# PHYSICAL ACTIVITY ASSESSMENTS THROUGHOUT PREGNANCY AND POSTPARTUM

By

Michelle Reese Conway

# A DISSERTATION

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

Kinesiology – Doctor of Philosophy

#### PUBLIC ABSTRACT

# PHYSICAL ACTIVITY ASSESSMENTS THROUGHOUT PREGNANCY AND POSTPARTUM

By

### Michelle Reese Conway

Participating in physical activity during and after pregnancy has many benefits including a reduced risk of preeclampsia, gestational diabetes, excessive weight gain, and cesarean delivery. To determine the amount of physical activity that is best for pregnant and postpartum women and their children, more research is needed that focuses on the consistency and accuracy of physical activity data collection techniques. The purpose of this dissertation was to assess the 1) consistency and accuracy of three physical activity devices, 2) agreement of how much physical activity is measured between a self-report questionnaire and a physical activity device worn at the hip and ankle, and 3) ability of women to recall their physical activity historically at three prior time points (21 and 32 weeks gestation, and 12 weeks postpartum).

Study participants included 48 women from the Mid-Michigan area who were  $\geq 18$  years, had the ability to read and speak English, and had low-risk pregnancies determined by a health care provider. For the first study, these women completed seven different activities (folding laundry, dusting, sweeping, picking up toys, walking on the treadmill, walking in the hallway, and a workout video) for five minutes each when they came to the laboratory, twice, one week apart at 21 and 32 weeks gestation and 12 weeks postpartum (total of six lab visits). Each time they came to the lab, the women wore an Oxycon Mobile portable metabolic analyzer (gold standard), ActiGraph accelerometer on their right hip and ankle, Omron pedometer on their right hip, and SenseWear armband over their left arm. To determine how consistent the devices were, the measurements from each device in lab visits 1, 3, and 5 were compared to the measurements in lab visits 2, 4, and 6, respectively. To determine accuracy of the ActiGraph, Omron, and SenseWear, their measurements were compared to the Oxycon. For the second study, the women wore the ActiGraphs for one week at home and then completed the Pregnancy Physical Activity Questionnaire (PPAQ) to estimate their prior week's physical activity, at 21 and 32 weeks gestation and 12 weeks postpartum. Time was spent in light, moderate, and vigorous physical activity was compared between the ActiGraphs and the PPAQ. Finally, for the third study, these 48 women were contacted again between two months and eight years after giving birth, based on when they entered our study. They completed three separate PPAQs to recall their physical activity at 21 and 32 weeks gestation and 12 weeks postpartum. How much time was spent in light, moderate, and vigorous physical activity and the total amount of physical activity was compared between the PPAQ completed between two months and eight years ago and the PPAQ completed recently.

Results showed the devices were consistent and accurate when worn by women completing activities they would normally do around the house and exercising. The PPAQ underestimated percent time spent in light physical activity and overestimated percent time spent in moderate physical activity compared to the ActiGraph. The women participated in very little vigorous physical activity according to both the PPAQ and ActiGraph. When assessing how well women could recall their activity when they were pregnant many months to years previously, women tended to underestimate their total and moderate activity and overestimate their light activity. This dissertation will help clarify the ideal amount of physical activity for women to participate in throughout and after pregnancy, and the effects of this physical activity on the health of the mother and child.

#### ABSTRACT

# PHYSICAL ACTIVITY ASSESSMENTS THROUGHOUT PREGNANCY AND POSTPARTUM

By

## Michelle Reese Conway

Benefits from physical activity participation during pregnancy include a reduced risk of preeclampsia, gestational diabetes, excessive weight gain, and cesarean delivery. More research about the reliability and validity of physical activity data collection techniques is needed to clarify the optimal intensity and frequency of physical activity for women to perform throughout and after pregnancy and the effects of this physical activity on maternal and birth outcomes. The purpose of this dissertation was to assess the 1) reliability and validity of three physical activity devices, 2) correlation between a self-report questionnaire and a physical activity device worn at the hip and ankle, and 3) validity of women recalling their physical activity at three time points (21 and 32 weeks gestation, and 12 weeks postpartum).

The sample consisted of 48 women from the Mid-Michigan area who were  $\geq$  18 years, had the ability to read and speak English, and had low-risk pregnancies determined by a health care provider. Women completed seven different activities of daily living and locomotor activities for five minutes each during two laboratory visits, one week apart, at 21 and 32 weeks gestation and 12 weeks postpartum (total of six lab visits). At each lab visit, the women wore an Oxycon Mobile portable metabolic analyzer (criterion), ActiGraph accelerometer on their right hip and ankle, Omron pedometer on their right hip, and SenseWear armband over their left triceps. The women wore the ActiGraphs for one week of free-living activity and completed the Pregnancy Physical Activity Questionnaire (PPAQ) to estimate the past week's physical activity at the same three time points. These 48 women were contacted again between five months and eight years after giving birth to complete three separate PPAQs to recall their physical activity at 21 and 32 weeks gestation and 12 weeks postpartum.

Multi and single trial intraclass correlations and standard errors of measurement were calculated to assess device reliability. Pearson correlations were used to compare device validity to the criterion Oxycon Mobile. A two-way repeated measured analysis of variance and Spearman correlations (SCC) were conducted to examine the relationship between the ActiGraphs and PPAQ at each gestational age. Paired sample t-tests or Wilcoxon Sign-Rank tests, and SCCs were conducted between the original and recall PPAQ responses for total metabolic (MET) minutes per week and percent time in light, moderate, and vigorous intensity activity. Paired samples t-tests, Wilcoxon Sign-Rank tests, and SCCs were again conducted with the participants separated by enrollment date via a median split.

Test-retest reliability of the three devices was moderate to high as 66% of interclass and 88% of intraclass correlations were between 0.6 - 1.0. Comparison between the Oxycon and devices showed 46% of validity coefficients were between 0.6 - 0.79.

The PPAQ underestimated percent time spent in light physical activity and overestimated percent time spent in moderate physical activity compared to the ActiGraph worn at the hip. Very little vigorous physical activity was measured on any instrument. Low to moderate correlations were calculated between the PPAQ and both hip and ankle ActiGraphs at all three time points. When assessing the validity of the PPAQ for long term recall, women tended to underestimate their total MET-minutes per week and percent time spent in moderate intensity activity and overestimate the percent time spent in light intensity activity. These relationships did not change significantly as a function of time since birth.

Copyright by MICHELLE REESE CONWAY 2018

## ACKNOWLEDGEMENTS

I would like to thank my advisor, James Pivarnik, for his wonderful mentoring, humor, and dedication throughout my Michigan State career; and for being "the most different" advisor.

Thank you to my committee members, Karin Pfeiffer, Nicole Talge, and Karl Erickson, for even agreeing to read this many page long document and being willing to share your wisdom with me.

Thank you to my friends, inside and outside of the program, for your support, always making me laugh, and keeping me social.

I, of course, have to thank my cats. Thank you, Sass, for being the best moving buddy to Michigan and the happiest kitten. And thank you to Mags and Murph for being the weirdest cats but making coming home every day a happy event.

