph16214244>
- Lee, E.-Y., Hesketh, K. D., Hunter, S., Kuzik, N., Rhodes, R. E., Rinaldi, C. M., Spence, J. C., & Carson, V. (2017). Meeting new Canadian 24-Hour Movement Guidelines for the Early Years and associations with adiposity among toddlers living in Edmonton, Canada. *BMC Public Health, 17*(S5), 840. <https://doi.org/10.1186/s12889-017-4855-x>
- Lettink, A., Altenburg, T. M., Arts, J., van Hees, V. T., & Chinapaw, M. J. M. (2022). Systematic review of accelerometer-based methods for 24-h physical behavior assessment in young children (0–5 years old). *International Journal of Behavioral Nutrition & Physical Activity, 19*(1), 1–63. <https://doi.org/10.1186/s12966-022-01296-y>
- Li, S., Howard, J. T., Sosa, E. T., Cordova, A., Parra-Medina, D., & Yin, Z. (2020). Calibrating Wrist-Worn Accelerometers for Physical Activity Assessment in Preschoolers: Machine Learning Approaches. *JMIR Formative Research, 4*(8), e16727. <https://doi.org/10.2196/16727>
- Matthews, C. E., Keadle, S. K., Berrigan, D., Lyden, K., & Troiano, R. P. (2021). Influence of Accelerometer Calibration Approach on Moderate–Vigorous Physical Activity Estimates for Adults—Corrigendum. *Medicine & Science in Sports & Exercise, 53*(9), 2018–2018. <https://doi.org/10.1249/MSS.0000000000002669>
- Matthews, C. E., Keadle, S. K., Sampson, J., Lyden, K., Bowles, H. R., Moore, S. C., Libertine, A., Freedson, P. S., & Fowke, J. H. (2013). Validation of a Previous-Day Recall Measure of Active and Sedentary Behaviors. *Medicine & Science in Sports & Exercise, 45*(8), 1629–1638. <https://doi.org/10.1249/MSS.0b013e3182897690>

- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & for the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, 339(jul21 1), b2535–b2535. <https://doi.org/10.1136/bmj.b2535>
- Montoye, A. H. K., Clevenger, K. A., Mackintosh, K. A., McNarry, M. A., & Pfeiffer, K. A. (2019). Cross-Validation and Comparison of Energy Expenditure Prediction Models Using Count-Based and Raw Accelerometer Data in Youth. *Journal for the Measurement of Physical Behaviour*, 2(4), 237–246. <https://doi.org/10.1123/jmpb.2018-0011>
- Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., Currie, M., Qureshi, R., Mattis, P., Lisy, K., & Mu, P.-F. (2020). Chapter 7: Systematic reviews of etiology and risk. In *JBI Manual for Evidence Synthesis*.
- Oftedal, S., Bell, K. L., Davies, P. S. W., Ware, R. S., & Boyd, R. N. (2014). Validation of Accelerometer Cut Points in Toddlers with and without Cerebral Palsy. *Medicine & Science in Sports & Exercise*, 46(9), 1808–1815. <https://doi.org/10.1249/MSS.0000000000000299>
- Pate, R. R., Almeida, M. J., McIver, K. L., Pfeiffer, K. A., & Dowda, M. (2006). Validation and Calibration of an Accelerometer in Preschool Children*. *Obesity*, 14(11), 2000–2006. <https://doi.org/10.1038/oby.2006.234>
- Pereira, J. R., Sousa-Sá, E., Zhang, Z., Cliff, D. P., & Santos, R. (2020). Concurrent validity of the ActiGraph GT3X+ and activPAL for assessing sedentary behaviour in 2-3-year-old children under free-living conditions. *Journal of Science & Medicine in Sport*, 23(2), 151–156. <https://doi.org/10.1016/j.jsams.2019.08.009>
- Pfeiffer, K. A., McIver, K. L., Dowda, M., Almeida, M. J. C. A., & Pate, R. R. (2006). Validation and Calibration of the Actical Accelerometer in Preschool Children. *Medicine & Science in Sports & Exercise*, 38(1), 152–157. <https://doi.org/10.1249/01.mss.0000183219.44127.e7>
- Pulakka, A., Cheung, Y., Ashorn, U., Penpraze, V., Maleta, K., Phuka, J., & Ashorn, P. (2013). Feasibility and validity of the ActiGraph GT3X accelerometer in measuring physical activity of Malawian toddlers. *Acta Paediatrica*, 102(12), 1192–1198. <https://doi.org/10.1111/apa.12412>
- Saint-Maurice, P. F., Welk, G. J., Bartee, R. T., & Heelan, K. (2017). Calibration of context-specific survey items to assess youth physical activity behaviour. *Journal of Sports Sciences*, 35(9), 866–872. <https://doi.org/10.1080/02640414.2016.1194526>
- Sarker, H., Anderson, L., Borkhoff, C., Abreo, K., Tremblay, M., Lebovic, G., Maguire, J., Parkin, P., & Birken, C. (2015). Validation of Parent-Reported Physical and Sedentary Activity by

- Accelerometry in Young Children. *Canadian Journal of Diabetes*, 39, S44–S44. rzh.
<https://doi.org/10.1016/j.jcjd.2015.01.169>
- Sirard, J. R., & Pate, R. R. (2001). Physical Activity Assessment in Children and Adolescents. *Sports Med.*
- Sirard, J. R., Trost, S. G., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2005). Calibration and Evaluation of an Objective Measure of Physical Activity in Preschool Children. *Journal of Physical Activity and Health*, 2(3), 345–357. <https://doi.org/10.1123/jpah.2.3.345>
- Timmons, B. W., LeBlanc, A. G., Carson, V., Connor Gorber, S., Dillman, C., Janssen, I., Kho, M. E., Spence, J. C., Stearns, J. A., & Tremblay, M. S. (2012). Systematic review of physical activity and health in the early years (aged 0–4 years). *Applied Physiology, Nutrition, and Metabolism*, 37(4), 773–792. <https://doi.org/10.1139/h2012-070>
- Trost, S. G. (2007). State of the Art Reviews: Measurement of Physical Activity in Children and Adolescents. *American Journal of Lifestyle Medicine*, 1(4), 299–314.
<https://doi.org/10.1177/1559827607301686>
- Trost, S. G., Fees, B. S., Haar, S. J., Murray, A. D., & Crowe, L. K. (2012). Identification and Validity of Accelerometer Cut-Points for Toddlers. *Obesity*, 20(11), 2317–2319.
<https://doi.org/10.1038/oby.2011.364>
- Trost, S. G., Mciver, K. L., & Pate, R. R. (2005). Conducting Accelerometer-Based Activity Assessments in Field-Based Research. *Medicine & Science in Sports & Exercise*, 37(11), S531–S543. <https://doi.org/10.1249/01.mss.0000185657.86065.98>
- Tulve, N. S., Jones, P. A., McCurdy, T., & Croghan, C. W. (2007). A Pilot Study Using an Accelerometer to Evaluate a Caregiver’s Interpretation of an Infant or Toddler’s Activity Level as Recorded in a Time Activity Diary. *Research Quarterly for Exercise & Sport*, 78(4), 375–383. trh.
- van Cauwenberghe E, Labarque V, Trost SG, de Bourdeaudhuij I, & Cardon G. (2011). Calibration and comparison of accelerometer cut points in preschool children. *International Journal of Pediatric Obesity*, 6(2–2), e582–9. rzh.
<https://doi.org/10.3109/17477166.2010.526223>

CHAPTER 4: DEVELOPMENT AND TESTING OF THE OBSERVATIONAL SYSTEM FOR RECORDING PHYSICAL ACTIVITY IN CHILDREN – TODDLERS

Abstract

Direct observation tools developed for preschoolers have been used to describe toddler physical activity (PA); however, authors have concluded that these tools are not inclusive to toddler-specific behavior. No direct observation system that can be used to assess both PA levels and the childcare environment in toddlers exists. **Objective:** The purpose of this study was to develop the Observational System for Recording Physical Activity in Children - Toddlers (OSRAC-T) and assess the interrater reliability of the newly developed tool. **Methods:** This tool is an extension of the Observational System for Recording Physical Activity in Children – Preschool. Tool content was established through identifying similar research, consulting with experts, and conducting informal observations. Reliability was assessed in a sample of toddlers (12 – 36 months) that attended one of three childcare centers. Video data were recorded and analyzed later using a focal child, time sampling system (5-second observation, 25-second recording). Interrater reliability was assessed in 39% of observations. **Results:** Thirty-one toddlers were included in the study (25.5 ± 6.0 months). The final instrument included nine categories that described physical activity level and type, social and environmental context, and support relevant to toddlers. Observers completed 124 observation sessions resulting in 7,757 30-second observation intervals. Interval-by-interval agreement was moderate to high (59.0 – 95.3%) for all categories and kappa values were moderate (0.46-0.69). **Conclusion:** The OSRAC-T is a reliable observation system to assess the PA behaviors of toddlers. This instrument can be used to measure behaviors relevant to toddlers to better inform early childcare center design,

or to inform future intervention studies. This tool could also be used to assess correlates or relationships between PA behavior and health outcomes in toddlers.

Introduction

Despite increasing attention in recent years, childhood obesity continues to be a nationwide epidemic. According to the Centers for Disease Control and Prevention, the prevalence of obesity for youth ≤ 19 years old was 19.7% between 2017 and 2020 while 12.7% of 2 – 5-year-olds were impacted by obesity (Centers for Disease Control and Prevention, 2022). Childhood obesity is associated with numerous negative health outcomes, both physical and psychological (Lee et al., 2017). It has been linked to an increased risk of cardiovascular disease, high blood pressure, anxiety, and depression in adulthood (Karnik & Kanekar, 2012). Although obesity is a complex disease, it is well known that obesity is linked to a caloric intake without a matched or higher caloric expenditure. This imbalance can be related to nutritional intake or physical inactivity. Specifically, higher levels of physical activity (PA) have been associated with a lower risk of obesity (Centers for Disease Control and Prevention, 2022). Due to the benefits of PA, it is imperative that young children are engaging in sufficient amounts.

Toddlers spend 55 – 65% of the day sedentary, suggesting that there is plenty of sedentary time (ST) that could be converted to physically active time (Borkhoff et al., 2015; Gubbels et al., 2011; Johansson et al., 2015). Sources have estimated that over half (59%) of young children (1 – 5 years old) attend a childcare center, making this an ideal setting for PA interventions (Corcoran & Steinley, 2017). Previous research proposes that childcare centers are a strong predictor of PA and that opportunities (e.g., large spaces, portable equipment) or locations (e.g., blocks, art, or manipulative areas) within the classroom or outdoors may encourage higher PA levels in preschoolers (Clevenger et al., 2022; Finn et al., 2002; McWilliams et al., 2009). These factors can then be used to inform childcare center classroom or

playground design in ways that promote PA. However, similar research needs to be conducted in toddlers.

Because toddlers have different motor and cognitive skills compared to preschoolers (Colson & Dworkin, 1997), there is a need for toddler specific PA measures. While most young children start walking independently by 18 months, those aged 12 to 18 months may require assistance or might not be able to walk at all. Activities like hopping, kicking, or jumping typically emerge between 24 and 36 months. Moreover, finer motor abilities such as block-building, writing, or drawing usually develop around the age of 36 months (Colson & Dworkin, 1997). These notable developmental variations imply that evaluating toddler activities may necessitate a different approach than assessing those of preschoolers. Thus, it is imperative that we can accurately assess PA levels and context in toddlers in the childcare context.

PA can be assessed in a multitude of ways (Sirard & Pate, 2001; Trost, 2007). Methods include survey-based instruments, such as questionnaires or recall diaries, accelerometry, or direct observation (DO). DO is considered the criterion measure of PA (Trost, 2007). DO requires a trained researcher to observe and code participant behavior using a specific tool. This technique allows for additional environmental and social contexts of activity to be accounted for in PA assessment. Several observational tools exist for use in children (Brown et al., 2006; McKenzie et al., 2002; McKenzie, Sallis, & Nader, 1991; McKenzie, Sallis, Nader, et al., 1991; Puhl et al., 1990). One of these tools is the Observational System for Recording Physical Activity in Children – Preschool (OSRAC – P) (Brown et al., 2006).

The OSRAC-P is an observational tool developed in preschoolers to assess PA behavior. The tool includes the following categories: (1) physical activity level (2) physical activity type (3)

location (4) indoor educational/play context (5) outdoor/gym educational/play context (6) initiator of activity (7) group composition (8) prompt for physical activity. The OSRAC-P is a focal-child system, meaning that one child is the focus of each observation, and decisions made about coding refer to that single child (Brown et al., 2006).

In toddlers, the OSRAC-P has been utilized to assess PA (Dinkel et al., 2019; Fees et al., 2015; Gubbels et al., 2011; Van Cauwenberghe et al., 2011). In these studies, only specific categories (i.e., activity level and type indoor) were assessed; no study has assessed all categories of PA simultaneously. Additionally, when assessing PA, two of the four studies concluded that the OSRAC-P would have to be adapted to be used in a toddler population (Dinkel et al., 2019; Fees et al., 2015). In these studies, additional location categories or activity codes and contexts were created. Fees et al. (2015) reported that the added activity context, onlooking, was more common among toddlers than some of the existing activity contexts. Although these studies proposed changes to the OSRAC-P, they did not create an adapted observational tool for use by future researchers. Therefore, this study's purpose is to develop an Observational System for Recording Physical Activity in Children - Toddlers (OSRAC-T) and assess the reliability of the newly developed tool. We hypothesized that that OSRAC-T would be reliable in assessing toddler PA.

Methods

Development of the OSRAC-T

The OSRAC-T was created as an extension of the collection of OSRAC direct observation systems (McIver et al., 2009, 2016), which is a focal-child system in which one child is the focus of each observation, and decisions made about coding refer to that single child (Brown et al.,

2006). The OSRAC-T was developed using the OSRAC-P as a guide and aimed to collect information on the physical and social environmental contexts of physical activity specific to toddlers. The goal was to include certain categories that were consistent across OSRAC tools (e.g., physical activity levels), but to include new categories or codes specific to toddlers and childcare settings. During development, similar research was referenced, and relevant codes were created or extracted.

Additionally, in a previous study conducted by our research team, the OSRAC-P was used to describe physical activity behavior in a group of toddlers (unpublished). Thirty toddlers were observed at childcare centers or at home for at least 90 minutes, both indoor and outdoor. When analyzing video data, notes were taken identifying behavior codes that were not present in the OSRAC-P but may be relevant to toddlers. These notes were used to make modifications to the observation tool. Once these codes and categories were created, the tool was informally tested during observations at other non-participating schools to ensure completeness of capture for behaviors and settings. Researchers obtained classroom schedules and observed normal classroom behaviors that occurred throughout the school day. Based on these preliminary observations, a final version of the OSRAC-T was developed for testing.

Testing of the OSRAC-T

To test the OSRAC-T, families were recruited from three childcare centers in the Greater Lansing and East Lansing, Michigan area in the United States. Potential childcare centers were contacted via phone or email to ask if they were interested in participating as a research site. Upon agreement to become a site, leadership at the childcare center was asked to send informational emails to parents of toddlers (12 – 36 months). Emails included information

regarding the study, in addition to links to an eligibility survey. If found eligible, parents were invited to enroll their child in the present study using an online consent form. Eligibility and consent were obtained using REDCap unless it was not feasible, then a paper consent form was provided. This study was approved by the Michigan State Institutional Review Board.

To be eligible for the present study, toddlers needed be 12-36 months old and attend one of the enrolled childcare centers. No additional criteria were needed for inclusion. To be comparable to other OSRAC adaptations (Cohen et al., 2014; McIver et al., 2009; Schenkelberg et al., 2021) we expected a reliability (ICC) of 0.7 ± 0.2 (CI: 0.95, $k=2$). To achieve this and account for a ~10% dropout rate, we needed twenty-nine toddlers to achieve a minimum sample size of twenty-six toddlers (Arifin, 2024).

Toddlers were videorecorded at childcare centers on two separate occasions for at least 60 minutes each. During each visit, two 30-minute observations were completed, one indoor and one outdoor. Each toddler was recorded during four observations, resulting in 120 minutes of observation per toddler. Children were recorded using a GoPro Hero5 Session video camera (GoPro, San Mateo, CA). Video was collected at 1080p using a wide view at 30 frames/s. Universal time was synchronized to the camera using the website time.is. Two to three cameras were placed in optimal settings to capture the best view of each classroom or playground. Data were collected during scheduled free play at each center; therefore, observations including snack and nap times were avoided.

While the OSRAC-P was originally designed to be done in-person, recent research has used video because one can pause, rewind, or re-watch observations. Therefore, a subset of toddlers ($n=10$) was randomly assigned to be live coded during video collection. The app

Cybertracker (Cape town, South Africa) was used on a mobile cellular device and the OSRAC-T was created as a project in the app. Time was synced using time.is, showing the universal time to the cameras. Another mobile app was used to inform coders when to start observing and when to start recording. These sounds were able to be heard in all recorded videos. Live coding was done by only one rater. This was to ensure that the OSRAC-T could be used during live sessions as the rest of the OSRAC tools are. Photos were taken of each toddler holding their ID number so that PA assessment could be conducted for the correct toddlers. Participating families received \$30 at the conclusion of the study.

Data Analysis

Descriptive statistics were performed to assess demographic data. Collected videos were downloaded from the GoPro devices and uploaded to Behavioral Observation Research Interactive Software (BORIS, Torino, Italy). BORIS is an event logging software that allows for the coding of video observations. Once in BORIS, the video was coded using a 5-second observation, followed by a 25-second record interval. The OSRAC-P was developed for in-person coding, so it utilizes an observe/record cycle. To maintain consistency with the OSRAC-P, ensure other researchers were still able to use the OSRAC-T in-person, and to reduce time and burden, an observe/record cycle was used even though video observations were being collected. As in the OSRAC-P, the highest level of physical activity in the 5-second observation period was coded. The additional categories were then coded in relation to that activity intensity. Only one code per category was determined for each observation interval.

Each video was coded by one of two trained coders. Prior to data collection, coders engaged in a training protocol which included reviewing the training manual and completing

four hours of video observation and coding practice. Discussions were held after each practice session. Coding occurred until IOA was at least 80% in all categories (Brown et al., 2006).

Observation intervals where “Cannot Tell” was coded (6.3% of all intervals) were removed from this analysis resulting in 7,265 intervals (3,699 indoor and 3,566 outdoor). Once removed, the percentage of intervals of each code in each category was described and reported. ST was defined as activities coded as 1 and 2, light (LPA) 3, and moderate-to-vigorous (MVPA) 4 and 5. Additionally, the percentage of intervals spent in each activity level was described for specific categories of environmental and social factors (i.e., activity type, indoor context, outdoor context, group composition, etc.). No statistical comparisons were conducted for these variables.

Inter-rater reliability for two observers coding the same videos was assessed for 39% of intervals. This was similar to previous adaptations of OSRAC systems (Cohen et al., 2014; McIver et al., 2009; Schenkelberg et al., 2021). Interobserver agreement was assessed for two observers using kappa coefficients, Cohen’s weighted kappa coefficients (Cohen, 1960), and percent agreement. Weighted kappa is a more acceptable measure for assessing ordinal variables, so it was used to assess PA level. To determine the percentage agreement, interval-by-interval agreement was assessed. The total numbers of agreements within a category were divided by the sum of agreements and disagreements for that category. This ratio was then multiplied by 100 to get percent agreement (Berk, 1979). Agreement was determined if the two observers coded the same code within a category for a given interval. If the two observers coded different codes within a category, this was considered a disagreement. The reliability of the OSRAC-T was considered acceptable if percent agreement was at least 70% and kappa was

classified as poor ($k = \leq 0.40$), moderate ($> 0.40 - \leq 0.60$), or strong (> 0.60). A kappa value of 0.60 was considered acceptable for the purpose of this study. These variables were also used to assess the intra-rater reliability between live coded sessions and video coding for one coder. All analyses were performed using SPSS 28.0 for Windows.

Results

Thirty-one toddlers were included in the study (Table 5).

Table 5. Participant Demographics

Gender Female, n (%)	13 (41.9)
Age, months (SD)	25.47 (6.0)
Length, cm (SD)	86.43 (5.8)
Weight, kg (SD)	12.92 (1.5)
Race/Ethnicity, n (%)	
American Indian or Alaska Native	1 (3.2)
Asian	2 (6.5)
Black or African American	3 (9.7)
Hispanic or Latino	1 (3.2)
Native Hawaiian or Pacific Islander	1 (3.2)
White/Caucasian	19 (61.3)
Other	1 (3.2)
Prefer not to Answer	3 (9.7)

SD; standard deviation

The final version of the OSRAC-T was comprised of nine coding categories: 1) Physical Activity Level, 2) Physical Activity Type, 3) Location, 4) Indoor/Educational Context, 5) Outdoor Play Context, 6) Group Composition, 7) Physical Activity Prompt, 8) Activity Initiator, and 9) Support. Table 6 describes the additional categories and codes used in the OSRAC-T. A complete list of categories and codes can be found in Appendix B.

Table 6. OSRAC-T Novel Categories and Codes

Category and Codes	Definition
Indoor Educational/Play Context	Tantrum - when a focal child is experiencing an uncontrolled outburst of emotion. Can be anger, frustration, etc.
Outdoor/Gym Education/Gym Context	Preacademic - when a focal child is engaging in pre-reading, pre-writing, or preacademic activities. When a focal child is located in a center containing books, writing, listening, science or math materials Tantrum - when a focal child is experiencing an uncontrolled outburst of emotion. Can be anger, frustration, etc.
Activity Type	Creep/Crawl/Scoot - crawling, creeping, scooting. Refers to a child translocating on their hands and knees (creeping), on their stomach (crawling), or while sitting down (scooting). Could include bear-crawling.
Support	Non-Weight Bearing - no weight or force being placed. Includes riding in strollers and other wheel devices or being carried Partial Weight Bearing - partial weight or force being placed. May include use of a supportive device such as pull/push devices Weight Bearing - full force or weight being places Cannot Tell Support - cannot tell support

There were 124 observation sessions which resulted in 7,757 30-s observational intervals (3,851 outdoor and 3,906 indoor). Inter-rater reliability was assessed during 48 observation sessions (38.7% of sessions), resulting in 2,814 observation intervals. Kappa values were moderate to strong in most categories (0.460-0.693) except activity initiation (0.349) and PA prompt (0.276). Percent agreement was greater than 70% in most categories of the OSRAC-T (76.9-95.3%) except PA level and Indoor activity context, which were 65.1% and 58.9% respectively. Average kappa and percent agreement, and standard deviations are presented in Table 7.

Table 7. Average Percent Agreement Between Raters

		Mean	SD
Physical Activity Level	Kappa	0.608	0.133
	Percent Agreement	0.651	0.116
Physical Activity Type	Kappa	0.693	0.094
	Percent Agreement	0.769	0.073
Indoor Activity Context	Kappa	0.488	0.128
	Percent Agreement	0.589	0.117
Outdoor Activity Context	Kappa	0.687	0.094
	Percent Agreement	0.782	0.082
Activity Initiator	Kappa	0.349	0.267
	Percent Agreement	0.930	0.058
Group Composition	Kappa	0.460	0.201
	Percent Agreement	0.823	0.123
Physical Activity Prompt	Kappa	0.276	0.232
	Percent Agreement	0.913	0.081
Support	Kappa	0.557	0.325
	Percent Agreement	0.953	0.048

There were 26 observation sessions that were both live and video coded. This resulted in 1,578 30-s observational intervals (20% of intervals). Kappa values were moderate-to-strong in all categories (0.465-0.831). Percent agreement was greater than 70% in all categories (74.6-98.7%). Only 1-3% of intervals across categories were coded as “cannot tell” during video coding, but not during live coding. Average kappa and percent agreement between live and video coding methods can be found in Table 8.

Table 8. Average Percent Agreement among Live and Video Coding

		Mean	SD
Physical Activity Level	Kappa	0.633	0.194
	Percent Agreement	0.746	0.103
Physical Activity Type	Kappa	0.796	0.078
	Percent Agreement	0.867	0.072
Indoor Activity Context	Kappa	0.744	0.142
	Percent Agreement	0.820	0.114
Outdoor Activity Context	Kappa	0.830	0.126
	Percent Agreement	0.919	0.072
Activity Initiator	Kappa	0.528	0.393
	Percent Agreement	0.931	0.088
Group Composition	Kappa	0.465	0.254
	Percent Agreement	0.851	0.113
Physical Activity Prompt	Kappa	0.814	0.199
	Percent Agreement	0.987	0.025
Support	Kappa	0.831	0.376
	Percent Agreement	0.972	0.025

Overall, toddlers spent 70.0% of time in sedentary time and 10.0% of time engaged in MVPA while at childcare (Table 9).

Table 9. Observed OSRAC-T Codes and Percentages of Intervals by Activity Level

Categories, Observed Codes	Observed Intervals (n)	Sedentary (%)	Light (%)	MVPA (%)
Total Observed Intervals	7,265	70.0	19.0	10.0
Activity Type				
Climb	31	19.0	19.0	61.0
Crawl/Creep/Scoot	21	52.0	38.0	10.0
Dance	50	24.0	48.0	28.0
Jump	23	00.0	22.0	78.0
Lie Down	76	100.0	0.0	0.0
Pull/Push	114	6.0	42.0	52.0
Ride	250	2.0	48.0	51.0
Rock	19	5.0	16.0	79.0
Roll	5	0.0	80.0	20.0
Rough & Tumble	4	25.0	75.0	0.0
Run	316	0.0	00.0	100.0
Sit/Squat	2,707	100.0	00.0	0.0
Stand	2,290	99.0	01.0	0.0
Throw	59	47.0	27.0	25.0
Walk	1,255	0.0	87.0	13.0
Indoor Educational Contexts				
Art	1	0.0	100.0	0.00
Books/Pre-Academic	392	92.0	8.0	1.0
Gross Motor	320	65.0	21.0	13.0
Group Time	398	94.0	4.0	2.0
Large Blocks	29	66.0	31.0	3.0
Manipulative	865	91.0	7.0	1.0
Music	158	65.0	23.0	12.0
Nap	4	100.0	0.0	0.0
Other	173	88.0	10.0	2.0
Self-Care	28	96.0	0.0	4.0
Snacks	69	94.0	6.0	0.0
Sociodramatic	1,016	79.0	19.0	3.0
Tantrum	35	97.0	3.0	0.0
Time Out	1	100.0	0.0	0.0
Transition	208	8.0	75.0	17.0
Videos	2	100.0	0.0	0.0
Outdoor Play Context				
Ball	255	46.0	27.0	27.0
Fixed	386	65.0	17.0	18.0

Table 9 (cont'd)

	Open Space	1,126	46.0	34.0	21.0
	Other	4	75.0	25.0	0.0
	Portable	94	82.0	17.0	1.0
	Pre-Academic	1	0.0	100.0	0.0
	Sandbox	161	93.0	5.0	2.0
	Snacks	10	60.0	30.0	10.0
	SocioProps	258	65.0	22.0	13.0
	Tantrum	13	100.0	0.0	0.0
	Wheel	1,258	68.0	17.0	16.0
Activity Initiation					
	Adult	1,013	82.0	13.0	6.0
	Child	6,252	69.0	20.0	11.0
Prompt					
	None	7,258	70.0	19.0	10.0
	Peer Prompt - Increase PA	2	0.0	0.0	100.0
	Teacher Prompt - Decrease PA	1	0.0	100.0	0.0
	Teacher Prompt - Increase PA	4	0.0	75.0	25.0
Support					
	Non-Weight Bearing	461	100.0	0.0	0.0
	Partial Weight Bearing	1	0.0	100.0	0.0
	Weight-Bearing	6,803	68.0	20.0	11.0
Group					
	1-1 Adult	309	77.0	17.0	6.0
	1-1 Peer	178	81.0	11.0	8.0
	Group Adult	6,156	70.0	19.0	10.0
	Group Child	326	63.0	20.0	17.0
	Solitary	296	66.0	24.0	10.0

The most observed types of physical activity were sit/squat (37.3%), stand (31.5%), and walk (17.3%). Indoor, the most common educational or play contexts were sociodramatic (27.5%) and manipulative (23.4%), followed by group time, books, and gross motor (8.7 – 10.8%). Time spent in these contexts was primarily sedentary (65.3 - 94.5%). Over half of outdoor play sessions were spent engaging with wheel toys (35.3%) or in open space (31.6%), followed by fixed equipment (10.8%). Again, time spent in these contexts was primarily

sedentary (64.8 - 67.7%), except open space which was spent primarily in light or MVPA (54.1%). Most activity was child initiated (86.1%), done without prompting (99.9%), and in a group with an adult (84.7%). Support was primarily weight bearing (93.6%). Non-weight bearing (NWB) occurred only 6.3% of the time. NWB was more common outdoor (97.2% of NWB intervals) and of these intervals, 94.2% occurred during the wheel context.

Discussion

Overall, the OSRAC-T is a reliable tool for assessing physical activity behavior in toddler aged children. There was moderate to strong inter-rater reliability in all categories except for activity initiation and PA prompts. Reliability in this study was lower compared to other variations of the original OSRAC tool (Brown et al., 2006; McIver et al., 2009; Schenkelberg et al., 2021). However, a lower level of reliability in activity initiation and PA prompt categories was reported across studies. Similar to previous adaptations, most of the disagreement from PA level occurred with discrepancies between coding 1-Stationary and 2-Stationary with limb or trunk movement. As previously mentioned, these categories were combined to determine sedentary time, making this disagreement less concerning. In this study PA prompts were agreed on in over 90% of intervals. As stated by the authors of the original OSRAC observational tool, low levels of variation in PA prompt make it difficult for kappa values to be calculated (Brown et al., 2006).

Levels of agreement in indoor and outdoor context were lower in this study compared to previous studies, especially the lower reliability of indoor activity context (0.488 v. 0.90 – 0.99; McIver et al., 2009; McIver et al., 2016; Schenkelberg et al., 2021). This may be due to the different motor and play development of toddlers compared to preschool-aged children (3 – 5

years old) who were the focus of prior reliability studies. For example, a previous study assessing play in children 2 – 5 years found that children 2 – 3 years commonly participated in idly looking or idly sitting (onlooking), meaning that younger children act as spectators rather than engaging (Parten, 1993). More recently, Fees et al. (2015) observed that 15% of toddlers' time was spent engaging in onlooking behavior. This disengagement may have led to discrepancies in categorizing indoor activity context. Specifically, when indoors, transition was frequently coded by one rater (during onlooking) while a different activity context was coded by the other. To address the lower levels of reliability between coders in this OSRAC tool compared to the others, changes will be made to the descriptions in the training manual. For example, it will be clarified that although toddlers are engaged in onlooking, the context of activity should be coded based on the education center or location that the toddler is if not engaged with any objects.

Consistent with other studies assessing PA in toddlers, children in our sample spent most of their time sedentary. Only about 10% of intervals were spent engaging in MVPA. The collection of OSRAC tools allow for the assessment of contextual information regarding physical activity behavior; however, as noted, toddlers are in different developmental states compared to preschoolers. Independent walking is met by most children by 18-months; however, children with developmental delays or children 12 – 18 months may not have been walking yet (Colson & Dworkin, 1997). Therefore, a support category was added to the OSRAC-T to account for toddlers' need full or partial support. These types of assistance accounted for about 7% of intervals. Additionally, toddlers are in a stage where they are developing temperament and socio-emotional skills such as the desire for autonomy and impulse control (Colson & Dworkin,

1997; Soini et al., 2014). The struggle to regulate these emotions results in temper tantrums, or outbursts of intense emotion. A tantrum context was added to the OSRAC-T to quantify this behavior and was coded during about 1.0% of intervals. Many of these changes are supported by previous observation studies (Fees et al., 2015; Soini et al., 2014).

There are both strengths and limitations to the current study. The additional category and codes provided in the OSRAC-T allow for a more comprehensive assessment of toddlers' physical activity behavior that could not be captured using the OSRAC-P. However, by adapting the OSRAC-P, rather than creating a brand-new observational tool, information regarding typical early childcare settings are included. This study is not without limitations. First, observation only occurred during free-play at the childcare center. This means that toddler specific behaviors that do not occur during this time may have been missed. The results of this study come from a small sample (N=31) that may not be representative of all toddlers. Our sample was primarily white/Caucasian, and all toddlers came from a family where the parent reported completing at least some college or technical school. Similar to other observation tools, another limitation is the 5-s observation, 25-s record interval that was used to mirror the original OSRAC tool. This results in estimations of physical activity levels and may not capture the sporadic bouts of activity in which young children are known to engage or total levels of physical activity. Lastly, inter-rater reliability was only examined for video-coded observations rather than live-coded sessions. This is different than other adaptations of the OSRAC system which utilized live coding sessions only (Brown et al., 2006; McIver et al., 2009; Schenkelberg et al., 2021). Therefore, it is unknown if this tool would be reliable if used for live coding.