Finally, thank you to my mom, Lynn, who has listened to me no matter how dramatic I can be and always given helpful advice, and for supporting me unconditionally throughout my education and life.

# TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	X
CHAPTER ONE: INTRODUCTION	1
Research Aims	6
APPENDIX	8
REFERENCES	10
CHAPTER TWO: REVIEW OF THE LITERATURE	14
Introduction	14
Reliability of Physical Activity Devices During Pregnancy	16
Reliability of the SenseWear	17
Reliability of the Omron	17
Reliability of the ActiGraph	18
Summary	18
Validity of Physical Activity Devices During Pregnancy	19
Validity of the SenseWear	19
Validity of the Omron and Other Pedometers	20
Validity of the AcitGraph and Other Accelerometers Using Counts per	
Minute	23
Validity of the ActiGraph and Other Accelerometers for Step Counting	24
Summary	25
Correlations Between Self-Report Methods and Physical Activity Devices	
During Pregnancy	26
Correlations Between the PPAQ and ActiGraph	26
Correlations Between the PPAQ and Devices Other Than the ActiGraph	28
Correlations Between the ActiGraph and Self-Report Methods Other Than	
the PPAQ	29
Correlations Between all Self-Report Methods and Devices Other Than	
the PPAQ and ActiGraph at one Time Point During Pregnancy	31
Correlations Between all Self-Report Methods and Devices Other Than	
the PPAQ and ActiGraph Throughout Pregnancy	32
Summary	34
Validity of Recall Measures for Maternal Physical Activity	35
Reliability of the PPAQ for Recalling Maternal Physical Activity	36
Reliability of Other Self-Report Physical Activity Methods for Recalling	
Maternal Physical Activity	37
Long Term Physical Activity Recall in Non-Pregnant Adults	38
Summary	39
Conclusions	40
APPENDIX	42

REFERENCES	46
CHAPTER THREE PHYSICAL ACTIVITY DEVICE RELIABLITY AND	
VALDIITY DURING PREGNANCY AND POSTPARTUM	54
Abstract	54
Introduction	56
Methods	
Study Population and Recruitment	
Fauinment	58
Lyuphent	
Data Collection and Reduction	
Statistical Analysis	
Paliahility	
Validity	
Valluty	01
Deliebility	01
Validity	01
Valially	
APPENDIX	
QUESTIONNAIRE AND ACCELEROMETERS WORN DURING PREGNANCY	7
AND POSTPARTUM	
Abstract	
Introduction	
Methods	
Study Sample and Recruitment	
Equipment and Data Collection	
Data Reduction	
Statistical Analysis	
Results	
Discussion	
Conclusion	
APPENDIX	93
REFERENCES	
CHAPTER FIVE: VALIDITY OF THE PREGNANCY PHYSICAL ACTIVITY	
QUESTIONNAIRE FOR MATERNAL PHYSICAL ACTIVITY RECALL	
Abstract	101
Introduction	
Methods	
Study Sample and Recruitment	

Data Collection	105
Data Reduction	105
Statistical Analysis	106
Results	106
Discussion	108
Conclusion	110
APPENDIX	113
REFERENCES	116
CHAPTER SIX: SUMMARY AND CONCLUSIONS	119
Summary	119
Reliability and Validity of Physical Activity Devices	119
Correlations Between the Pregnancy Physical Activity Questionnaire	
and ActiGraph Accelerometers	120
Historical Recall Validity of the Pregnancy Physical Activity	
Questionnaire	123
Limitations	125
Strengths	126
Conclusions	127
REFERENCES	129

# LIST OF TABLES

Table 1.1. Devices and units of analysis used in the studies included in this dissertation	.9
Table 2.1.       Summary table of studies examining the validity of physical activity device when worn during pregnancy and postpartum	.43
Table 2.2. Spearman correlations between the Pregnancy Physical Activity Questionnaire (PPAQ) and ActiGraph data analyzed with various intensity cut points	44
Table 2.3.       Spearman correlations between various physical activity questionnaires and ActiGraph activity cut points during pregnancy	45
Table 3.1. Means and standard deviations (SD) at each time point, for each device	70
Table 3.2. Interclass reliability coefficients (via Pearson correlation) for the entire         35 minute visit, at each time points, for each device	71
Table 3.3.Multi-trial intraclass reliability coefficients (via ANOVA) for the entire 35 minute visit, at each time point, for each device	72
Table 3.4. Single trial reliability coefficients for the entire 35 minute visit, at each time point, for each device	73
Table 3.5. Standard error of measurement expressed in units and percent of the mean units (%) for the entire 35 min visit, at each time point, for each device	74
Table 3.6 Validity coefficients (via Pearson correlation) for the entire visit, for each time point, for each device when compared to relative VO <sub>2</sub>	75
Table 4.1. Spearman correlation coefficients between the Pregnancy Physical Activity Questionnaire (PPAQ) and ActiGraph worn at the hip at three time points during pregnancy and postpartum	97
Table 5.1. Means and standard deviations of total metabolic equivalent minutes per week (MET Min/Wk) and percent time spent in light, moderate, and vigorous physical activity from the original Pregnancy Physical Activity Questionnaire (PPAQ) and recall PPAQ at three time points during pregnancy and postpartum and by time intervals	14
Table 5.2. Spearman correlation coefficients between the original Pregnancy Physical Activity Questionnaire (PPAQ) and recall PPAQ and by time intervals	15

# LIST OF FIGURES

Figure 4.1. Percent time (standard error of the mean) spent in light, moderate, and vigorous physical activity at 21 weeks gestation. Light: < 2.9 METs, Moderate: 3.0 – 5.9 METs, Vigorous: ≥ 6.0. n = 36	94
Figure 4.2. Percent time (standard error of the mean) spent in light, moderate, and vigorous physical activity at 32 weeks gestation. Light: < 2.9 METs, Moderate: $3.0 - 5.9$ METs, Vigorous: $\geq 6.0$ . n = 35	95
Figure 4.3. Percent time (standard error of the mean) spent in light, moderate, and vigorous physical activity at 12 weeks postpartum. Light: < 2.9 METs, Moderate: $3.0 - 5.9$ METs, Vigorous: $\geq 6.0$ . n = 30	96

### **CHAPTER ONE: INTRODUCTION**

The United States Department of Health and Human Services (DHHS) recommends women to participate in 150 minutes of moderate intensity aerobic activity per week before, during, and after pregnancy, and the American College of Obstetricians and Gynecologists (ACOG) recommends 20 to 30 minutes of moderate exercise per day, most or all days of the week (1, 2). Benefits from physical activity participation during pregnancy include a reduced risk of preeclampsia, gestational diabetes, excessive weight gain, and cesarean delivery, while no negative effects have been established on the maternal-fetal dyad (1, 3–6). Physical activity during the postpartum period has also been shown to improve maternal cardiovascular fitness without affecting infant growth and milk production, and help women continue lifelong healthy habits (1). Although physical activity has been shown to benefit most pregnant women, only 13.8% of pregnant women in the United States meet the DHHS physical activity guidelines (7). To clarify the optimal intensity and frequency of physical activity for women to participate in throughout and after pregnancy and the effects of this physical activity on maternal and birth outcomes, more research is needed (1). Therefore, reliable (the instrument's consistency of measurement) and valid (the extent to which the instrument is measuring what it is designed to measure) data collection techniques are required.