Conclusion

The OSRAC-T is a reliable observation system to assess the physical activity behaviors of toddlers. Physical activity level, activity type, indoor/outdoor activity context, activity initiation, physical activity prompt, group composition, and support were recorded during 30-minutes of free play, both indoor and outdoor at early childcare centers. This adaptation allows for the unique behaviors of toddlers to be accounted for and can be used in future studies to better assess toddler physical activity. Future research should aim to use this instrument in more underrepresented populations and communities (i.e. low socioeconomic status). This instrument can be used to determine behaviors unique to toddlers to better inform early childcare center design or to inform future intervention studies that may require different approaches than those taken with preschoolers. Identifying toddler specific design methods may result in better developmental trajectories across childhood. This tool could also be used to assess correlates or relationships between physical activity behavior and health outcomes in toddlers.

REFERENCES

- Arifin, W. N. (2024). Sample size calculator (web). Retrieved from <http://wnarifin.github.io>
- Berk, R. A. (1979). Generalizability of behavioral observations: A clarification of interobserver agreement and interobserver reliability. *American Journal of Mental Deficiency, 83*(5), 460–472.
- Borkhoff, C. M., Heale, L. D., Anderson, L. N., Tremblay, M. S., Maguire, J. L., Parkin, P. C., & Birken, C. S. (2015). Objectively measured physical activity of young Canadian children using accelerometry. *Applied Physiology, Nutrition, and Metabolism, 40*(12), 1302–1308. <https://doi.org/10.1139/apnm-2015-0164>
- Brown, W. H., Pfeiffer, K. A., McIver, K. L., Dowda, M., Almeida, J. M. C. A., & Pate, R. R. (2006). Assessing Preschool Children's Physical Activity: The Observational System for Recording Physical Activity in Children-Preschool Version. *Research Quarterly for Exercise and Sport, 77*(2), 167–176. <https://doi.org/10.1080/02701367.2006.10599351>
- Centers for Disease Control and Prevention. (2022). *Prevalence of Childhood Obesity in the United States*. U.S. Department of Health and Human Service. <https://www.cdc.gov/obesity/data/childhood.html>
- Clevenger, K. A., McKee, K. L., & Pfeiffer, K. A. (2022). Classroom Location, Activity Type, and Physical Activity During Preschool Children's Indoor Free-Play. *Early Childhood Education Journal, 50*(3), 425–434. <https://doi.org/10.1007/s10643-021-01164-7>
- Cohen, A., McDonald, S., McIver, K., Pate, R., & Trost, S. (2014). Assessing Physical Activity During Youth Sport: The Observational System for Recording Activity in Children: Youth Sports. *Pediatric Exercise Science, 26*(2), 203–209. <https://doi.org/10.1123/pes.2013-0095>
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 20*(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Colson, E., & Dworkin, P. (1997). *Toddler Development, 18*(8), 255–259.
- Corcoran, L., & Steinley, K. (2017). *Early Childhood Program Participation, Results from the National Household Education Surveys Program of 2016*.
- Dinkel, D., Snyder, K., Patterson, T., Warehime, S., Kuhn, M., & Wisneski, D. (2019). An exploration of infant and toddler unstructured outdoor play. *European Early Childhood Education Research Journal, 27*(2), 257–271. <https://doi.org/10.1080/1350293X.2019.1579550>

- Fees, B. S., Fischer, E., Haar, S., & Crowe, L. K. (2015). Toddler Activity Intensity During Indoor Free-Play: Stand and Watch. *Journal of Nutrition Education and Behavior, 47*(2), 170–175. <https://doi.org/10.1016/j.jneb.2014.08.015>
- Finn, K., Johannsen, N., & Specker, B. (2002). Factors associated with physical activity in preschool children. *The Journal of Pediatrics, 140*(1), 81–85. <https://doi.org/10.1067/mpd.2002.120693>
- Gubbels, J. S., Kremers, S. P., Stafleu, A., de Vries, S. I., Goldbohm, R. A., Dagnelie, P. C., de Vries, N. K., Buuren, S. van, & Thijs, C. (2011). Association between parenting practices and children’s dietary intake, activity behavior and development of body mass index: The KOALA Birth Cohort Study. *International Journal of Behavioral Nutrition and Physical Activity, 8*(1 pp.414–414), 414–414. [agr. https://doi.org/10.1186/1479-5868-8-18](https://doi.org/10.1186/1479-5868-8-18)
- Johansson, E., Hagströmer, M., Svensson, V., Ek, A., Forssén, M., Nero, H., & Marcus, C. (2015). Objectively measured physical activity in two-year-old children – levels, patterns and correlates. *International Journal of Behavioral Nutrition and Physical Activity, 12*(1), 3. <https://doi.org/10.1186/s12966-015-0161-0>
- Karnik, S., & Kanekar, A. (2012). Childhood Obesity: A Global Public Health Crisis. *International Journal of Preventive Medicine, 3*(1).
- Lee, E.-Y., Hesketh, K. D., Hunter, S., Kuzik, N., Rhodes, R. E., Rinaldi, C. M., Spence, J. C., & Carson, V. (2017). Meeting new Canadian 24-Hour Movement Guidelines for the Early Years and associations with adiposity among toddlers living in Edmonton, Canada. *BMC Public Health, 17*(S5), 840. <https://doi.org/10.1186/s12889-017-4855-x>
- McIver, K. L., Brown, W. H., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2009). ASSESSING CHILDREN’S PHYSICAL ACTIVITY IN THEIR HOMES: THE OBSERVATIONAL SYSTEM FOR RECORDING PHYSICAL ACTIVITY IN CHILDREN-HOME. *Journal of Applied Behavior Analysis, 42*(1), 1–16. <https://doi.org/10.1901/jaba.2009.42-1>
- McIver, K. L., Brown, W. H., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2016). Development and Testing of the Observational System for Recording Physical Activity in Children: Elementary School. *Research Quarterly for Exercise and Sport, 87*(1), 101–109. <https://doi.org/10.1080/02701367.2015.1125994>
- McKenzie, T. L., Marshall, S. J., Sallis, J. F., & Conway, T. L. (2002). *System for Observing Play and Leisure Activity in Youth* [dataset]. American Psychological Association. <https://doi.org/10.1037/t72617-000>
- McKenzie, T. L., Sallis, J. F., & Nader, P. R. (1991). *SOFIT: System for Observing Fitness Instruction Time. 11*, 195–205.

- McKenzie, T. L., Sallis, J. F., Nader, P. R., Patterson, T. L., Elder, J. P., Berry, C. C., Rupp, J. W., Atkins, C. J., Buono, M. J., & Nelson, J. A. (1991). BEACHES: AN OBSERVATIONAL SYSTEM FOR ASSESSING CHILDREN'S EATING AND PHYSICAL ACTIVITY BEHAVIORS AND ASSOCIATED EVENTS. *Journal of Applied Behavior Analysis, 24*(1), 141–151.
<https://doi.org/10.1901/jaba.1991.24-141>
- McWilliams, C., Ball, S. C., Benjamin, S. E., Hales, D., Vaughn, A., & Ward, D. S. (2009). Best-Practice Guidelines for Physical Activity at Child Care. *Pediatrics, 124*(6), 1650–1659.
<https://doi.org/10.1542/peds.2009-0952>
- Parten, M. B. (1993). *SOCIAL PLAY AMONG PEESCHOOL CHILDKEN.*
- Puhl, J., Greaves, K., Hoyt, M., & Baranowski, T. (1990). Children's Activity Rating Scale (CARS): Description and Calibration. *Research Quarterly for Exercise and Sport, 61*(1), 26–36.
<https://doi.org/10.1080/02701367.1990.10607475>
- Schenkelberg, M. A., Brown, W. H., McIver, K. L., & Pate, R. R. (2021). An observation system to assess physical activity of children with developmental disabilities and delays in preschool. *Disability and Health Journal, 14*(2), 101008.
<https://doi.org/10.1016/j.dhjo.2020.101008>
- Sirard, J. R., & Pate, R. R. (2001). Physical Activity Assessment in Children and Adolescents. *Sports Med.*
- Soini, A., Villberg, J., Sääkslahti, A., Gubbels, J., Mehtälä, A., Kettunen, T., & Poskiparta, M. (2014). Directly Observed Physical Activity among 3-Year-Olds in Finnish Childcare. *International Journal of Early Childhood, 46*(2), 253–269.
<https://doi.org/10.1007/s13158-014-0111-z>
- Trost, S. G. (2007). State of the Art Reviews: Measurement of Physical Activity in Children and Adolescents. *American Journal of Lifestyle Medicine, 1*(4), 299–314.
<https://doi.org/10.1177/1559827607301686>
- Van Cauwenberghe, E., Gubbels, J., De Bourdeaudhuij, I., & Cardon, G. (2011). Feasibility and validity of accelerometer measurements to assess physical activity in toddlers. *International Journal of Behavioral Nutrition and Physical Activity, 8*(1), 67.
<https://doi.org/10.1186/1479-5868-8-67>

CHAPTER 5: ACCURACY OF HIP AND WRIST COLLECTED ACCELEROMETER DATA COMPARED TO DIRECT OBSERVATION IN TODDLERS

Abstract

Inconsistencies in toddler physical activity (PA) assessment exists due to the differences in methodological decisions made between studies. It is currently unknown how individual classification methods developed to assess toddler behavior compare. Performing simultaneous cross-validation of all existing methods will indicate which should be considered for use in future studies. **Purpose:** To cross-validate various cut-points for assessing sedentary time (ST) and PA in toddlers using hip- and wrist-placed accelerometers. **Methods:** Thirty-one toddlers (13 girls) wore two accelerometers, one on the non-dominant wrist and one on the right hip, during four, 30-minute free-play sessions. Twenty-five cut-points were applied, for both the vertical axis and vector magnitude data, using an epoch length of 5 - 60-secs. Sessions were video recorded and coded using the rules outlined in the Observational System for Recording Physical Activity in Children - Toddler Version in 30-s intervals. Mean absolute difference (MAD), Pearson's r correlation coefficient, and equivalence testing were calculated for all classification methods, for each activity intensity. **Results:** MAD (%) values ranged from 23.2 – 42.4% for SB, 26.6 – 51.8% for sedentary + light, 10.6 – 56.9% for light PA, and 7.4 – 32.4% for moderate-to-vigorous PA (MVPA). Percentage of time classified as sedentary, light, or MVPA varied greatly between cut-points. Accelerometer placement or data axes used did not influence accuracy. No set of cut-points, applied to any intensity, were determined to be equivalent to direct observation. **Conclusion:** Lack of comparability of cut-points with direct observation may be due to the wide range of epochs that differed from the direct observation interval, and the way SB was classified according to the direct observation protocol. Future

research should assess more advanced methods of measurement, such as machine learning, to determine if more accurate methodologies could be used to assess PA in toddlers.

Introduction

Despite well-known evidence of the benefits associated with physical activity (PA), we continue to see a decline in PA levels during childhood (Telama et al., 2005). Evidence suggests that low levels of PA are associated with an increased risk of negative health outcomes (Cliff et al., 2009; Lee et al., 2017). These outcomes include higher body mass index (BMI) and risk of obesity along with higher risk of chronic diseases such as cardiovascular and metabolic disease (Lee et al., 2017).

Early childhood (0 – 5 years old) is a critical period for child development (Timmons et al., 2012). This period is comprised of rapid physical and cognitive growth. Habits established during this time have been found to track over time, indicating that a habit developed in these years will continue through childhood and into adulthood (Telama et al., 2005). Thus, early childhood is an ideal period for promoting PA engagement and preventing long bouts of sedentary time (ST).

Several current guidelines recommend that toddlers engage in 180-minutes of PA across the day (Okely et al., 2017; Tremblay et al., 2017; Willumson & Bull, 2020). Rather than encouraging PA of different intensities, this recommendation is for 180-minutes of total PA, including light (LPA) and moderate-to vigorous PA (MVPA). These guidelines promote PA and movement rather than high intensity PA.

There have been conflicting results reported regarding toddler PA levels. Several studies report that toddlers are engaging in elevated levels of PA and that they are meeting the 180-minutes per day recommendation (Hnatiuk et al., 2012; Johansson, Hagströmer et al., 2015; Lee et al., 2017). Other studies conclude that toddler PA levels are quite low and that they do not

meet the recommendations (Herzig et al., 2017; Vanderloo & Tucker, 2015; Wijtzes et al., 2013). A recent systematic review concluded that across studies, toddlers engaged in 72.9 – 636.5 min/day of total PA, 48.5 – 582.4 min/day of LPA, 6.5 – 89.9 min/day of MVPA, and 172.7 – 545.0 min/day of ST (Bruijns et al., 2020). These inconsistencies could be due to differences in methodological decisions made between studies. Researchers suggest there should be a consensus on decisions which influence quantification of PA levels, and this applies to all age groups, including toddlers.

Accelerometers are a frequently chosen option for assessing PA levels (Troost, 2007). Accelerometers can be uniaxial or multiaxial, measuring accelerations in multiple planes. Accelerations can then be used to quantify PA levels. Several studies have calibrated cut-points for toddlers based on hip-mounted ActiGraph accelerometers (Costa et al., 2014; Kelly & Villalpando, 2016; Pulakka et al., 2013; Trost et al., 2012). These studies have resulted in a large variety of cut-points ranging from ≤ 5 counts per 5s (Costa et al., 2014) to ≤ 43 counts per 5s (Johansson et al., 2015) for sedentary time (ST) and ≥ 165 counts per 5s (Costa et al., 2014) to ≥ 290 counts per 5s (Johansson et al., 2015) for moderate-to-vigorous PA (MVPA). Two studies have calibrated cut-points based on accelerometer data from a wrist-placed or ankle-placed device (Hager et al., 2016; Johansson et al., 2015). In older children, wrist placed device wear has been found to have a higher compliance compared to hip placed (Fairclough et al., 2016; Nyberg et al., 2009). Studies reported differences in estimates of PA and ST in which toddlers were engaging depending on the cut-points that were applied (Bruijns et al., 2020).

It is unknown which existing method is most accurate or valid because it is difficult to compare the original validation studies due to different statistical analyses, sample

characteristics, and settings. Cross-validation in an independent sample provides the best indication of how methods will perform in new samples (Browne, 2000). Performing simultaneous cross-validation of all existing methods could indicate which should be considered for use in future studies. Therefore, the purpose of this study was to cross-validate numerous cut-points for assessing PA and ST in toddlers using direct observation as the criterion measure. The secondary purpose of this study was to assess differences between hip and wrist data, using both vertical axis (VA) and vector magnitude (VM) data.

Methods

Participants

Toddlers were recruited from three childcare centers in Michigan. Potential childcare centers were contacted via phone or email. Upon agreement of becoming a site, leadership at the childcare center was asked to send informational emails to parents of toddlers (12 – 36 months). Emails included information regarding the study, in addition to links to an eligibility survey. To be eligible, toddlers had to be between 12 and 36 months old and attend one of the included childcare centers. If found eligible, parents were invited to enroll their child in the study using an online consent form. Eligibility and consent were obtained using REDCap, unless not feasible, then a paper consent form was provided. This study was approved by the Michigan State Institutional Review Board.

Procedures

Height and weight were measured using a portable stadiometer (ShorrBoards®) and electric scale (Tanita®). Childcare staff informed researchers of times toddlers were scheduled to have free-play for observation. This allowed for nap times and snacks to be avoided.

Observations occurred both indoors and outdoors for 30-minutes per session on two separate days. During observation periods toddlers wore two accelerometers, one on their non-dominant wrist and one on their right hip. Dominance was reported by parents during the initial consenting process and then assessed in the field using a high-five protocol. Toddlers were asked for a high-five two separate times. If the same hand was used, that was determined to be dominant. If separate hands were used, toddlers were given a crayon and asked to draw on a piece of paper. The preferred hand was considered dominant. If a toddler did not express dominance, the device was placed on the left wrist. All observations were video recorded by GoPro Hero5 Session (GoPro, San Mateo, CA) cameras placed to record physical activity data in the classroom and on the playground. Video was collected at 1080p using a wide view at 30 frames/s. Universal time was synchronized to the camera using the website time.is.