Questionnaires and recall methods are the most commonly used data collection techniques for examining physical activity during pregnancy and postpartum as they are simple and cost effective methods for collecting data on large samples (8–11). However, questionnaires may pose limitations such as recall biases, inaccurate reporting, and challenges with question interpretation (12). Physical activity devices, such as the SenseWear armband (SenseWear), Omron pedometer (Omron), and ActiGraph accelerometer (ActiGraph), are not influenced by

recall bias or self-report error and are relatively non-invasive to participants. However, the wear location of the device may affect the accuracy of the assessment (13–15). For example, a device worn at the hip may not accurately measure stationary activities, such as cycling or weight lifting, and it has been suggested that the tilt angle could affect the accuracy of devices. This is important to consider in pregnant women as it is likely that the tilt angle of devices worn at the hip will increase as pregnancy progresses. Whether physical activity is measured via self-report or devices, it is important that these methods are reliable and valid (16). In addition, high correlations among measurement modalities would facilitate comparison of results across studies.

There are currently no published studies that have examined the reliability of physical activity devices when worn during pregnancy, including the SenseWear, Omron, and ActiGraph. This is an important omission, as poor reliability significantly affects a device's potential validity. In non-pregnant individuals, low reliability results have been described when a device was worn while subjects walked at slow speeds (17). This finding has relevance to pregnant women who have been shown to lower physical activity intensity as gestation progresses and postpartum women whose lifestyles, including physical activity, might change significantly (18).

The validity of the SenseWear, Omron, and ActiGraph, has been examined in pregnant women, but these studies have only assessed their validity at one or two time points during pregnancy or utilized a second physical activity measurement device as the criterion measure, rather than using a gold standard such as indirect calorimetry (19–22). It is important to consider how the reliability and validity of these devices can be affected by anatomical and physiological changes that occur in women during pregnancy. For example, Crouter et al. (23) determined that

a pedometer's validity was lower in an obese population due to the pedometer tilt angle at the hip. This finding has implications for pregnant women as any physical activity device's tilt angle will likely change throughout pregnancy when worn at the hip. Devices' validity has also been found to be affected by slower, compared to faster, walking speeds and during activities of daily living (24–26). This is important to note as stated previously, women's walking intensity typically decreases throughout pregnancy, and a majority of pregnant and postpartum women's physical activity is composed of activities of daily living, rather than leisure time physical activity (18, 27).

In addition, it would be ideal if self-report and device based physical activity tools had high correlations with pregnant and postpartum women's physical activity levels in free-living environments. This will allow for comparison of results across large, epidemiological studies utilizing self-report questionnaires and studies measuring physical activity with devices. The Pregnancy Physical Activity Questionnaire (PPAQ) is a popular questionnaire in pregnancy research that has been shown to have low to moderate correlations for assessing a week of sedentary, light, moderate, and vigorous physical activity with the ActiGraph (28, 29). In comparison to a different accelerometer (Actical), the PPAQ significantly overestimated number of minutes per week women participated in physical activity of various intensities (30). These studies were conducted at only one time point during pregnancy, and no published studies have evaluated the level of agreement between the PPAQ and ActiGraph at multiple time points throughout gestation and postpartum. Also, perceptions of physical activity intensity may change throughout gestation and after pregnancy, which may affect the women's answers to the questionnaires. Women may view the same intensity activity as more difficult later in gestation, potentially affecting their responses to the questionnaires (31).

In many studies examining the effects of physical activity during pregnancy, researchers require their participants to recall their physical activity for time periods of weeks, months, or years after the physical activity occurred using instruments such as the PPAQ. Although the historical recall reliability or validity of this questionnaire has not been examined in published studies, it has been found to have good short term reliability (one or two weeks), and the Modifiable Activity Questionnaire (MAQ) has moderate to strong correlations with a physical activity diary completed by women six years previously at three times points throughout pregnancy and postpartum (29, 32, 33). However, other self-report methods have shown that women's memory of their physical activity declines over time, potentially affecting the accuracy of the questionnaires (34). It is important that the long-term validity of the PPAQ is assessed if it is to be used as a recall tool in future studies.

Limited current research indicates that more information on the reliability and validity of various device based and self-report physical activity methods is needed when worn at multiple time points during pregnancy and postpartum. Researchers have analyzed similar data in many ways, but the units of analysis we have used are shown in Table 1 (Appendix). These gaps must be addressed for more specific conclusions on the effects of physical activity completed during pregnancy on maternal and fetal health. Therefore, the purposes of this dissertation are to determine the 1) reliability and validity of three popular physical activity monitors (SenseWear, Omron, and ActiGraph), 2) correlations between the ActiGraph worn at the hip, the ankle, and the PPAQ in a free-living environment, and 3) historical recall validity of the PPAQ; all at three times points during pregnancy and postpartum (21 and 32 weeks gestation and 12 weeks postpartum).

The organization of this dissertation is as follows: chapter 2 is a literature review focusing on the various devices and questionnaires used in physical activity during pregnancy research. Chapter 3 is a manuscript published in *Medicine in Science and Sports and Exercise* on the reliably and validity of the SenseWear, Omron, and ActiGraph when worn during pregnancy and postpartum. Chapter 4 is an unpublished manuscript on the correlations between ActiGraphs and the PPAQ. Chapter 5 is an unpublished manuscript on the long-term recall validity of the PPAQ. Finally, chapter 6 is focused on the overall conclusions of this dissertation.

### **Research Aims**

*Specific Aim 1:* To determine the reliability and validity of three popular physical activity monitors: SenseWear armband (SenseWear) placed on the left triceps, Omron pedometer (Omron) placed on the right hip, and ActiGraph accelerometer (ActiGraph) placed on the right hip and ankle, when worn twice at 21 and 32 weeks gestation and 12 weeks postpartum, one week apart in a laboratory environment.

H 1.1. The reliability of these devices will be moderate to strong (r = 0.60 - 1.0) at all time points (21 weeks, 32 weeks, 12 weeks postpartum) during pregnancy and postpartum.

H 1.2. The validity of the devices placed at the hip (Omron and ActiGraph) will be lower when worn during the third trimester (32 weeks), compared to second trimester (21 weeks) and 12 weeks postpartum values. The validity of all devices will be lower during the third trimester (32 weeks), compared to second trimester (21 weeks) and 12 weeks postpartum values.

*Specific Aim 2:* To determine the associations between the Pregnancy Physical Activity Questionnaire (PPAQ) and the ActiGraph placed on the right hip and ankle for measuring the percent time spent in light, moderate, and vigorous physical activity when worn in free-living conditions for one week at 21 and 32 weeks gestation and 12 weeks postpartum.