Observational System for Recording Physical Activity in Children - Toddler

The physical activity levels outlined in the Observational System for Recording Physical Activity in Children (OSRAC) were used as the criterion measure of physical activity intensity (Brown et al., 2006). These systems have previously been used to assess PA level in toddlers (Dinkel et al., 2019; Fees et al., 2015; Van Cauwenberghe et al., 2011). The OSRAC-T is a focal child system used to code information regarding that child's physical activity level or intensity and type, and the child's physical and social environments. In the present study, physical activity level was coded using a 5-second observation window to represent each 30-second interval on a 1 to 5 scale where 1 = stationary/motionless, 2 = stationary with movement of limbs or trunk, 3 = slow/easy movement, 4 = moderate movement, and 5 = fast movement. The highest PA level exhibited during each 5-second interval was coded for that observation. If a 1

or 2 was coded for the interval, this was categorized as sedentary. If it was coded 3 it was categorized as light, and 4 and 5 were categorized as moderate-to-vigorous PA. If toddlers were not visible, physical activity level was coded as “can’t tell.” Coding was performed using Behavioral Observation Research Interactive Software (BORIS). BORIS is a free event logging software that allows for the coding of video observations. Each child was observed individually. Researchers were trained to use the OSRAC-P protocols and exhibited reliability between coders ($k=0.608$).

ActiGraph wGT3X-BT

The ActiGraph wGT3X-BT triaxial accelerometer (Pensacola, FL) was also used to assess physical activity levels. This is a small, water-resistant monitor that collects data in three axes. Accelerometers were initialized to record raw acceleration data at a sampling rate of 30 Hz using ActiLife software (version 6.13.4). Data were downloaded as ‘gt3x’ files using ActiLife.

Statistical Analysis

Descriptives of demographic information were calculated (Table 10). Raw data in ‘gt3x’ file format were imported to RStudio (Vienna, Austria; version 1.3.1056) and activity counts were generated for both devices using ActiGraph’s algorithm using the *agcounts* package (version 0.6.6). Applied cut-points, identified using a previously conducted systematic review, included both those calibrated for toddlers and those calibrated for preschoolers that have been previously applied to toddlers. Cut-points identified included Butte (2014), Costa (2014), Dobell (2019), Evenson (2008), Johansson (2015), Kelly (2016), Oftedal (2014), Pulakka (2013), Pate (2006), Reilly (2003), Sirard (2005), Trost (2012), and Van Cauwenberghe (2011). Rather than scaling previously developed cut-points to match the 30-second epoch in which DO data

were collected, cut-points were applied at the epoch used in the original validation studies. Percentage of time spent in each intensity was calculated and compared to the criterion of direct observation.

The percentage of intervals classified as each intensity level mentioned above was compared between the criterion (direct observation) and the aforementioned individual classification approaches using mean absolute difference, Pearson's r correlation coefficient, equivalence testing, and Bland-Altman plots generated using the 'blandr' package (version 0.5.1). When no difference between outcomes is expected, equivalence testing is more appropriate than standard hypothesis testing (e.g., paired t-tests). The two one-sided tests of equivalence were conducted using the 'TOSTER' package (version 0.4.0). If the 90% confidence interval around the mean difference did not overlap or exceed the equivalence bounds, the methods were considered equivalent ($p < 0.05$). Equivalence bounds were set as 10% of the mean MVPA according to the criterion (O'Brien, 2021). Normality was verified for all variables using histograms. All analyses were conducted in RStudio (version 2023.06.1).

Results

Observation data were obtained for 31 toddlers (41.9% girls) who were on average 25.5 \pm 6.0 months old (Table 10).

Table 10. Participant Demographics

Gender Female, n (%)	13 (41.9)
Age, mean months (SD)	25.5 (6.0)
Length, mean cm (SD)	86.4 (5.8)
Weight, mean kg (SD)	12.9 (1.5)
Race/Ethnicity, n (%)	

Table 10. (cont'd)

American Indian or Alaska Native	1 (3.2)
Asian	2 (6.5)
Black or African American	3 (9.7)
Hispanic or Latino	1 (3.2)
Native Hawaiian or Pacific Islander	1 (3.2)
White/Caucasian	19 (61.3)
Other	1 (3.2)
Prefer not to Answer	3 (9.7)

In total, 7,757 30-second observation intervals were included in the analysis. Table 11 provides an overview of the cut-points for vertical axis and vector magnitude data of wrist- and hip-worn accelerometers that were applied in the current study. Only two wrist-based cut-points were applied, while the remaining methods were hip-based.

Table 11. Individual classification methods

	Epoch (s)	ST	MVPA
Hip-based, vertical axis			
Butte et al., 2014	60	≤ 240	≥ 4450
Costa et al., 2014	5	≤ 5	≥ 165
Dobell et al., 2019	5	≤ 11	≥ 83
Dobell et al., 2019	15	≤ 32	≥ 249
Dobell et al., 2019	30	≤ 65	≥ 498
Evenson et al., 2008	15	≤ 25	≥ 1003
Johannson et al., 2015	5	≤ 43	≥ 290
Kelly et al., 2016	15	≤ 181	≥ 435
Oftedal et al., 2014*	5	< 2	≥ 2
Pulakka et al., 2013*	15	< 35	≥ 35
Pate et al., 2006	15	≤ 37	≥ 420
Reilly et al., 2003*	60	< 1100	≥ 1100
Sirard et al., 2005	15	≤ 301	≥ 615
Trost et al., 2012	15	≤ 48	> 418
Van Cauwenberge et al., 2011	15	≤ 372	≥ 585
Hip-based, vector magnitude			
Butte et al., 2014	60	≤ 820	≥ 6112

Table 11. (cont'd)

Costa et al., 2014	5	≤ 96.12	≥ 361.94
Dobell et al., 2019	5	≤ 78	≥ 183
Dobell et al., 2019	15	≤ 221	≥ 517
Dobell et al., 2019	30	≤ 443	≥ 1029
Johannson et al., 2015	5	≤ 105	≥ 512
Oftedal et al., 2014*	5	≤ 40	> 40
Pulakka et al., 2013*	15	< 208	≥ 208
Wrist-based, vertical axis			
Johannson et al., 2015	5	≤ 89	≥ 440
Wrist-based, vector magnitude			
Johannson et al., 2015	5	≤ 221	≥ 729

SB, sedentary time; MVPA, moderate-to-vigorous physical activity; *cut-point developed for ST+Light PA and MVPA

Tables 12 – 15 present the equivalence of assessing the various levels of PA using the different hip- and wrist- based cut-points. Three studies only included cut-points to distinguish between MVPA and non-MVPA (Oftedal et al., 2014; Pulakka et al., 2013; Reilly, 2008). The equivalence of non-MVPA is presented in Table 13. No methods were determined to be equivalent to direct observation regardless of wear location (hip v. wrist) or data axes (1 axis v. vector magnitude).

Sedentary Time

Only two estimates of sedentary time were determined to be equivalent (10% of mean: ± 6.7%) including Dobell et al., 2019 (15-s hip-based VM) and Johannson et al., 2015 (5-s hip-based VA). Overall, most cut-points underestimated time spent sedentary (Figure 2a.) The cut-points that exhibited the lowest bias were Dobell et al., 2019 (15-s hip-based VM; 2.9%; lower limit: $p < 0.001$, upper limit: $p = 0.02$; Table 3). MAD values ranged from 15.4% (Dobell et al., 2019; 15-s hip-based VM data) to 42.4% (Butte et al., 2014, 60-s hip-based VM data). All but four sets of cut-points were moderately to highly correlated with direct observation (Johannson

et al. 2015, 5-s wrist-based; Sirard et al., 2005; 15-s hip-based; Van Cauwenberge et al., 2011; 15-s hip-based).

Table 12. Comparison of sedentary time percentage according to criterion versus each individual classification methods

Method	Loc	Epoch (s)	SED (%)				Absolute Difference			Equivalence Test		
			Mean	SD	Min	Max	Mean	SD	<i>r</i>	Bias	90% CI	Equivalent
Criterion	-	-	66.7	20.0	3.3	100.0	-	-	-	-	-	-
Butte et al. (2014)	Hip	60	35.8	25.5	0.0	100.0	33.3	17.9	0.56	-30.9	-34.2, -27.6	N
Butte et al. (2014)*	Hip	60	25.8	25.7	0.0	100.0	41.9	18.9	0.60	-40.8	-44.1, -37.6	N
Costa et al. (2014)	Hip	5	45.7	21.2	3.1	100.0	23.2	13.6	0.67	-21.0	-23.6, -18.5	N
Costa et al. (2014)*	Hip	5	45.6	21.0	3.3	100.0	22.8	13.2	0.70	-21.1	-23.5, -18.6	N
Dobbell et al. (2019)	Hip	5	51.5	20.5	0.0	100.0	19.1	12.1	0.65	-15.1	-17.7, -12.5	N
Dobbell et al. (2019)*	Hip	5	41.5	21.5	3.1	100.0	26.6	14.0	0.69	-25.2	-27.7, -22.7	N
Dobbell et al. (2019)	Hip	15	40.0	23.0	0.0	100.0	24.8	14.7	0.62	-21.7	-24.6, -18.8	N
Dobbell et al. (2019)*	Hip	15	34.6	23.4	1.6	100.0	15.4	13.9	0.66	2.9	-0.2, 6.1	Y
Dobbell et al. (2019)	Hip	30	33.2	24.1	0.0	100.0	30.7	16.3	0.58	-29.2	-32.1, -26.3	N
Dobbell et al. (2019)*	Hip	30	31.0	24.7	0.0	100.0	33.2	16.0	0.64	-32.1	-34.8, -29.3	N
Evenson et al. (2008)	Hip	15	37.5	23.0	0.0	100.0	17.7	17.8	0.62	14.2	11.0, 17.3	N
Johansson et al. (2015)	Wrist	5	37.5	23.1	6.3	100.0	34.4	17.6	0.32	-29.4	-33.2, -25.6	N
Johansson et al. (2015)*	Wrist	5	39.3	23.1	7.7	100.0	33.1	17.7	0.31	-27.5	-31.3, -23.6	N
Johansson et al. (2015)	Hip	5	78.2	17.2	15.3	100.0	23.2	25.8	0.53	-5.5	-10.4, 0.2	Y
Johansson et al. (2015)*	Hip	5	47.7	20.8	3.6	100.0	24.8	16.5	0.70	16.8	13.1, 20.4	N
Kelly et al. (2016)	Hip	15	69.6	18.8	13.9	100.0	27.1	15.4	0.43	-24.9	-27.8, -22.1	N
Pate et al. (2006)	Hip	15	41.8	22.8	0.8	100.0	19.7	19.0	0.62	18.6	15.5, 21.7	N
Sirard et al. (2005)	Hip	15	80.8	15.7	23.8	100.0	35.0	17.9	0.34	-33.5	-36.6, -30.4	N
Trost et al. (2012)	Hip	15	45.0	22.5	0.8	100.0	28.6	15.8	0.61	-26.7	-29.6, -23.8	N
Van Cauwenberghe et al. (2011)	Hip	15	85.3	13.3	31.2	100.0	36.9	17.1	0.32	-35.7	-38.7, -32.8	N

*Vector Magnitude; Loc: location; SED: sedentary; SD: standard deviation; N: not equivalent

Non-MVPA

When assessing sedentary + light behavior (non-MVPA), estimates were found not to be equivalent to DO (10% of mean: $\pm 8.5\%$; Table 13). Time spent in non-MVPA was underestimated (Figure 2b.). The cut-points that exhibited the lowest bias were Reilly et al., 2003 (60-s hip-based VA; -0.9% ; lower limit: $p = 0.006$, upper limit: $p = < 0.001$). MAD values were higher than ST ranging from 15.2% (Reilly et al., 2003, 60-s hip-based VA data) to 52.6% (Ofstedal et al., 2014, 5-s hip-based VM data) with low to moderate correlations ranging from 0.38 (Reilly et al., 2003, 60-s hip-based VA data) to 0.50 (Pulakka et al., 2019, 15-s hip-based VM data).

Table 13. Comparison of non-MVPA time percentage according to criterion versus each individual classification method

Method	Loc	Epoch (s)	SED+LPA (%)				Absolute Difference			Equivalence Test		
			Mean	SD	Min	Max	Mean	SD	<i>r</i>	Bias	90% CI	Equivalent
Criterion	-	-	84.2	14.1	31.2	100.0	-	-	-	-	-	-
Ofstedal et al. (2014)	Hip	5	42.4	21.5	2.5	100.0	42.5	18.0	0.51	-42.2	-45.1, -39.3	N
Ofstedal et al. (2014)*	Hip	5	32.3	22.0	1.9	100.0	52.6	18.8	0.48	-52.3	-55.3, 49.3	N
Pulakka et al. (2013)	Hip	15	41.1	23.0	0.0	100.0	43.9	19.2	0.50	-43.5	-46.6, -40.5	N
Pulakka et al. (2013)*	Hip	15	33.4	23.3	1.6	100.0	51.4	19.6	0.50	-51.1	-54.3, -48.0	N
Reilly et al. (2003)	Hip	60	78.9	20.5	6.7	100.0	15.2	13.9	0.38	-5.7	-8.7, -2.6	N

*Vector Magnitude; Loc: location; SED+LPA: sedentary + light physical activity; SD: standard deviation; N: not equivalent

Light PA

When assessing light physical activity behavior, estimates were also not equivalent to DO (10% of mean: $\pm 1.8\%$). Overall, cut-points overestimated time spent in light PA (Figure 2c.).

The cut-points that exhibited the lowest bias were Dobell et al., 2019 (15-s hip-based VM; 0.6%; lower limit: $p = 0.03$, upper limit: $p = 0.17$). MAD values had a larger range than SB ranging from 9.0% (Dobell et al., 2005, 15-s hip-based VM data) to 56.9% (Butte et al., 2014, 60-s hip-based VM data) and very low to moderate correlations ranging from 0.11 (Van Cauwenberghe et al., 2011, 15-s hip-based VA data) to 0.67 (Dobell et al., 2019, 5-s hip-based VA data).

Table 14. Comparison of light physical activity percentage according to criterion versus each individual classification method

Method	Loc	Epoch (s)	LPA (%)				Absolute Difference			Equivalence Test		
			Mean	SD	Min	Max	Mean	SD	<i>r</i>	Bias	90% CI	Equivalent
Criterion	-	-	17.9	11.4	0.0	95.8	-	-	-	-	-	-
Butte et al. (2014)	Hip	60	59.4	23.0	0.0	100.0	42.5	18.7	0.43	41.5	38.4, 44.7	N
Butte et al. (2014)*	Hip	60	68.0	24.0	0.0	100.0	50.9	18.8	0.50	50.1	47.0, 53.3	N
Costa et al. (2014)	Hip	5	43.8	16.6	0.0	96.7	26.7	11.1	0.63	25.9	23.9, 27.9	N
Costa et al. (2014)*	Hip	5	43.7	16.9	0.0	96.1	26.6	11.0	0.66	25.8	23.9, 27.8	N
Dobbell et al. (2019)	Hip	5	27.2	12.3	0.0	85.8	11.1	7.5	0.67	9.3	7.9, 10.8	N
Dobbell et al. (2019)*	Hip	5	23.2	9.5	0.0	66.5	9.2	6.0	0.59	5.3	3.8, 6.7	N
Dobbell et al. (2019)	Hip	15	36.9	15.5	0.0	95.8	25.7	11.9	0.58	24.6	22.5, 26.7	N
Dobbell et al. (2019)*	Hip	15	27.3	11.6	0.0	66.9	9.0	10.1	0.47	0.6	-1.4, 2.7	N
Dobbell et al. (2019)	Hip	30	43.0	17.5	0.0	66.9	38.7	14.7	0.51	37.9	35.4, 40.4	N
Dobbell et al. (2019)*	Hip	30	31.5	14.5	0.0	75.4	13.0	8.4	0.42	9.4	7.6, 11.2	N
Evenson et al. (2008)	Hip	15	55.8	19.7	0.0	98.3	9.9	10.9	0.54	-4.4	-6.6, -2.3	N
Johansson et al. (2015)	Wrist	5	44.1	15.2	0.0	62.3	32.2	18.0	0.23	-29.1	-32.5, -25.7	N
Johansson et al. (2015)*	Wrist	5	43.5	15.2	0.0	66.5	31.4	18.5	0.25	-28.4	-31.8, -25.0	N
Johansson et al. (2015)	Hip	5	28.0	14.4	0.0	69.4	16.9	14.4	0.28	-11.4	-14.2, -8.5	N
Johansson et al. (2015)*	Hip	5	48.6	19.1	0.0	96.1	36.4	23.8	0.60	-33.6	-37.7, -29.5	N

Table 14. (cont'd)

Kelly et al. (2016)	Hip	15	18.5	9.8	0.0	42.9	28.8	12.6	0.19	27.9	25.7, 30.1	N
Pate et al. (2006)	Hip	15	45.8	17.2	0.0	95.8	10.7	11.6	0.56	-9.7	-11.6, -7.8	N
Sirard et al. (2005)	Hip	15	13.5	9.8	0.0	37.7	26.0	13.6	0.13	25.1	22.7, 27.4	N
Trost et al. (2012)	Hip	15	42.5	16.4	0.0	95.0	19.9	11.3	0.54	19.0	17.0, 21.0	N
Van Cauwenberghe et al. (2011)	Hip	15	8.2	6.5	0.0	26.23	16.5	10.5	0.11	13.6	11.4, 15.8	N

*Vector Magnitude; Loc: location; LPA: light physical activity; SD: standard deviation; N: not equivalent

MVPA

Lastly, when assessing MVPA, no estimates were deemed equivalent (10% of mean: \pm 0.96%; Table 6.). Over 2/3 of cut-points overestimated time spent in MVPA (Figure 2d.). Both sets of Costa (2014) cut-points exhibited the lowest bias (5-s hip-based VA; 1.0%; lower limit: $p = 0.04$, upper limit: $p = 0.51$ and 5-s hip-based VM; 1.1%; lower limit: $p = 0.03$, upper limit: $p = 0.54$). Overall, MAD values were lower than other intensities, but ranged greatly from 7.0% (Trost et al., 2012, 15-s hip-based VA data) to 58.4% (Van Cauwenberghe et al., 2011, 15-s hip-based VA data) and low to moderate correlations ranging from 0.24 (Johansson et al., 2015, 5-s, hip-based VM data) to 0.61 (Dobell et al., 2019, 5-s hip-based VM data).

Table 15. Comparison of moderate-to-vigorous physical activity percentage according to criterion versus each individual classification method

Method	Loc	Epoch (s)	MVPA (%)				Absolute Difference			Equivalence Test		
			Mean	SD	Min	Max	Mean	SD	r	Bias	90% CI	Equivalent
Criterion	-	-	9.8	12.1	0.0	68.9	-	-	-	-	-	-
Butte et al. (2014)	Hip	60	4.8	9.7	0.0	60.0	49.7	19.3	0.37	49.4	46.3, 52.4	N
Butte et al. (2014)*	Hip	60	6.1	10.6	0.0	61.3	57.3	19.7	0.44	57.0	53.9, 60.2	N
Costa et al. (2014)	Hip	5	10.6	9.3	0.0	51.9	7.8	8.8	0.41	1.0	-0.8, 2.8	N

Table 15. (cont'd)

Costa et al. (2014)*	Hip	5	10.6	9.1	0.0	45.1	7.7	8.6	0.42	1.1	0.7, 2.8	N
Dobbell et al. (2019)	Hip	5	21.2	14.2	0.0	73.0	9.6	9.6	0.46	3.0	1.0, 5.0	N
Dobbell et al. (2019)*	Hip	5	35.4	17.7	0.0	80.9	9.1	9.6	0.61	2.3	0.3, 4.3	N
Dobbell et al. (2019)	Hip	15	23.1	17.3	0.0	81.2	7.2	9.9	0.42	-2.8	-4.6, -1.0	N
Dobbell et al. (2019)*	Hip	15	38.1	20.5	0.0	88.5	28.9	15.7	0.60	28.5	26.0, 31.0	N
Dobbell et al. (2019)	Hip	30	23.9	19.1	0.0	86.9	16.9	12.9	0.39	13.6	11.1, 16.1	N
Dobbell et al. (2019)*	Hip	30	37.6	22.2	0.0	95.1	9.5	9.6	0.60	2.9	0.9, 4.9	N
Evenson et al. (2008)	Hip	15	6.8	7.9	0.0	44.3	8.1	10.4	0.34	-4.7	-6.6, -2.9	N
Johansson et al. (2015)	Wrist	5	18.4	12.4	0.0	53.3	12.9	9.8	0.41	8.8	6.7, 10.8	N
Johansson et al. (2015)*	Wrist	5	17.2	12.4	0.0	55.3	12.1	10.2	0.40	7.6	5.5, 9.7	N
Johansson et al. (2015)	Hip	5	3.9	4.0	0.0	19.1	7.3	10.7	0.24	-5.7	-7.5, -3.9	N
Johansson et al. (2015)*	Hip	5	3.7	3.9	0.0	20.5	7.2	10.6	0.29	-5.8	-7.6, -4.1	N
Kelly et al. (2016)	Hip	15	11.9	11.7	0.0	62.3	7.2	9.8	0.39	-3.0	-4.8, -1.2	N
Oftedal et al. (2014)	Hip	5	57.6	21.5	0.0	97.5	14.7	10.4	0.50	11.7	9.6, 13.8	N
Oftedal et al. (2014)*	Hip	5	67.7	22.0	0.0	98.1	26.2	13.2	0.45	25.8	23.7, 27.9	N
Pate et al. (2006)	Hip	15	12.4	11.9	0.0	63.1	7.5	9.9	0.38	-3.4	-5.2, -1.6	N
Pulakka et al. (2013)	Hip	15	58.9	23.0	0.0	100.0	18.1	14.4	0.49	14.3	11.5, 17.1	N
Pulakka et al. (2013)*	Hip	15	66.6	23.3	0.0	98.4	28.6	16.9	0.47	28.0	25.3, 30.7	N
Reilly et al. (2003)	Hip	60	21.1	20.5	0.0	93.3	16.2	15.5	0.40	11.5	8.6, 14.5	N
Sirard et al. (2005)	Hip	15	5.7	7.1	0.0	38.5	48.4	17.9	0.34	48.1	45.2, 50.9	N
Trost et al. (2012)	Hip	15	12.5	12.0	0.0	63.9	7.0	10.1	0.38	-3.9	-5.6, -2.1	N
Van Cauwenberghe et al. (2011)	Hip	15	6.5	7.7	0.0	42.6	58.4	18.8	0.34	58.1	55.1, 61.1	N

*Vector Magnitude; Loc: location MVPA: moderate-to-vigorous physical activity; SD: standard deviation; N: not equivalent

The 95% Limits of agreement and respective Bland Altman plots for cut-points across intensities with the lowest and highest bias can be seen in Figure 3a. and b. respectively.

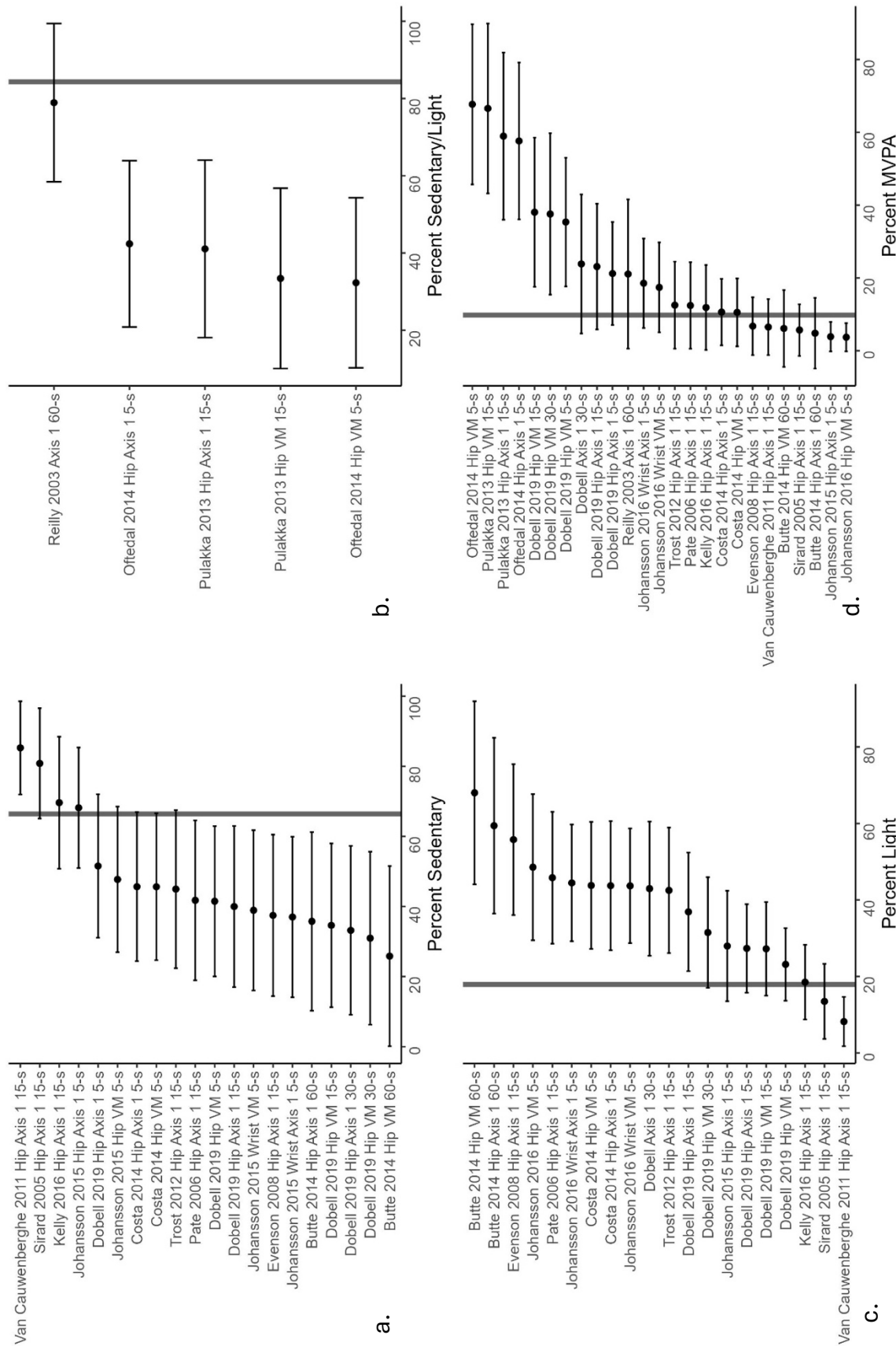
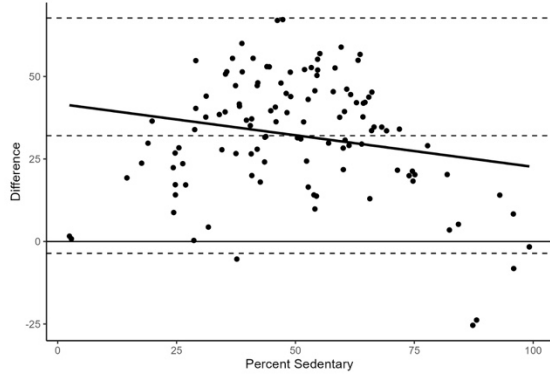
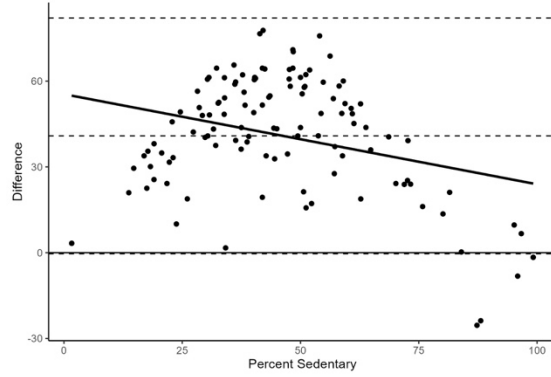


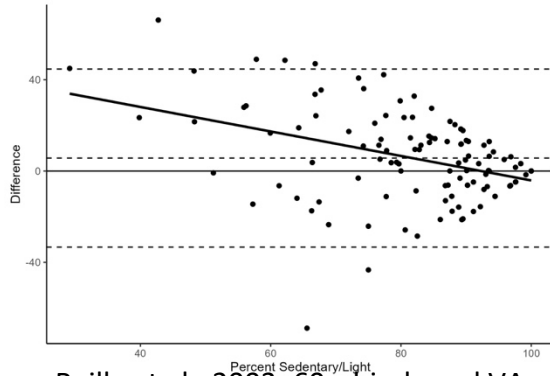
Figure 2. Individual classification approach equivalence to criterion measure across activity intensities. a. Percent Sedentary; b. Percent Sedentary/Light; c. Percent Light; d. Percent MVPA



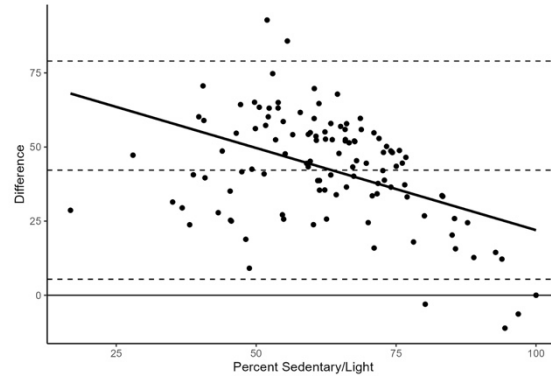
Dobell et al., 2019; 15s, hip-based VM
(bias = 2.9)



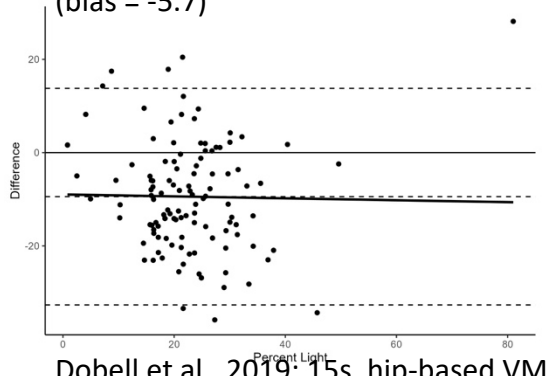
Butte et al., 2014; 60s, hip-based VM
(bias = -40.8)



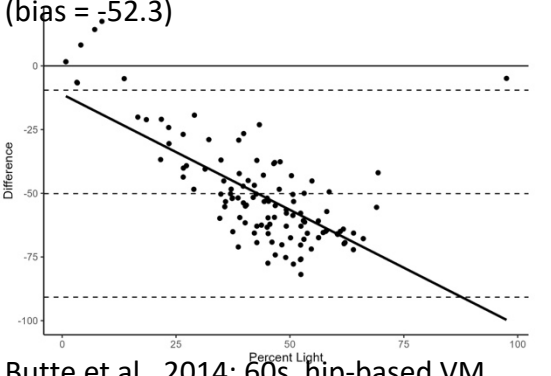
Reilly et al., 2003; 60s, hip-based VA
(bias = -5.7)



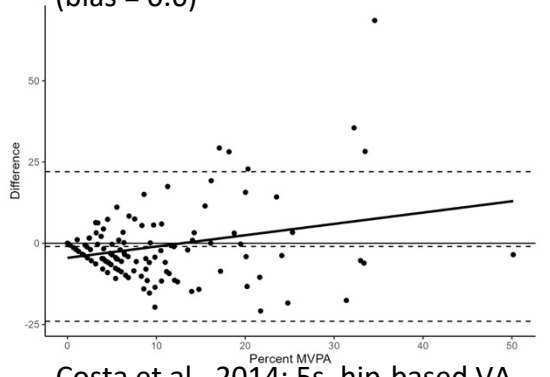
Oftedal et al., 2014; 5s, hip-based VM
(bias = -52.3)



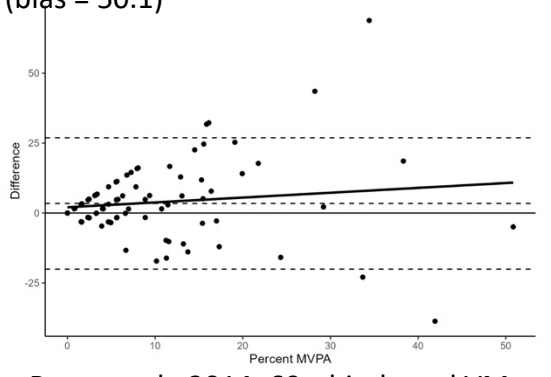
Dobell et al., 2019; 15s, hip-based VM
(bias = 0.6)



Butte et al., 2014; 60s, hip-based VM
(bias = 50.1)



a. Costa et al., 2014; 5s, hip-based VA
(bias = 1.0)



b. Butte et al., 2014; 60s, hip-based VM
(bias = 57.0)

Figure 3. Cut-points with the lowest (a.) and highest (b.) bias across all intensities

Discussion

The present study is the first to cross-validate various cut-points developed for toddlers or previously applied to toddlers using both hip- and wrist- worn accelerometers and vertical axis and vector magnitude data. Specifically, the sample in this study included only those in the toddler age range, 12 – 36 months. Previous cross-validation studies included children outside of the range (Altenburg et al., 2022) or did not include the entire age range (Pereira et al., 2020; Van Cauwenberghe et al., 2011). The current study provides a comprehensive overview of all cut-points identified in a previous systematic review to assess physical activity and sedentary time in toddlers. Across intensities, both placements, and data axis types, only two methods were found to be within $\pm 10\%$ equivalence of DO. These methods both were used to classify sedentary time (Dobell et al., 2019; Johansson et al., 2015). Both cut-points were calibrated at the hip but had no other similarities. One calibrated cut-points in preschoolers using 15-s VM data, while the others were toddler calibrated using 5-s VA data.