H 2.1. There will be low correlations (r < 0.30) between the PPAQ and ActiGraph at all test time points (21 weeks, 32 weeks, 12 weeks postpartum) for low, moderate, and vigorous physical activity.

*Specific Aim 3:* To test the historical (2 months -7 years) recall validity of the PPAQ when completed by women between two months and seven years after giving birth.

H 3.1. The historical recall validity of the PPAQ will be low (r < 0.30) for all time points (21 weeks and 32 weeks gestation, 12 weeks postpartum).

H 3.2. The recall validity of the PPAQs completed by women who gave birth less than five years ago will be higher compared to the PPAQs completed by women who gave birth five or more years ago.

APPENDIX

Device	Unit of Analysis	Calculated Using	Type of Variable	Strengths	Limitations
Oxycon Mobile Indirect Calorimeter	ml·kg·min <sup>1</sup>	-	Continuous	Gold standard	Laboratory based
ActiGraph Accelerometer	1. counts per second $^1$	Freedson (2011) cut points	Continuous		
	<ol> <li>minutes per day spent in sedentary, light, moderate, and vigorous PA<sup>2</sup></li> </ol>	Freedson (2011) cut points	Continuous	Records intensity, frequency, and time of PA	Complicated output, no cut points specifically for pregnant women
	<ol> <li>percent time per day spent in sedentary, light, moderate, and vigorous PA<sup>2</sup></li> </ol>	-	Continuous		
Omron Pedometer	steps per minute <sup>1</sup>	-	Continuous	Simple output	Does not record intensity or time of PA, potentially influenced by tilt angle and slower walking speeds
SenseWear Armband	1. kcals per minute <sup>1</sup>	-	Continuous	Combines accelerometry with	Can be uncomfortable, no
	2. steps per minute <sup>1</sup>	-	Continuous	biological variables (skin temperature)	longer available for purchase
Pregnancy Physical Activity Questionnaire	<ol> <li>percent time per day spent in sedentary, light, moderate, and vigorous PA<sup>2,3</sup></li> </ol>	Metabolic equivalents from Compendium of PA	Continuous	Easy to use, cost effective, developed	Potential for recall bias,
	<ol> <li>MET hours per week spent in sedentary, light, moderate, and vigorous PA<sup>3</sup></li> </ol>	Metabolic equivalents from Compendium of PA	Continuous	in pregnant women	with question interpretation

# Table 1.1. Devices and units of analysis used in the studies included in this dissertation

Note. 1: Unit of analysis for aim 1, 2: Unit of analysis for aim 2, 3: Unit of analysis for aim 3.

REFERENCES

# REFERENCES

- 1. ACOG. ACOG Committee Opinion No. 650: Physical Activity and Exercise During Pregnancy and the Postpartum Period. *Obstet Gynecol*. 2015;126(6):e135-142.
- 2. United States Department of Health and Human Services. 2008 Physical Activity Guidelines for Americans. 2008;
- 3. Aune D, Saugstad OD, Henriksen T, Tonstad S. Physical activity and the risk of preeclampsia: a systematic review and meta-analysis. *Epidemiol Camb Mass*. 2014;25(3):331–43.
- 4. Aune D, Sen A, Henriksen T, Saugstad OD, Tonstad S. Physical activity and the risk of gestational diabetes mellitus: a systematic review and dose–response meta-analysis of epidemiological studies. *Eur J Epidemiol*. 2016;31(10):967.
- 5. da Silva SG, Ricardo LI, Evenson KR, Hallal PC. Leisure-Time Physical Activity in Pregnancy and Maternal-Child Health: A Systematic Review and Meta-Analysis of Randomized Controlled Trials and Cohort Studies [Internet]. *Sports Med Auckl NZ*. 2016; doi:10.1007/s40279-016-0565-2.
- 6. Domenjoz I, Kayser B, Boulvain M. Effect of physical activity during pregnancy on mode of delivery. *Am J Obstet Gynecol*. 2014;211(4):401.e1-11.
- 7. Evenson KR, Wen F. National trends in self-reported physical activity and sedentary behaviors among pregnant women: NHANES 1999–2006. *Prev Med.* 2010;50(3):123–8.
- 8. Juhl M, Olsen J, Andersen PK, Nøhr EA, Andersen A-MN. Physical exercise during pregnancy and fetal growth measures: a study within the Danish National Birth Cohort. *Am J Obstet Gynecol*. 2010;202(1):63.e1-63.e8.
- 9. Oken E, Ning Y, Rifas-Shiman SL, Radesky JS, Rich-Edwards JW, Gillman MW. Associations of physical activity and inactivity before and during pregnancy with glucose tolerance. *Obstet Gynecol*. 2006;108(5):1200–7.
- 10. Owe KM, Nystad W, Skjaerven R, Stigum H, Bø K. Exercise during pregnancy and the gestational age distribution: a cohort study. *Med Sci Sports Exerc*. 2012;44(6):1067–74.
- 11. Wolf HT, Owe KM, Juhl M, Hegaard HK. Leisure time physical activity and the risk of pre-eclampsia: a systematic review. *Matern Child Health J*. 2014;18(4):899–910.
- 12. Klesges RC, Eck LH, Mellon MW, Fulliton W, Somes GW, Hanson CL. The accuracy of self-reports of physical activity: *Med Sci Sports Exerc*. 1990;22(5):690–7.

- 13. Ozemek C, Kirschner MM, Wilkerson BS, Byun W, Kaminsky LA. Intermonitor reliability of the GT3X+ accelerometer at hip, wrist and ankle sites during activities of daily living. *Physiol Meas.* 2014;35(2):129–38.
- 14. Swartz AM, Strath SJ, Bassett DR, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc*. 2000;32(9 Suppl):S450-456.
- 15. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J, Intille S. Estimating Activity and Sedentary Behavior from an Accelerometer on the Hip or Wrist: *Med Sci Sports Exerc*. 2013;45(5):964–75.
- 16. Hancock GR, Mueller RO, Stapleton LM. *The Reviewer's Guide to Quantitative Methods in the Social Sciences*. Routledge; 2010. 449 p.
- 17. De Cocker KA, De Meyer J, De Bourdeaudhuij IM, Cardon GM. Non-traditional wearing positions of pedometers: validity and reliability of the Omron HJ-203-ED pedometer under controlled and free-living conditions. *J Sci Med Sport*. 2012;15(5):418–24.
- 18. Borodulin KM, Evenson KR, Wen F, Herring AH, Benson AM. Physical activity patterns during pregnancy. *Med Sci Sports Exerc*. 2008;40(11):1901–8.
- 19. Connolly CP, Coe DP, Kendrick JM, Bassett DR, Thompson DL. Accuracy of physical activity monitors in pregnant women. *Med Sci Sports Exerc*. 2011;43(6):1100–5.
- 20. Harrison CL, Thompson RG, Teede HJ, Lombard CB. Measuring physical activity during pregnancy. *Int J Behav Nutr Phys Act*. 2011;8:19.
- 21. Kinnunen TI, Tennant PWG, McParlin C, Poston L, Robson SC, Bell R. Agreement between pedometer and accelerometer in measuring physical activity in overweight and obese pregnant women. *BMC Public Health*. 2011;11:501.
- 22. Smith KM, Lanningham-Foster LM, Welk GJ, Campbell CG. Validity of the SenseWear® Armband to predict energy expenditure in pregnant women. *Med Sci Sports Exerc*. 2012;44(10):2001–8.
- 23. Crouter SE, Schneider PL, Bassett DR. Spring-levered versus piezo-electric pedometer accuracy in overweight and obese adults. *Med Sci Sports Exerc*. 2005;37(10):1673–9.
- 24. Brazeau A-S, Karelis AD, Mignault D, Lacroix M-J, Prud'homme D, Rabasa-Lhoret R. Test–retest reliability of a portable monitor to assess energy expenditure. *Appl Physiol Nutr Metab.* 2011;36(3):339–43.
- 25. Machač S, Procházka M, Radvanský J, Slabý K. Validation of physical activity monitors in individuals with diabetes: energy expenditure estimation by the multisensor SenseWear Armband Pro3 and the step counter Omron HJ-720 against indirect calorimetry during walking. *Diabetes Technol Ther*. 2013;15(5):413–8.