Using direct observation, this study reported 66.7%, 17.9%, and 9.8% of toddler free-play was spent engaging in sedentary time, light physical activity, and MVPA respectively. This is similar to other studies assessing PA in toddlers using direct observation (Altenburg et al., 2022; Fees et al., 2015; Gubbels et al., 2011; Van Cauwenberghe et al., 2011). While not the purpose of the present study, children were observed during indoor and outdoor free-play at child care. Thus, these values reflect the relatively high proportion of time children spend being sedentary, and the need for activity-promoting interventions in this setting and population.

Overall, when assessing sedentary time, most cut-points resulted in underestimations. Percent time spent sedentary varied greatly across cut-points and at most underestimated

sedentary time by 40% (Butte et al., 2014; 60-s hip-based VM). This underestimation is similar to values reported in other cross-validation studies (Altenburg et al., 2022) using direct observation, but different to those using another accelerometer brand or model as a comparison which reported overestimations of sedentary time (Pereira et al., 2020). Only the two sets of cut-points found to be equivalent had the smallest mean bias (3-5%; Dobell et al., 2019; Johansson et al., 2015). Sedentary time may be underestimated due to unique behavior seen with toddlers such as being carried by an adult which would be captured as sedentary by direct observation, but result in accelerometer-recorded count levels similar to walking. Specifically, when comparing hip-worn accelerometer counts during toddler walking v. being carried by an adult, counts were 49/5s and 144/5s respectively (data not shown). When assessing wrist-worn data, counts were not higher than walking, but were still higher than sedentary counts. This could result in an underestimation of sedentary time.

With respect to MVPA, bias was smaller compared to other intensities, but cut-points tended to overestimate the percentage of time spent in MVPA. Four cut-points overestimated MVPA by 52.0 – 61.7%; however, it is important to note that these four studies calibrated cut-points for non-MVPA versus MVPA (Oftedal et al., 2014; Pulakka et al., 2013). Unlike ST for which there was consistently high bias, numerous cut-points for MVPA exhibited smaller levels of bias (Costa et al., 2014; Evenson et al., 2008; Dobell et al., 2019; Johansson et al., 2015; Kelly & Villalpando, 2016; Pate et al., 2006; Trost et al., 2012). These findings are similar to other studies (Altenburg et al., 2022). These methods were all calibrated with a hip-worn device but include both VM and VA data. Four methods were calibrated in toddlers (Costa et al., 2014;

Johansson et al., 2015; Kelly & Villalpando, 2016; Trost et al., 2012), while the remaining were preschool cut-points (cite?).

Like MVPA, LPA was overestimated by all but two sets of cut-points. LPA was overestimated by as much as 54.2%. Four cut-points resulted in lower levels of bias (Dobell et al., 2019; Evenson et al., 2008). These were both hip-based methods, but again included VM and VA data. All cut-points were calibrated in preschoolers. Again, these results were like those reported by Altenburg et al., 2022, who reported poor precision across cut-points with constant overestimation. Overall, the accuracy of estimations varied greatly across cut-points for all sedentary time and physical activity cut-points.

As mentioned, this study applied all cut-points calibrated using toddlers in addition to any cut-point calibrated in preschoolers that has been included in previous cross-validation studies. Nine cut-points were toddler specific, while the remaining 16 sets were originally from a preschool population. A previous cross-validation studies reported similar poor to fair accuracy across cut-points applied to toddlers (Altenburg et al., 2022). In the present study, only two methods were equivalent to direct observation, however there were no consistent characteristics of these methods. The only similarity was that across intensities, those that had the smallest mean bias were all calibrated at the hip. Another cross-validation study concluded that the only cut-point that resulted in small bias (5%) of sedentary time were the hip-based Trost (2012) cut-points (Pereira et al., 2020). Similar to the current study, other cross-validation of toddler/preschool cut-points have concluded that future research is warranted for more accurate estimations of physical activity (Altenburg et al., 2022; Pereira et al., 2020; Van Cauwenberghe et al., 2011).

In the current study, equivalency did not vary between vector magnitude and vertical axes data. When assessing uniaxial v. triaxial monitors in preschoolers, researchers have reported that there were no clear advantages to either type of monitor (Adolph et al., 2012). Similarly, other studies in toddlers have reported similar classification accuracy between cut-points developed using both one and three axes (Johansson et al., 2015; Pulakka et al., 2013). On the contrary, Oftedal et al. (2014) reported that in toddlers, use of triaxial monitors resulted in better classification accuracy than uniaxial monitors. Since only two cut-points existed for wrist placement in toddlers, differences in equivalences by wear location cannot be discussed. However, it should be noted that wrist cut-points did display larger bias or higher deviations from the mean when compared to hip-developed methods.

These findings suggest that the Johansson (2015) 5-s hip-based VA and Dobell (2019) 15-s hip-based VM cut-points are valid for assessing sedentary time. However, no other cut-point resulted in equivalent estimations of sedentary time or physical activity in toddlers. We cannot conclude that using hip versus wrist or vector magnitude data rather than vertical axes data, or vice versa, has any advantages over the other. There are several considerations to highlight regarding this study that may explain these findings. The present study collected direct observation data using a 5-second observe, 25-second record cycle. This resulted in a 30-second epoch being used for the criterion measure. However, rather than scaling the cut-points, this study applied various cut-points at their developed epoch and compared those values to the 30-second direct observation. This is similar to Altenburg et al. (2021), but different than Van Cauwenberghe et al. (2011) who collected direct observation in 15-second intervals and applied 15-second cut-points. Calibration studies commonly use a direct observation interval that

matches their cut-point epoch (Ofstedal et al., 2014; Sirard et al., 2005; Trost et al., 2012). This could explain the lack of equivalency between direct observation and cut-point methods. However, when looking at cut-points with the lowest bias, there was no pattern in relation to the epoch in which the cut-points were calibrated. This is similar to Altenburg et al. (2021) who reported no differences in accuracy across cut-points scaled to 5, 15, or 30-seconds.

There are also differences in how sedentary time is classified using the OSRAC tools. To be comparable to other studies using direct observation, and to be in sync with those who developed the OSRAC systems, the present study classified sedentary as intervals coded as 1 (stationary) or 2 (limb or trunk movement) (Brown et al., 2006; Dinkel et al., 2019; Fees et al., 2015; McIver et al., 2009). Previous studies have classified sedentary time as only those intervals coded as 1 (Trost et al., 2012), while others have classified sedentary time as those intervals coded as 1 and 2 (Van Cauwenberghe et al., 2011). However, as an exploratory action, we classified sedentary time as intervals coded as only 1 and then re-ran the analyses (data not shown). This resulted in sedentary time being overestimated and light physical activity being underestimated, which is the opposite to what was concluded in this study. This suggests that in validation work, researchers should not only consider the methods applied in accelerometry, but the criteria used if direct observation is being utilized.

Strengths of this study include that, to our knowledge, this is the first study to cross-validate all toddler developed cut-points in a toddler population. Additionally, this study applied all preschool developed cut-points that have been previously cross-validated in toddlers. This makes our study a comprehensive assessment of methods used to measure physical activity in toddlers.

The present study is not without limitations. As mentioned above, the criterion measure of activity in this study was collected using a 5-second observation to represent a 30-second interval, in line with the established OSRAC-T protocol. It has been recognized that young children participate in short bursts of intense physical activity (Bailey et al., 1995). It is unknown if a 5-second window of observation can be applied to a 30-second interval in toddlers. Potentially, this observation window is not representative of the 30-seconds. A second limitation is the short duration of activity that was collected for each participant. Due to the intensive nature of using direct observation, only four 30-minute observation sessions were captured for each child. We cannot generalize that these same results would be concluded during 24-hour physical activity monitoring.

We conclude that across accelerometer placements and data axes, there is no set of cut-points that can accurately classify sedentary time, light, and moderate-to-vigorous physical activity in toddlers. Future research should aim to assess if there are more advanced methods of assessing physical activity, such as machine learning approaches, that may result in higher levels of accuracy. However, it should be recognized that machine learning may not be a better option, we may just need better guides on developing cut-points.

REFERENCES

- Adolph, A. L., Puyau, M. R., Vohra, F. A., Nicklas, T. A., Zakeri, I. F., & Butte, N. F. (2012). Validation of Uniaxial and Triaxial Accelerometers for the Assessment of Physical Activity in Preschool Children. *Journal of Physical Activity and Health, 9*(7), 944–953. <https://doi.org/10.1123/jpah.9.7.944>
- Altenburg, T. M., de Vries, L., op den Buijsch, R., Eyre, E., Dobell, A., Duncan, M., & Chinapaw, M. J. M. (2022). Cross-validation of cut-points in preschool children using different accelerometer placements and data axes. *Journal of Sports Sciences, 40*(4), 379–385. <https://doi.org/10.1080/02640414.2021.1994726>
- Bailey, R. C., Olson, J., Pepper, S. L., Porszasz, J., Barstow, T. J., & Cooper, D. M. (1995). *The level and tempo of children's physical activities: An observational study.* 1033–1041.
- Brown, W. H., Pfeiffer, K. A., McIver, K. L., Dowda, M., Almeida, J. M. C. A., & Pate, R. R. (2006). Assessing Preschool Children's Physical Activity: The Observational System for Recording Physical Activity in Children-Preschool Version. *Research Quarterly for Exercise and Sport, 77*(2), 167–176. <https://doi.org/10.1080/02701367.2006.10599351>
- Browne, M. W. (2000). Cross-Validation Methods. *Journal of Mathematical Psychology, 44*(1), 108–132. <https://doi.org/10.1006/jmps.1999.1279>
- Bruijns, B. A., Truelove, S., Johnson, A. M., Gilliland, J., & Tucker, P. (2020). 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, 17*(1), 14. <https://doi.org/10.1186/s12966-020-0912-4>
- Butte, N. F., Wong, W. W., Lee, J. S., Adolph, A. L., Puyau, M. R., & Zakeri, I. F. (2014). Prediction of Energy Expenditure and Physical Activity in Preschoolers. *Medicine & Science in Sports & Exercise, 46*(6), 1216–1226. <https://doi.org/10.1249/MSS.0000000000000209>
- Costa, S., Barber, S. E., Cameron, N., & Clemes, S. A. (2014). Calibration and validation of the ActiGraph GT3X+ in 2–3 year olds. *Journal of Science and Medicine in Sport, 17*(6), 617–622. <https://doi.org/10.1016/j.jsams.2013.11.005>
- Dinkel, D., Snyder, K., Patterson, T., Warehime, S., Kuhn, M., & Wisneski, D. (2019). An exploration of infant and toddler unstructured outdoor play. *European Early Childhood Education Research Journal, 27*(2), 257–271. <https://doi.org/10.1080/1350293X.2019.1579550>
- Dobell, A. P., Eyre, E. L. J., Tallis, J., Chinapaw, M. J. M., Altenburg, T. M., & Duncan, M. J. (2019). Examining accelerometer validity for estimating physical activity in pre-schoolers during

- free-living activity. *Scandinavian Journal of Medicine & Science in Sports*, 29(10), 1618–1628. rzh. <https://doi.org/10.1111/sms.13496>
- Fairclough, S. J., Noonan, R., Rowlands, A. V., Van Hees, V., Knowles, Z., & Boddy, L. M. (2016). Wear Compliance and Activity in Children Wearing Wrist- and Hip-Mounted Accelerometers. *Medicine & Science in Sports & Exercise*, 48(2), 245–253. <https://doi.org/10.1249/MSS.0000000000000771>
- Fees, B. S., Fischer, E., Haar, S., & Crowe, L. K. (2015). Toddler Activity Intensity During Indoor Free-Play: Stand and Watch. *Journal of Nutrition Education and Behavior*, 47(2), 170–175. <https://doi.org/10.1016/j.jneb.2014.08.015>
- Gubbels, J. S., Kremers, S. P. J., van Kann, D. H. H., Stafleu, A., Candel, M. J. J. M., Dagnelie, P. C., Thijs, C., & de Vries, N. K. (2011). Interaction between physical environment, social environment, and child characteristics in determining physical activity at child care. *Health Psychology*, 30(1), 84–90. <https://doi.org/10.1037/a0021586>
- Hager, E. R., Gormley, C. E., Latta, L. W., Treuth, M. S., Caulfield, L. E., & Black, M. M. (2016). Toddler physical activity study: Laboratory and community studies to evaluate accelerometer validity and correlates. *BMC Public Health*, 16(1), 936. <https://doi.org/10.1186/s12889-016-3569-9>
- Johansson, E., Ekelund, U., Nero, H., Marcus, C., & Hagströmer, M. (2015). Calibration and cross-validation of a wrist-worn ActiGraph in young preschoolers: Calibration of ActiGraph in toddlers. *Pediatric Obesity*, 10(1), 1–6. <https://doi.org/10.1111/j.2047-6310.2013.00213.x>
- Kelly, L. A., & Villalpando, J. (2016). Development of ActiGraph GT1M Accelerometer Cut-Points for Young Children Aged 12-36 Months. *Journal of Athletic Enhancement*, 5(4). <https://doi.org/10.4172/2324-9080.1000233>
- Mclver, K. L., Brown, W. H., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2009). Assessing Children's Physical Activity in their Homes: The Observational System for Recording Physical Activity in Children-Home. *Journal of Applied Behavior Analysis*, 42(1), 1–16. eax.
- Nyberg, G. A., Nordenfelt, A. M., Ekelund, U., & Marcus, C. (2009). Physical Activity Patterns Measured by Accelerometry in 6- to 10-yr-Old Children. *Medicine & Science in Sports & Exercise*, 41(10), 1842–1848. <https://doi.org/10.1249/MSS.0b013e3181a48ee6>
- Oftedal, S., Bell, K. L., Davies, P. S. W., Ware, R. S., & Boyd, R. N. (2014). Validation of Accelerometer Cut Points in Toddlers with and without Cerebral Palsy. *Medicine & Science in Sports & Exercise*, 46(9), 1808–1815. <https://doi.org/10.1249/MSS.0000000000000299>

- Pate, R. R., Almeida, M. J., McIver, K. L., Pfeiffer, K. A., & Dowda, M. (2006). Validation and Calibration of an Accelerometer in Preschool Children*. *Obesity, 14*(11), 2000–2006. <https://doi.org/10.1038/oby.2006.234>
- Pereira, J. R., Sousa-Sá, E., Zhang, Z., Cliff, D. P., & Santos, R. (2020). Concurrent validity of the ActiGraph GT3X+ and activPAL for assessing sedentary behaviour in 2–3-year-old children under free-living conditions. *Journal of Science and Medicine in Sport, 23*(2), 151–156. <https://doi.org/10.1016/j.jsams.2019.08.009>
- Pulakka, A., Cheung, Y., Ashorn, U., Penpraze, V., Maleta, K., Phuka, J., & Ashorn, P. (2013). Feasibility and validity of the ActiGraph GT3X accelerometer in measuring physical activity of Malawian toddlers. *Acta Paediatrica, 102*(12), 1192–1198. flh.
- Reilly, J. J. (2008). Physical activity, sedentary behaviour and energy balance in the preschool child: Opportunities for early obesity prevention. *Proceedings of the Nutrition Society, 67*(3), 317–325. flh.
- Sirard, J. R., Trost, S. G., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2005). Calibration and Evaluation of an Objective Measure of Physical Activity in Preschool Children. *Journal of Physical Activity and Health, 2*(3), 345–357. <https://doi.org/10.1123/jpah.2.3.345>
- Telama, R., Yang, X., Viikari, J., Välimäki, I., Wanne, O., & Raitakari, O. (2005). Physical activity from childhood to adulthood. *American Journal of Preventive Medicine, 28*(3), 267–273. <https://doi.org/10.1016/j.amepre.2004.12.003>
- Trost, S. G., Fees, B. S., Haar, S. J., Murray, A. D., & Crowe, L. K. (2012). Identification and Validity of Accelerometer Cut-Points for Toddlers. *Obesity, 20*(11), 2317–2319. <https://doi.org/10.1038/oby.2011.364>
- Van Cauwenberghe, E., Gubbels, J., De Bourdeaudhuij, I., & Cardon, G. (2011). Feasibility and validity of accelerometer measurements to assess physical activity in toddlers. *International Journal of Behavioral Nutrition and Physical Activity, 8*(1), 67. <https://doi.org/10.1186/1479-5868-8-67>

CHAPTER 6: OVERALL DISCUSSION AND CONCLUSIONS

Discussion

Physical activity (PA) engagement as early as three years old is related to PA levels in early and late adulthood (Telama et al., 2005). Furthermore, engaging in sufficient levels of PA in the early years (0 – 4 years) is associated with positive physical, cognitive and socioemotional outcomes (Lee et al., 2017). These positive associations have recently led to governing bodies, including the World Health Organization, publishing movement guidelines for children younger than 5 years old (Willumsen & Bull, 2020). Along with sleep and screen time recommendations, it is recommended that toddlers (1 – 2 years old) engage in 180-minutes of PA at any intensity (light, moderate, or vigorous). These recommendations make it imperative that we can accurately assess PA in toddlers.