- 26. Van Remoortel H, Giavedoni S, Raste Y, et al. Validity of activity monitors in health and chronic disease: a systematic review. *Int J Behav Nutr Phys Act*. 2012;9:84.
- 27. Ainsworth BE. Issues in the Assessment of Physical Activity in Women. *Res Q Exerc Sport*. 2000;71(sup2):37–42.
- 28. Chasan-Taber L, Schmidt MD, Roberts DE, Hosmer D, Markenson G, Freedson PS. Development and validation of a Pregnancy Physical Activity Questionnaire. *Med Sci Sports Exerc*. 2004;36(10):1750–60.
- 29. Matsuzaki M, Haruna M, Nakayama K, et al. Adapting the Pregnancy Physical Activity Questionnaire for Japanese Pregnant Women. *J Obstet Gynecol Neonatal Nurs*. 2014;43(1):107–16.
- 30. Brett KE, Wilson S, Ferraro ZM, Adamo KB. Self-report Pregnancy Physical Activity Questionnaire overestimates physical activity. *Can J Public Health Rev Can Sante Publique*. 2015;106(5):e297-302.
- 31. Marshall MR, Pivarnik JM. Perceived Exertion of Physical Activity During Pregnancy. J *Phys Act Health*. 2015;12(7):1039–43.
- 32. Stein AD, Rivera JM, Pivarnik JM. Measuring energy expenditure in habitually active and sedentary pregnant women. *Med Sci Sports Exerc*. 2003;35(8):1441–6.
- 33. Chandonnet N, Saey D, Alméras N, Marc I. French Pregnancy Physical Activity Questionnaire Compared with an Accelerometer Cut Point to Classify Physical Activity among Pregnant Obese Women [Internet]. *PLoS ONE*. 2012 [cited 2017 Apr 19];7(6) available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3372468/. doi:10.1371/journal.pone.0038818.
- Winters-Hart CS, Brach JS, Storti KL, Trauth JM, Kriska AM. Validity of a questionnaire to assess historical physical activity in older women. *Med Sci Sports Exerc*. 2004;36(12):2082–7.

#### **CHAPTER TWO: REVIEW OF THE LITERATURE**

## Introduction

In 1985, the first American College of Obstetricians and Gynecologists (ACOG) physical activity recommendations for pregnant women were published (1). They were based on a consensus opinion of a panel of obstetricians, rather than peer-reviewed research. Since then, many studies have been conducted on pregnancy and exercise, and most have determined a positive or neutral effect on labor and birth outcomes (2). Experts now recommend that in the absence of obstetric or medical contraindications, such as persistent second or third trimester bleeding, anemia, or poorly controlled type 1 diabetes, women participate in 150 minutes of moderate physical activity per week and continue this physical activity into the postpartum period. ACOG has also stated that physical *inactivity* during pregnancy is associated with complications such as maternal obesity, preeclampsia, and gestational diabetes (3). It is important that researchers utilize appropriate data collection techniques to continue to understand the relationship between physical activity and maternal and birth outcomes and to create more specific physical activity recommendations.

Various data collection methodologies have been utilized in studies examining the effects of physical activity during pregnancy including physical activity devices and self-report methods. Device measured physical activity allows for detailed information on the frequency, time, and intensity of physical activity to be collected without the issue of recall bias, but devices are often expensive and have high participant burden. National surveillance systems and large epidemiologic studies often rely upon self-report measures as they are cost effective and easy to complete, but they do have the potential for participant recall and response bias. The reliability and validity of the devices and questionnaires used in physical activity research must be carefully

examined at multiple time points during pregnancy to help determine current trends throughout gestation, and future interventions that include the most optimal frequency, intensity, time, and type of physical activities. Also, although most investigators agree that physical activity during pregnancy is not harmful to the mother and child, direct comparisons among studies is difficult as many different brands and versions of physical activity devices and types of questionnaires have been used. More data are needed on the correlations between various devices and self-report methods when worn by pregnant women in free-living environments to ensure correct conclusions about the effects of physical activity are being made. Finally, due to the practical advantages of a shortened study period and ease of data collection, many researchers utilize a retrospective study design to collect physical activity data via maternal recall. Yet, the validity of women to report long term physical activity (> two weeks) has not been examined by many self-report techniques. It is important to address these gaps in the literature to most effectively understand the relationship between physical activity and maternal and fetal outcomes.

The SenseWear armband (SenseWear; BodyMedia, Inc., Pittsburgh, PA, USA), Omron pedometer (Omron; Omron Healthcare, Inc., Bannockburn, IN, USA), ActiGraph accelerometer (ActiGraph; LLC, Fort Walton Beach, FL, USA), and Pregnancy Physical Activity Questionnaire (PPAQ) are all popular data collection techniques used in physical activity and pregnancy research. The SenseWear is a two-axis accelerometer worn on the triceps that considers physiological measures (e.g. skin temperature) to estimate energy expenditure. This device was discontinued in January 2016; however it has been used in many physical activity and pregnancy focused studies prior to this time. The Omron pedometer features two piezoelectric accelerometers capable of detecting vertical and horizontal accelerations and can be worn around the neck, in a pocket, or on a belt clip. The ActiGraph is a uni- or tri-axial

accelerometer that detects acceleration, while filtering out high-frequency movement (such as vibration from driving...), most commonly placed at the hip, but has also been studied when placed on the ankle and wrist. Finally, the PPAQ is a self-report questionnaire which requires women to report the amount of time spent completing various physical activities in the current trimester. The purpose of this literature review is to examine 1) the reliability and validity of physical activity devices, 2) the correlations between the various self-report methods and physical activity devices, and 3) the validity of self-report methods, all used in pregnancy studies. Because of their popularity, this literature review focuses on the devices and questionnaire listed above. Also, throughout the dissertation, low, moderate, and high correlations will be defined as  $\pm 0 - 0.29$ ,  $\pm 0.30 - 0.69$ , and  $\pm 0.70 - 1.0$ , respectively (4).

## **Reliability of Physical Activity Devices During Pregnancy**

No published reports of any physical activity devices' test-retest reliabilities when worn during pregnancy or postpartum were found. Although some investigators have asked women to wear various physical activity devices in free-living environments at multiple time points throughout gestation, reliability was not assessed (5–9). Most device measured studies showed that pregnant women's physical activity significantly decreases from the second through late in the third trimester, (5–12). In contrast, Stein et al. (13) determined there was no difference in physical activity from the second trimester through postpartum. Symons Downs et al. (12) found moderate reliability (r = 0.41) between mean steps taken per day at 20 and 32 weeks gestation. It is difficult to form conclusions on the reliability of devices from these studies when there were weeks to months between wear times. In addition, Symons Downs et al. (12) calculated interclass reliability using Pearson correlation, which does not account for mean differences in time points. It is likely that decreases in physical activity found in these studies were due to pregnancy-specific physical changes that occur and affect physical activity behaviors, rather than due to the devices' lack of reliability. Support for this comes from DiNallo et al. (14) who determined that women self-selected a significantly lower treadmill walking speed at 32, compared to 20 weeks gestation. Because of the lack of research performed with pregnant women, this section will focus on the reliability of the SenseWear, Omron, and ActiGraph when worn by non-pregnant participants.

### Reliability of the SenseWear

The SenseWear appears to be reliable when worn by adults during a variety of activities. Brazeau et al. (15) and DeCocker et al. (16) determined the SenseWear has high test-retest reliability in healthy and obese adults at rest (Intraclass correlation coefficients (ICC) > 0.88), but did not assess its reliability during treadmill or free-living activities. When adults completed a puzzle at rest and performed a walking circuit, the SenseWear had moderate to high reliability with Pearson correlation coefficients of 0.63 and 0.96, respectively (15). The device was also reliable when worn during cycling at 50% of participants' VO<sub>2</sub>peak in two different studies and during two identical resistance training trials (ICCs = 0.94, 0.85, 0.95, respectively) (15, 17, 18). *Reliability of the Omron* 

The reliability of the Omron may be affected by the speed at which adults are walking. During five increasing treadmill walking speeds, Omron reliability was lowest at the slowest speed and highest at the highest speed (ICC = 0.25 at 3.2 km/h, 0.96 at 6.4 km/h) (16). This finding is important for researchers examining physical activity device reliability when worn by pregnant women at different gestational stages that might result in a decrease in activity intensity (19). The intra-instrument reliability of the Omron worn in the pants pocket, as a necklace, and in a carrier bag has also been tested while participants walked up stairs and walked down stairs.

Weak correlations of 0.14 and 0.26, respectively, were calculated representing low reliability. However, during one day of free-living, high intra-instrument reliability was found between the Omron worn in the pants pocket and as a necklace with a correlation of 0.94 (16).

### *Reliability of the ActiGraph*

Both the uni-axial (7164 and GT1M) and tri-axial ActiGraph (GT3X and GT3X+) have shown high reliability when tested on vibration tables at various intensities (ICCs > 0.96 and coefficient of variation (CV) = 4.1%) (20, 21). When pairs of ActiGraphs were placed at the wrist, hip, and ankle during a variety of sedentary, housework, yard work, locomotion, and recreational activities, the tri-axial ActiGraph generally showed high reliability with 228 of the 240 ICCs above 0.80. Significant ICCs (p < 0.05) were reported at all three placements for all activities, except for axis 3 of the ankle ActiGraph while subjects were walking at a comfortable pace (ICC = 0.55), axis 3 of the hip ActiGraph while subjects carried boxes (ICC = 0.10), and axis 2 of the hip ActiGraph while subjects were climbing up and down stairs (ICC = 0.45) (22). Also, Santos-Lazano et al. (23) asked one participant to wear multiple ActiGraphs while walking and running at four treadmill speeds and found the device was very reliable in all three axes with ICCs above 0.90. Unlike the Omron, the ActiGraph did not appear to be affected by treadmill speed (23).

#### Summary

With non-pregnant participants, the SenseWear and ActiGraph have high reliability when examined during a variety of physical activities and activities of daily living (ADLs), but the Omron's ability to reliably monitor physical activity appears mixed depending on the intensity and type of activity. As stated previously, there is a gap in the literature with respect to the reliability of these popular devices when worn during pregnancy and postpartum. Additional

studies are needed, as a physical activity measurement device's validity is affected by its reliability (24). Also, because of a woman's changing anatomy, physiology, and physical activity intensity through pregnancy to postpartum, the reliability of these devices must be studied at multiple time points throughout gestation to ensure reliable and valid results are being reported.

# Validity of Physical Activity Devices During Pregnancy

The validity of physical activity devices when worn by pregnant women has been examined, but very few investigators have studied their validity when worn at multiple time points during pregnancy. Similar to the reliability of these devices, it is important to assess this throughout pregnancy as women's anatomy, physiology, and physical activity intensity changes, potentially affecting their measurement ability.

### Validity of the SenseWear

A limited number of studies have reported the SenseWear's validity compared to indirect calorimetry when worn during pregnancy and postpartum. Berntsen et al. (25) found when women in their second or third trimesters completed 90 minutes of sedentary physical activity and aerobic exercise, the ICC between the SenseWear and indirect calorimetry for the 90 minutes was 0.85. Smith et al. (26) had women complete various ADLs and treadmill walking and reported inconsistent validity coefficients between indirect calorimetry and the SenseWear (Pearson r = 0.08 through 0.99), but did not state which tasks had the higher or lower correlations. Although the authors did not report details on the correlations, they did report the difference in kilocalories per minute (kcals/min) between indirect calorimetry and the SenseWear and indirect calorimetry for folding laundry and walking at 2, 2.5, and 3 mph, 0% grade, but not for typing, sweeping, or walking at 3 mph, 3% grade (26). Brazeau et al. (15) determined that

when the SenseWear was worn by healthy adults during three treadmill speeds, validity coefficients increased with walking speed (r = 0.44 at 3.1 km/h, 0.79 at 4.3 km/h, 0.89 at 6.4 km/h). Also, a weak ICC of 0.03 was calculated when the SenseWear was worn by obese adults walking at 3 km/h (27). These studies suggest that the device may be less valid at slower, compared to faster, walking speeds. In a review by Van Remoortel et al. (28), the SenseWear underestimated total energy expenditure (TEE) at slower walking speeds and overestimates TEE at higher walking speeds (compared to indirect calorimetry) when worn by adults. Again, this is important to consider for a pregnant population who may alter physical activity intensity over time.