Previous studies have utilized methods of activity assessments in preschoolers and applied them to toddlers (Bruijns et al., 2020; Dinkel et al., 2019; Fees et al., 2015). However, toddlers are at unique stage of motor and cognitive skill development (Colson & Dworkin, 1997). Most young children begin walking independently by 18-months, however, those 12 – 18 months may need assistance walking or may not be able to translocate at all. Motor skills such as hopping, kicking, or jumping do not occur until 24 – 36 months. Additionally, fine motor skills, such as successfully building with blocks, writing, or drawing are not achieved until around 36-months (Colson & Dworkin, 1997). These salient differences in development suggest that PA in toddlers may need to be assessed differently than preschoolers.

The focus of this dissertation was the assessment of PA in toddlers, which may enhance how we promote PA in this age group. The first aim of this dissertation was to systematically review the literature on PA assessment methods that have been developed, validated, or cross-

validated in toddlers (Chapter 3). We hypothesized that the literature would be limited, and accelerometry would be the most common assessment method, paired with cut-point-based data reduction. The reviewed studies (N=16) provided extensive evidence to support our hypotheses that the methods available were limited. Of the methods available, ten were developed using accelerometers (Costa et al., 2014; Haar et al., 2013; Hager et al., 2016; Johansson et al., 2015; Kelly & Villalpando, 2016; Klesges & Klesges, 1987; Oftedal et al., 2014; Pulakka et al., 2013; Trost et al., 2012; van Cauwenberghe E et al., 2011) and only four were survey based (Bingham et al., 2016; Henriksson et al., 2020; Sarker et al., 2015; Tulve et al., 2007). Lastly, three studies aimed to cross-validate preschool developed accelerometer cut-points in toddlers (Altenburg et al., 2022; Pereira et al., 2020; Van Cauwenberghe et al., 2011).

To our knowledge, this is the first study to systematically review all literature assessing PA in toddlers using any method (i.e., survey, device-based) rather than a single method (i.e., device-based). In doing so, we concluded that there are limited studies that focus on toddler activity. This review will allow researchers to easily determine next steps in toddler activity assessment. In addition to highlighting the limited amount of research in this area, this literature review called to attention several other gaps in the literature on this topic.

One of the gaps is that a direct observation (DO) tool has not yet been developed for use in toddlers. DO is commonly used as a criterion measure of PA in young children (Trost, 2007). The Observational System for Recording Physical Activity in Children – Preschool version (OSRAC-P) (Brown et al., 2006) has been used to assess toddler PA previously. In addition to assessing PA levels, the OSRAC systems include categories to describe other aspects of behavior such as activity type and context, group composition, and activity prompts (Brown et al., 2006;

McIver et al., 2009; Schenkelberg et al., 2021). Previously, the OSRAC-P was used to assess toddler behavior, however, research concluded that it was insufficient, and that additional categories and codes had to be added (Dinkel et al., 2019; Fees et al., 2015). Therefore, the second aim of this dissertation was to develop the Observational System for Recording Physical Activity in Children – Toddlers (OSRAC-T) and examine its reliability (Chapter 4).

The final version of the OSRAC-T included several additional codes and categories. To account for the unique stage of development, a ‘tantrum’ code was added to indoor and outdoor play context and the ‘crawl’ activity type was adapted to ‘crawl/creep/scoot’. Lastly, a ‘support’ category with the following codes: non-weight bearing, partial weight-bearing, weight-bearing, and cannot tell support was added. Previous research has assessed accelerometry counts when walking v. being carried (Kwon et al., 2019). When walking independently, counts were significantly lower than when a toddler was carried (49 counts/5-seconds v. 144 counts/5-seconds). This trend did not exist when a toddler was riding in a wagon or stroller, but the importance of noting external support is crucial in this population (Kwon et al., 2019). Overall, toddlers spent 70.0% of time in sedentary time and 10.0% of time engaged in MVPA while at childcare. Across 7,757 30-second observation intervals, inter-rater reliability was reported to be moderate – strong across all categories ($k = 0.460-0.693$) except activity initiation ($k = 0.349$) and PA prompt ($k = 0.276$). Reliability in this study was lower compared to other variations of the original OSRAC tool (Brown et al., 2006; Schenkelberg et al., 2021).

Inter-rater reliability for indoor activity context was lower in our study compared to other OSRAC systems ($\kappa = 0.488$ v. $0.90-0.99$). The discrepancy in classification may be due to the activity of onlooking. Onlooking is a behavior that is defined by idly looking or sitting,

meaning toddlers (2 – 3 years) are commonly acting as spectators rather than engaging (Parten, 1993). Previous studies have reported that toddlers spent 15% of time engaging in onlooking (Fees et al., 2015). When looking further at the differences in agreement, many included transition being coded by one observer and varying other contexts being coded by the other. The rules in the OSRAC systems state that the indoor/outdoor context a child is engaging in must be coded by what the child is doing or by what center the child is located in if they are not engaging with any material. Transition includes wandering, mean a child is not engaging with material or in a defined center. Because toddlers were frequently observing other children play, rather than engaging, it was difficult to decipher what context to code which may account for the differences in categorization. This commonly manifested as one coder selecting transition, while the other selected a more specific category. Outdoor, there is no transition code, only open space (child in an open area not defined and not engaging with any material), making it easier to classify this type of behavior outside. To mitigate these differences, we plan to better define the transition code as it relates to toddler PA in the OSRAC -T training manual.

This is the first observation tool created for toddlers that includes age-specific items. The ability to assess the unique behaviors and patterns of toddler PA has many implications for the field. As stated previously, we know that toddlers are at varying stages of development compared to not only older preschool children, but each other. Guidelines currently exist for classroom design in younger and older toddler rooms at childcare centers (*Child Care Center Design Guide*, 2003). These guidelines are based on developmental stages and are different than preschool guidelines. However, no study has assessed the differences in activity levels promoted by suggestions (i.e., classroom locations, play objects). This tool can be used to

assess daily behavior and better inform spaces that not only optimize fine motor and social skills, but also motor skills and physical activity levels. Later research could use the results from this tool to inform toddler activity interventions.

Despite the clear benefits of utilizing DO, due to the rich contextual information that can be collected, the process of DO is time intensive and requires extensive training prior to use (Trost, 2007). Additionally, due to the differing observation windows, commonly a 5-second observation followed by 25-seconds to record, the sporadic nature of child's PA is difficult to capture. Device-based methods, such as accelerometers, are better able to capture these sporadic movement patterns and can be used to assess multiple children at once. This brings us to the third aim of this dissertation, to cross-validate various cut-points for assessing sedentary time (ST) and PA in toddlers using hip- and wrist-placed accelerometers and vertical axis and vector magnitude data (Chapter 5).

Using the findings from our systematic review, twenty-five individual classification methods were identified. These were all cut-point methods, comprised of twenty-three hip-based and two wrist-based methods. Nine methods used vertical axis data, while fourteen used vector magnitude data. All cut-points were used in their calibrated epoch and compared to DO (5-second observe/25-second record intervals). Only two methods were found to be with $\pm 10\%$ equivalence of DO in assessing sedentary time. Similar to other studies, in the current study, ST was underestimated, while light and moderate-to-vigorous physical activity (MVPA) were overestimated compared to DO (Altenburg et al., 2022). These findings are opposite to those that used accelerometry as a criterion measure, where sedentary time was overestimated (Pereira et al., 2020).

Estimates across both placements and data axes varied greatly depending on the cut-point applied. This is not entirely surprising because previous literature has reported no clear advantage to using uniaxial or triaxial monitors in preschoolers (Adolph et al., 2012; Johansson et al., 2015; Pulakka et al., 2013). Previous literature suggests that hip-worn monitors are more feasible for use in toddlers, but accuracy compared to a wrist-worn device could not be determined due to the lack of available wrist cut-points (Kwon et al., 2019). This is further supported by our study. To date, only two sets of cut-points have been developed using wrist-worn data for in toddlers (Johansson et al., 2015). Which wear location is more accurate is also unclear in older children. Wrist cut-points have been reported to be more valid; however, there was also reasonable agreement with hip-worn monitors (Hislop et al., 2012). Because these methods are limited, we were unable to draw conclusions between placements. However, bias in wrist-worn devices compared to DO and hip-worn devices and DO appeared similar.

This is the first study to compare toddler and preschool accelerometer cut-points identified using a systematic review to DO. This study suggests that, in toddlers, a hip- or wrist-worn device is not more advantageous than the other. This could have implications for future studies performing 24-hour monitoring. Our study emphasizes the complexity in comparing cut-point methods due to the different calibration methods across studies. The lack of equivalency in our study is thought to be because of these differences. As in older children, we can hypothesize that more advanced methods that utilize data collected in a raw form may be valuable in toddlers. Despite the many novel strengths of this dissertation, it and the field of activity assessment, are not without limitations.

Limitations

Future work should incorporate toddlers from more diverse groups and different geographical regions. These data were collected using a small, homogenous sample from three childcare centers in a Midwest city. The sample was primarily white/Caucasian and all toddlers came from a family where the parent reported completing at least some college or technical school. This could be because two of the three childcare centers were affiliated with Michigan State University. These methods should be implemented in other types of childcare centers (i.e., nature-based, in-home, daycare) or national programs such as Head Start programs that provide childcare to low income families (Office of the Administration for Children and Families, 2023). These methods should also be implemented in other countries and regions across the world.

Data for Chapters 3 and 4 utilized were collected using a 5-second observation to represent each 30-second interval to be comparable to other OSRAC adaptations. There are several limitations to this decision. As discussed, young children typically engage in sporadic patterns of vigorous activity. Researchers frequently choose to use smaller epoch lengths, such as 5- or 15-seconds, to better capture this activity when using accelerometry. Studies with direct observation have also used video coding to eliminate the need for recording time. This allows for each 15-second interval to be observed to better assess activity (Troost et al., 2012; van Cauwenberghe E et al., 2011). It is unknown if using the 5-second observation to represent each 30-second is a valid assessment method in toddlers. Future studies should assess direct observation using smaller observation intervals to determine if shorter intervals result in more

accurate assessments. Our data suggest that this could be done using video observations that can be paused during coding rather than in-person sessions.

Another consideration that may limit our results is the rules outlined in the OSRAC tools. When using the OSRAC the highest PA level in the 5-second observation window is coded and all categories are coded according to that activity level (Brown et al., 2006). That highest level of PA may mask other levels of activity in that interval. Another issue arises with PA levels 1-stationary and 2-stationary w/limbs. It is common to miscode one of the categories as the other (Brown et al., 2006; McIver et al., 2009; Schenkelberg et al., 2021). Many studies consider 1 & 2 codes as sedentary; however, it has been suggested that 2-stationary w/ limbs requires an energy expenditure over resting, leading some researchers to classify 2-stationary w/ limbs as light activity (van Cauwenberghe E et al., 2011). Whichever decision is made should be considered when comparing DO data to other metrics. The Children's Activity Rating Scale, which is utilized to classify PA levels in the OSRAC, was validated using heart rate and indirect calorimetry in preschoolers (Puhl et al., 1990). Future studies should determine metabolic demand of daily activities (i.e., sitting, standing, walking, running, jumping) in toddlers to determine which classification method is best.

One last consideration is the metrics used to assess the third aim of this dissertation. We know that accelerometer cut-points are difficult to compare due to the differences in methodological approaches (Bruijns et al., 2020; Trost, 2007). In Chapter 5, we decided to use cut-points as they were calibrated and compare the percentage of time spent in each activity intensity to the 5-second DO intervals. Aside from the previously mentioned limitations with the classification of DO, it is difficult to compare PA levels when using different epoch levels. In

2 – 5-year-old children, significantly different estimates of MVPA between 5s and 60s epochs have been reported (Vale et al., 2009). Therefore, we may not be able to compare activity collected at 5-seconds to activity collected at 15- or 30-seconds. Ultimately, we should consider the consistent report that cut-points are not equivalent to criterion measures, regardless of data axes or monitor placement (Altenburg et al., 2022; Pereira et al., 2020; van Cauwenberghe E et al., 2011). It is important to consider that machine learning approaches have been shown to be advantageous compared to cut-points in preschoolers (Ahmadi & Trost, 2022). However, we should not eliminate cut-point methods, we may just need better guidance on the best way to calibrate cut-points. For example, standardization across methods would be beneficial in both application and development. Future research should aim to standardize approaches to cut-point calibration. Additionally, future studies should implement these more advanced methods of activity assessment in toddlers to determine if that could improve accuracy. Finally, direct observation and device-based methods should be used simultaneously. The rich information we can collect about toddler behavior, in addition to the less intensive assessment of PA levels can provide insightful information that could be used to inform interventions to improve activity engagement (Carson et al., 2024)

Conclusions and Future Directions

Recently, Carson et al. (2024) released *Future Directions for Movement Behavior Research in the Early Years* assessing the current progress of research on children in the early years (0 – 4 years). They highlighted the need for a developed instrument to assess movement behaviors and standardized protocols for measurement instruments to enable comparison across studies. This dissertation aimed to begin addressing these shortcomings. We

systematically reviewed the current literature to determine where the field of toddler activity assessment is currently at and what methods exist. This highlighted the limited methods to assess activity in this population. There is a need for valid and reliable survey-based instruments that can be used in large epidemiological studies. Future research should focus on the development of these surveys in diverse population in different geographical regions.

We also determined there was a lack of DO methods to assess toddler activity. The OSRAC-T is the first tool developed that is age-appropriate for those 12-36 months old. This instrument can be used to assess various aspects of toddler physical activity. The OSRAC-T should be used in more diverse groups and geographic regions to begin assessing validity. We can also look at childcare center design and begin to work with specialists to better optimize both development and activity levels. Lastly, we highlighted the complexity in applying device-based assessment methods to an independent sample. It may be more important to consider calibration metrics, rather than device wear location or data axes. When using accelerometers in this age group, it may be beneficial to consider methods other than cut-points that utilize RAW accelerometer data.

PA in early childhood has been shown to predict activity engagement in early-to-late adulthood (Telama et al., 2005). The importance of toddler activity has been highlighted in recent literature. Young children meeting 24-hour activity guidelines is associated with better physical, cognitive, and socio-emotional health (Kuzik et al., 2015; Rollo et al., 2020). This dissertation provides new methodologies and highlights methodological considerations for assessing PA in toddlers. When used together, these can both inform future research directions.