Van Remoortel et al.'s (28) review also pooled studies that assessed the validity of the SenseWear during non-pregnant adults' ADLs, compared to indirect calorimetry. Although two studies examining the validity of other multisensor devices were included in their analysis, 66% of the studies (n = 4/6) included the SenseWear, and results showed high validity (pooled r = 0.76) (28). Overall, the results of lab validation studies show higher correlations for walking activities (pooled r = 0.89) compared to other ADLs (pooled r = 0.76). Bernsten et al. (25) found a similar result with an ICC of 0.73 for the SenseWear's ability to measure TEE during 120 minutes of various lifestyle and sporting activities, compared to indirect calorimetry. This result is important as a majority of pregnant and postpartum women's physical activity consists of ADLs, rather than leisure time physical activity (29).

### Validity of the Omron and Other Pedometers

Omron validity has been shown to be high when worn by pregnant and non-pregnant populations when compared to manual counting as a criterion measure (30, 31). However, studies examining the validity of the pedometer for step counting at different treadmill speeds

when worn at the hip by adults have found lower correlations for slower, compared to faster, walking speeds (31–33). One study showed lower ICCs for slower treadmill walking speeds when the Omron was worn by non-pregnant adults in the pants pocket, but not if it was worn as a necklace or in a carrier bag (16). Crouter et al. (32) determined the Omron overestimated steps taken at the two slowest treadmill speeds (54 and 67 m/min), but had good agreement for the three faster speeds (80, 94, and 107 m/min). Lee et al. (2015) found similar results using Lin's concordance coefficients (Lin's), as the lowest coefficient between the Omron and manual counting was for the slowest treadmill speed; however, the correlation was still strong (Lin's = 0.90) (4, 31). High validity coefficients were reported in the only study which has examined the validity of the Omron as compared to manual counting during pregnancy and no significant speed effects were found (30). However, this study also examined the validity of the Yamax pedometer and found it was significantly less accurate at the slowest walking speed than all other speeds, and was significantly less accurate compared to the Omron for three of the four walking speeds. This is important to consider as Kinnunen et al. (34) and Harrison et al. (35) used the Yamax pedometer as the criterion measure to assess the uni-axial ActiGraph step counting ability when worn by pregnant women. Kinnuenen et al. (34) found no significant difference in step counts and Harrison et al. (35) determined there was a moderate correlation between the two devices (SCC = 0.69). More research is needed to confirm these findings in a pregnant population.

The Omron's ability to calculate Caloric energy expenditure, as compared to indirect calorimetry, has been shown to be strong for healthy adults (4, 32, 33). During five treadmill speeds between 54 and 107 m/min, ICCs between 0.79 and 0.97 were calculated (33). However, a second study determined the Omron significantly (p < 0.05) overestimated net kcals at the four

slowest speeds, and again, accuracy improved with treadmill speed (32). Unfortunately, studies have not been conducted on the Omron's validity to estimate energy expenditure as compared to indirect calorimetry throughout pregnancy.

One potential limitation of this pedometer is that it only records steps taken in walking bouts of four seconds or longer. Therefore, the Omron is likely to record fewer steps during ADLs (36). Bassett et al. (37) determined that during activities such as ironing, cooking, and washing dishes, the Yamax pedometer predicted that the non-pregnant subjects were not participating in physical activity, with indirect calorimetry as the criterion. This has not been examined with the Omron or pregnant participants, but the Omron has been shown to have high agreement with the Yamax in free-living conditions, therefore it may have similar results with ADLs as the Yamax (16, 31). Again, this is an important omission in physical activity during pregnancy research because ignoring ADLs fails to capture a large proportion of women's daily energy expenditure (29).

Other brands of pedometers have been shown to have an increased step count error with an increasing body mass index (BMI) (38–40). Crouter et al. (40) suggests the primary reasoning for the decreasing accuracy is the pedometer tilt angle. It is plausible that the tilt angle of a device worn at the hip would change with the pregnancy related increasing hip and waist circumferences. DiNallo et al. (11) examined the effects of changing tilt angle and body girth circumference from the second to third trimester on the variation of the Yamax pedometer's output. When tilt angle and body girth circumference were not controlled for, a significant difference (p < 0.05) in step counts between the second and third trimesters were found. This difference disappeared when the two variables were accounted for (11). Findings from this study indicate that anatomical changes during the process of pregnancy may affect devices'

measurement of physical activity; however these preliminary findings must be replicated with other devices.

### Validity of the ActiGraph and Other Accelerometers Using Counts per Minute

Van Hees et al. (41) used linear regression to assess agreement between a tri-axial GENEA accelerometer worn at the wrist and doubly labeled water for seven days of free-living by pregnant women. The authors found non-significant correlations between the accelerometer and doubly labeled water regardless of body weight or if the accelerometer was worn on the dominant or non-dominant wrist (41). No other published studies examining the validity of the ActiGraph or other accelerometers during pregnancy compared to indirect calorimetry were located. Two studies have compared activity counts via the uni-axial ActiGraph, RT3 accelerometer, and NewLifestyles accelerometer to indirect calorimetry while pregnant women walked at various treadmill speeds (11, 14). However, the purposes of these studies were to compare the devices' measurements between the second and third trimesters, rather than between the devices and indirect calorimetry. Therefore, conclusions cannot be made on the accelerometers' validity during pregnancy as no validity statistics were reported. Stein et al. (13) assessed the convergent validity between a Caltrac accelerometer and a heart rate monitor at three time points during pregnancy and postpartum and the unadjusted energy expenditure estimates were weak to moderately correlated across the three periods.

Crouter et al. (42) examined the relationships between a variety of uni-axial ActiGraph regression equations and energy expenditure determined by indirect calorimetry when nonpregnant subjects completed an assortment of activities. The authors concluded that the equations are only valid for the activities and populations for which they were developed and do not work well for a wide range of intensities. A pregnancy-specific calibration equation relating

ActiGraph or other accelerometer counts per min to energy expenditure has not been published, and this must be considered when interpreting accelerometer results.

The validity of the ActiGraph has been examined when non-pregnant participants completed ADLs, when compared to indirect calorimetry. During 120 minutes of household and sport activities, the ICC between the uni-axial ActiGraph and indirect calorimetry was r = 0.55 (43). These results are similar to those reported by others examining the relationship between the uni-axial ActiGraph and indirect calorimetry during ADLs (37, 44, 45). Van Remoortel et al. (28) published a systematic review that evaluated the validity of activity monitors when worn by healthy adults. Overall, results showed the uni-axial ActiGraph is moderately correlated to TEE measured via indirect calorimetry during laboratory protocols based on ADLs (r's between 0.56 and 0.65) (28). The ActiGraph appears to be moderately accurate when compared to indirect calorimetry, however no studies have been published comparing indirect calorimetry results to ActiGraphs when worn during pregnancy.

Similar to the Omron, the effects of increasing the tilt angle of the ActiGraph has been questioned. Feito et al. (46) found no effects of BMI and tilt angle for count values recorded by the uni-axial ActiGraph during treadmill walking. One study showed there was a significant difference in activity counts between the second and third trimesters when BMI and body girth circumference were controlled for, but this relationship disappeared when the variables were not accounted for. As discussed earlier, these results represent that a device's validity could be affected by the tilt angle and should be considered when worn by pregnant women.