REFERENCES

- Adolph, A. L., Puyau, M. R., Vohra, F. A., Nicklas, T. A., Zakeri, I. F., & Butte, N. F. (2012). Validation of Uniaxial and Triaxial Accelerometers for the Assessment of Physical Activity in Preschool Children. *Journal of Physical Activity and Health, 9*(7), 944–953. <https://doi.org/10.1123/jpah.9.7.944>
- Ahmadi, M. N., & Trost, S. G. (2022). Device-based measurement of physical activity in preschoolers: Comparison of machine learning and cut point methods. *PLOS ONE, 17*(4), e0266970. <https://doi.org/10.1371/journal.pone.0266970>
- Altenburg, T. M., de Vries, L., op den Buijsch, R., Eyre, E., Dobell, A., Duncan, M., & Chinapaw, M. J. M. (2022). Cross-validation of cut-points in preschool children using different accelerometer placements and data axes. *Journal of Sports Sciences, 40*(4), 379–385. <https://doi.org/10.1080/02640414.2021.1994726>
- Bingham, D., Collings, P., Clemes, S., Costa, S., Santorelli, G., Griffiths, P., & Barber, S. (2016). Reliability and Validity of the Early Years Physical Activity Questionnaire (EY-PAQ). *Sports, 4*(2), 30. <https://doi.org/10.3390/sports4020030>
- Brown, W. H., Pfeiffer, K. A., McIver, K. L., Dowda, M., Almeida, J. M. C. A., & Pate, R. R. (2006). Assessing Preschool Children's Physical Activity: The Observational System for Recording Physical Activity in Children-Preschool Version. *Research Quarterly for Exercise and Sport, 77*(2), 167–176. <https://doi.org/10.1080/02701367.2006.10599351>
- Bruijns, B. A., Truelove, S., Johnson, A. M., Gilliland, J., & Tucker, P. (2020). 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, 17*(1), 14. <https://doi.org/10.1186/s12966-020-0912-4>
- Carson, V., Draper, C. E., Okely, A., Reilly, J. J., & Tremblay, M. S. (2024). Future Directions for Movement Behavior Research in the Early Years. *Journal of Physical Activity and Health, 21*(3), 218–221. <https://doi.org/10.1123/jpah.2023-0679>
- Child Care Center Design Guide*. (2003). U.S. General Services Administration.
- Colson, E., & Dworkin, P. (1997). *Toddler Development, 18*(8), 255–259.
- Costa, S., Barber, S. E., Cameron, N., & Clemes, S. A. (2014). Calibration and validation of the ActiGraph GT3X+ in 2–3 year olds. *Journal of Science and Medicine in Sport, 17*(6), 617–622. <https://doi.org/10.1016/j.jsams.2013.11.005>
- Dinkel, D., Snyder, K., Patterson, T., Warehime, S., Kuhn, M., & Wisneski, D. (2019). An exploration of infant and toddler unstructured outdoor play. *European Early Childhood*

Education Research Journal, 27(2), 257–271.
<https://doi.org/10.1080/1350293X.2019.1579550>

Fees, B. S., Fischer, E., Haar, S., & Crowe, L. K. (2015). Toddler Activity Intensity During Indoor Free-Play: Stand and Watch. *Journal of Nutrition Education and Behavior*, 47(2), 170–175. <https://doi.org/10.1016/j.jneb.2014.08.015>

Haar, S., Fees, B., Trost, S., Crowe, L. K., & Murray, A. (2013). Design of a Garment for Data Collection of Toddler Language and Physical Activity. *Clothing and Textiles Research Journal*, 31(2), 125–140. <https://doi.org/10.1177/0887302X13478161>

Hager, E. R., Gormley, C. E., Latta, L. W., Treuth, M. S., Caulfield, L. E., & Black, M. M. (2016). Toddler physical activity study: Laboratory and community studies to evaluate accelerometer validity and correlates. *BMC Public Health*, 16(1), 1–10. <https://doi.org/10.1186/s12889-020-09808-w>

Henriksson, H., Alexandrou, C., Henriksson, P., Henström, M., Bendtsen, M., Thomas, K., Müssener, U., Nilsen, P., & Löf, M. (2020). MINISTOP 2.0: A smartphone app integrated in primary child health care to promote healthy diet and physical activity behaviours and prevent obesity in preschool-aged children: Protocol for a hybrid design effectiveness-implementation study. *BMC Public Health*, 20(1), 1–11. <https://doi.org/10.1186/s12889-020-09808-w>

Hislop, J. F., Bulley, C., Mercer, T. H., & Reilly, J. J. (2012). Comparison of Epoch and Uniaxial Versus Triaxial Accelerometers in the Measurement of Physical Activity in Preschool Children: A Validation Study. *Pediatric Exercise Science*, 24(3), 450–460. <https://doi.org/10.1186/s12889-020-09808-w>

Johansson, E., Ekelund, U., Nero, H., Marcus, C., & Hagströmer, M. (2015). Calibration and cross-validation of a wrist-worn ActiGraph in young preschoolers: Calibration of ActiGraph in toddlers. *Pediatric Obesity*, 10(1), 1–6. <https://doi.org/10.1111/j.2047-6310.2013.00213.x>

Kelly, L. A., & Villalpando, J. (2016). Development of ActiGraph GT1M Accelerometer Cut-Points for Young Children Aged 12-36 Months. *Journal of Athletic Enhancement*, 5(4). <https://doi.org/10.4172/2324-9080.1000233>

Klesges, L. M., & Klesges, R. C. (1987). *The assessment of children's physical activity: A comparison of methods*. 19(5), 511–517.

Kuzik, N., Clark, D., Ogden, N., Harber, V., & Carson, V. (2015). Physical activity and sedentary behaviour of toddlers and preschoolers in child care centres in Alberta, Canada. *Canadian Journal of Public Health*, 106(4), e178–e183. <https://doi.org/10.17269/cjph.106.4794>

- Kwon, S., Honegger, K., & Mason, M. (2019). Daily Physical Activity Among Toddlers: Hip and Wrist Accelerometer Assessments. *International Journal of Environmental Research and Public Health*, *16*(21), 4244. <https://doi.org/10.3390/ijerph16214244>
- Lee, E.-Y., Hesketh, K. D., Hunter, S., Kuzik, N., Rhodes, R. E., Rinaldi, C. M., Spence, J. C., & Carson, V. (2017). Meeting new Canadian 24-Hour Movement Guidelines for the Early Years and associations with adiposity among toddlers living in Edmonton, Canada. *BMC Public Health*, *17*(S5), 840. <https://doi.org/10.1186/s12889-017-4855-x>
- McIver, K. L., Brown, W. H., Pfeiffer, K. A., Dowda, M., & Pate, R. R. (2009). Assessing Children's Physical Activity in their Homes: The Observational System for Recording Physical Activity in Children-Home. *Journal of Applied Behavior Analysis*, *42*(1), 1–16. eax.
- Office of the Administration for Children and Families. (2023, June 30). *Head Start Services*. Office of Head Start. <https://www.acf.hhs.gov/ohs/about/head-start>
- Oftedal, S., Bell, K. L., Davies, P. S. W., Ware, R. S., & Boyd, R. N. (2014). Validation of Accelerometer Cut Points in Toddlers with and without Cerebral Palsy. *Medicine & Science in Sports & Exercise*, *46*(9), 1808–1815. <https://doi.org/10.1249/MSS.0000000000000299>
- Parten, M. B. (1993). *Social play among peeschool children*.
- Pereira, J. R., Sousa-Sá, E., Zhang, Z., Cliff, D. P., & Santos, R. (2020). Concurrent validity of the ActiGraph GT3X+ and activPAL for assessing sedentary behaviour in 2–3-year-old children under free-living conditions. *Journal of Science and Medicine in Sport*, *23*(2), 151–156. <https://doi.org/10.1016/j.jsams.2019.08.009>
- Puhl, J., Greaves, K., Hoyt, M., & Baranowski, T. (1990). Children's Activity Rating Scale (CARS): Description and Calibration. *Research Quarterly for Exercise and Sport*, *61*(1), 26–36. <https://doi.org/10.1080/02701367.1990.10607475>
- Pulakka, A., Cheung, Y., Ashorn, U., Penpraze, V., Maleta, K., Phuka, J., & Ashorn, P. (2013). Feasibility and validity of the ActiGraph GT3X accelerometer in measuring physical activity of Malawian toddlers. *Acta Paediatrica*, *102*(12), 1192–1198. <https://doi.org/10.1111/apa.12412>
- Rollo, S., Antsygina, O., & Tremblay, M. S. (2020). The whole day matters: Understanding 24-hour movement guideline adherence and relationships with health indicators across the lifespan. *Journal of Sport and Health Science*, *9*(6), 493–510. <https://doi.org/10.1016/j.jshs.2020.07.004>
- Sarker, H., Anderson, L., Borkhoff, C., Abreo, K., Tremblay, M., Lebovic, G., Maguire, J., Parkin, P., & Birken, C. (2015). Validation of Parent-Reported Physical and Sedentary Activity by

- Accelerometry in Young Children. *Canadian Journal of Diabetes*, 39, S44–S44. rzh.
<https://doi.org/10.1016/j.jcjd.2015.01.169>
- Schenkelberg, M. A., Brown, W. H., McIver, K. L., & Pate, R. R. (2021). An observation system to assess physical activity of children with developmental disabilities and delays in preschool. *Disability and Health Journal*, 14(2), 101008.
<https://doi.org/10.1016/j.dhjo.2020.101008>
- Telama, R., Yang, X., Viikari, J., Välimäki, I., Wanne, O., & Raitakari, O. (2005). Physical activity from childhood to adulthood. *American Journal of Preventive Medicine*, 28(3), 267–273.
<https://doi.org/10.1016/j.amepre.2004.12.003>
- Trost, S. G. (2007). State of the Art Reviews: Measurement of Physical Activity in Children and Adolescents. *American Journal of Lifestyle Medicine*, 1(4), 299–314.
<https://doi.org/10.1177/1559827607301686>
- Trost, S. G., Fees, B. S., Haar, S. J., Murray, A. D., & Crowe, L. K. (2012). Identification and Validity of Accelerometer Cut-Points for Toddlers. *Obesity*, 20(11), 2317–2319.
<https://doi.org/10.1038/oby.2011.364>
- Tulve, N. S., Jones, P. A., McCurdy, T., & Croghan, C. W. (2007). A Pilot Study Using an Accelerometer to Evaluate a Caregiver’s Interpretation of an Infant or Toddler’s Activity Level as Recorded in a Time Activity Diary. *Research Quarterly for Exercise & Sport*, 78(4), 375–383. trh.
- Vale, S., Santos, R., Silva, P., Soares-Miranda, L., & Mota, J. (2009). *Preschool Children Physical Activity Measurement: Importance of Epoch Length Choice*. 21(4), 413–420.
<https://doi.org/10.1123/pes.21.4.413>
- Van Cauwenberghe, E., Gubbels, J., De Bourdeaudhuij, I., & Cardon, G. (2011). Feasibility and validity of accelerometer measurements to assess physical activity in toddlers. *International Journal of Behavioral Nutrition and Physical Activity*, 8(1), 67.
<https://doi.org/10.1186/1479-5868-8-67>
- van Cauwenberghe E, Labarque V, Trost SG, de Bourdeaudhuij I, & Cardon G. (2011). Calibration and comparison of accelerometer cut points in preschool children. *International Journal of Pediatric Obesity*, 6(2–2), e582–9. rzh.
<https://doi.org/10.3109/17477166.2010.526223>
- Willumsen, J., & Bull, F. (2020). Development of WHO Guidelines on Physical Activity, Sedentary Behavior, and Sleep for Children Less Than 5 Years of Age. *Journal of Physical Activity and Health*, 17(1), 96–100. <https://doi.org/10.1123/jpah.2019-0457>

APPENDIX A. NOVEL PHYSICAL ACTIVITY ASSESSMENT METHOD STUDY CHARACTERISTICS

Study	Sample N N male Age range OW/OB Health Status	Type of Validation	Novel Method	Criterion	Activities	Additional Information
Henriksson (2014)	75 44 18-36 m 30 Healthy	Concurrent	Activity Diary	Doubly- labelled water and indirect calorimetry	Free-living (14 days)	
Klesges (1985)	30 17 24-48 m 9 Healthy	One- Sample Cross Validation	Calectro S2-294	Direct observation – Fargo time sampling survey	Free-living	
Trost (2012)	22 8 16-35 m NR Healthy	Independe nt Sample Cross- Validation	Hip Cut- point	Direct observation – CARS	Free-play (20-min)	
Johansson (2015)	26 15 15-36 m 7 Healthy	One- Sample Cross Validation	Hip and Wrist Cut- point	Direct observation – CARS	watching cartoon, drawing, running obstacle course and 15- minutes outdoor free play	
Costa (2014)	26 13 24-47 m 4 Healthy	Independe nt Sample Cross- Validation	Hip Cut- point	Direct observation – CARS	sitting/lay ing, sitting while throwing balloons or drawing, walking or	

					rolling slowly, walking and dancing moderately, running and jumping	
Haar (2013)	40 18 24-47 m NR Healthy	Independent Sample Cross-Validation	ActiGraph and ActiCal	Direct observation - CARS	Free-play	
Kelly (2016)	23 10 12-36m NR Healthy	One-Sample Cross Validation	Hip Cut-points	Direct observation - CPAF	Resting, walking, dancing, jumping, crawling, skipping, running, climbing, swinging	
Pulakka (2013)	56 24 16-18 m NR Healthy	One-Sample Cross Validation	Hip Cut-points	Direct observation - CPAF	Free-play (120-min)	
Hager (2016)	24 14 12-36 m 3 Healthy	Independent Sample Cross-Validation	Ankle Cut-points	Direct observation - CARS	Watching tv (sitting), listening to book (sitting), table games, puzzles, and play doh (standing), , imaginary play (walking),	

					ball games (running), tag (running)	
Oftedal (2014)	103 61 18-36 m NR Healthy and Cerebral Palsy	Independent Sample Cross-Validation	Hip Cut-points	Direct observation – sedentary or non-sedentary	Free-play (20-30 min)	
Altenburg (2021)	63 28 24-48 m 14 Healthy	Cross-Validation	NA	Direct observation – OSRAC-P	Free-play (1-hour)	Applied Sirard (2005), Dobell (2019), Butte (2014), Johansson (2016), Pate (2006) cut-points
Pereira (2020)	60 30 22-42 m NR Healthy	Concurrent	Hip ActiGraph GT3X+ RAW data	activPAL	Free-play (1 hour)	Applied Costa (2014), Trost, Kelly (2016), Evenson (2008), Pate (2006), Reilly (2003), Sirard (2005), and Van Cauwenbergh (2011) cut-points
Van Cauwenbergh (2011)	31 17 12-26 m NR Healthy	Cross-Validation	Hip ActiGraph GT1M counts	Direct observation – OSRAC-P	Free-play 19.5-60 min)	Applied Pate (2006), Sirard (2005), Van Cauwenbergh (2011) cut-points
Bingham (2016)	196 99	Convergent	Questionnaire	ActiGraph GT3X+	Free-living (7-days)	

	18-48 m NR NR			(Costa, 2014 and Pate, 2006)		
Sarker (2015)	87 40 < 6 y 15 Healthy	Convergent	Questionnaire	Actical (Adolph, 2012)	Free-living (7-days)	
Tulve (2007)	9 5 4-17 m NR Healthy	Convergent	Activity diary	Actical (counts)	Free-play	

OW/OB: number overweight/obese; NR: not reported

APPENDIX B. CATEGORIES AND CODES OUTLINED IN THE OSRAC-T

Category	Code	Category	Code
<i>Activity Level</i>	1-Stationary 2-Limbs 3-Slow-Easy 4-Moderate 5-Fast Can't Tell		Gross Motor Group Time Large Blocks Manipulative Music Nap Self-Care Snacks Sociodramatic Tantrum Teacher Arranged Time Out Transition Videos Other N/A Can't Tell
<i>Activity Type</i>	Climb Crawl/Creep/Scout Dance Jump/Skip Lie Down Pull/Push R & T Ride Rock Roll Run Sit/Squat Stand Swim Swing Throw Walk Other Can't Tell	<i>Outdoor Context</i>	Ball Books/Preacademic Fixed Game OpenSpace Pool Portable Sandbox Snacks Sociodramatic Props Tantrum Teacher Arranged Time Out Wheel Other N/A Can't Tell
<i>Location</i>	Indoor Outdoor Transition Can't Tell		
<i>Group Composition</i>	Solitary 1-1 Adult 1-1 Peer Group Adult Group Can't Tell	<i>Prompt</i>	None TP-I TP-D PP-I PP-D
<i>Initiator of Activity</i>	Adult Child Can't Tell	<i>Support</i>	Non-weight Bearing Partial-weight Bearing Weight Bearing Can't Tell
<i>Indoor Context</i>	Art Books/Preacademic		