## Validity of the ActiGraph and Other Accelerometers for Step Counting

The level of agreement for step counting between accelerometers and pedometers or manual counting has been shown to be moderate to strong when worn during pregnancy (30, 34,
35). Connolly et al. (30) assessed the uni-axial ActiGraph's accuracy when worn by pregnant women compared to manual counting during treadmill walking and found the ActiGraph significantly (p < 0.05) undercounted number of steps taken at three of the four speeds, with the least accurate results at the slowest speed. Similar results were found in Lee et al.'s (31) study, but when worn by college-aged adults, rather than pregnant women. However, the New Lifestyles accelerometer showed good agreement compared to manual counting, and no significant speed effects were found when worn by pregnant women in their second trimester when walking at various treadmill speeds (30). When worn during free-living conditions in the first or second trimester, the uni-axial ActiGraph correlated significantly with pedometers for estimating daily steps in two different studies (Spearman correlation coefficient (SCC) = 0.78 and 0.69) (34, 35). There were also significant correlations between pedometer step counts and accelerometer measures of time spent in light, moderate or vigorous physical activity, and total activity time (34). However, these correlations were at best, only moderate (SCC = 0.36 – 0.51) (34).

## Summary

Studies have found that walking speed, tilt angle, and ADLs may affect the validity of various devices. Investigators who have studied pregnant participants have examined device validity at one or two, rather than multiple, time points during pregnancy (Table 2.1: Appendix). More research must be completed when the devices are worn at multiple time points during pregnancy as women's walking speed decreases, the tilt angle of devices worn at the hip likely increases, and women accumulate most of their physical activity via ADLs. The devices researchers use to assess the effects of physical activity on maternal and birth outcomes must be

judged with these variables under consideration to improve the specificity of the physical activity recommendations for pregnant women.

# Correlations Between Self-Report Methods and Physical Activity Devices During Pregnancy

A recent review by Evenson et al. (47) summarized 12 studies published prior to 2011 conducted with pregnant participants comparing nine different self-report physical activity questionnaires to accelerometers and pedometers. A majority assessed their relationship for measuring physical activity for one week of free-living, but at only one time during pregnancy and "the results of the comparisons ranged from poor to substantial agreement," depending on the device used, location of the device, length of time the device was worn, and cut points used to define physical activity intensity. The authors concluded these variables must be considered before choosing a questionnaire for individual studies (47). Because a variety of questionnaires and devices have been used in studies examining the relationship between physical activity and pregnancy, this section of the literature review will be organized by studies comparing:1) the PPAQ and the ActiGraph, 2) the PPAQ and devices other than the ActiGraph, 3) the ActiGraph and self-report methods other than the PPAQ, 4) all self-report methods and devices other than the PPAQ and ActiGraph at one time point during pregnancy, and finally, 5) all self-report methods and devices other than the PPAQ and ActiGraph at multiple time points during pregnancy.

## Correlations Between the PPAQ and ActiGraph

Chasen-Taber et al. (48) created the PPAQ because at the time, most physical activity questionnaires had been developed in men and did not include a large portion of pregnant and postpartum women's usual activities such as household or child care activities. The PPAQ is

composed of 32 questions which provides an assessment of four domains of physical activity: sports and exercise (n = 8), household and caregiving (n = 16), transportation (n = 3), and occupation (n = 5). The authors validated their questionnaire using the uni-axial ActiGraph and since then, the PPAQ has been translated to other languages, such as French, Turkish, and Vietnamese, and again, validated with the ActiGraph. However, because most investigators report the amount of time spent in various physical activity intensities in different units for the PPAQ and ActiGraph, it is difficult to directly compare previous study results.

A second challenge when comparing results of studies using ActiGraphs is the variety of activity cut points used to calculate physical activity intensity. The cut points used in studies comparing the PPAQ and ActiGraph include the Freedson (49), Hendelman (50), Swartz (45), and Matthews (51). These differ in the threshold used to classify intensity of physical activity based on the ActiGraph's counts per minute (e.g. Freedson's, Hendelman's, Swartz's, and Matthew's cut point for light activity is 1952, 191, 574, and 760 counts per minute, respectively). Table 2.2 (Appendix) shows weak to moderate Spearman correlations were found between the questionnaire and ActiGraph when Freedson's, Hendelman's, and Swartz's cut points were used in two different studies (48, 52). Overall, higher correlations were reported when the Hendelman's and the Swartz's cut points were used, compared to the Freedson's cut points (48, 52). The PPAQ and ActiGraph showed the highest and lowest levels of agreement on the amount of moderate and sedentary activity, respectively. However, all correlations for all cut points were less than 0.50, representing a low to moderate relationship between the PPAQ and ActiGraph, in general.

When the Matthews cut points were used to categorize activity by pregnant, obese women in either their first, second, or third trimesters, moderate Spearman correlations between

the French translated PPAQ and uni-axial ActiGraph were found (Table 2.2) (53). These results showed the Matthew's cut points may be better suited for analyses of pregnant women's activity, compared to the Freedson's, Hendelman's, or Swartz's cut points. Chandonnet et al. (53) actually justified using the Matthew's cut points over the other options because these cut points were developed using data which included locomotion and ADLs that are likely performed by pregnant women.

Schmidt et al. (54) asked women in one of their three trimesters to wear the ActiGraph for seven days, then complete the PPAQ and the Kaiser Physical Activity Survey (KPAS). The authors calculated correlations between the KPAS and ActiGraph, and the KPAS and PPAQ for total activity, but unfortunately, did not compare results between the PPAQ and ActiGraph. *Correlations Between the PPAQ and Devices Other Than the ActiGraph* 

Although the PPAQ has been validated with the ActiGraph, additional studies have examined the correlations between the PPAQ and other accelerometers and pedometers during pregnancy. Brett et al. (55) compared the number of minutes in various physical activity intensities measured by the Actical (uni-axial accelerometer) and the PPAQ when women were in their second trimester. Compared to the Actical, the PPAQ significantly overestimated the number of minutes per week the women participated in sedentary, light, moderate, and moderate to vigorous physical activity, but reported similar time spent in vigorous and leisure time physical activity (55). This was confirmed as the two methods were significantly correlated (p < 0.05) for only vigorous and leisure time physical activity (SCC = 0.43 and 0.56, respectively) (55). The Actical and PPAQ also showed a negative relationship for sedentary activity (SCC = -0.28), which is similar to the relationship shown previously between the ActiGraph and PPAQ (48, 52, 55). Xiang et al. (56) compared the Chinese translated PPAQ to a Kenz uni-axial accelerometer and found low to moderate Spearman correlations between the two for total, light, moderate, and vigorous activity (SCC = 0.35, 0.33, 0.19, 0.15, respectively), but did not include sedentary activity in their analyses.

A Turkish and a Vietnamese translated version of the PPAQ was validated against pedometers worn once during pregnancy for 10 and 7 days, respectively (57, 58). Because activity intensity cannot be measured well by pedometers, only the correlation coefficient between the pedometer and PPAQ for total activity could be assessed in each study. Cirak et al. (58) calculated a much higher Pearson correlation compared to Ota et al. (57) (r = 0.70 and 0.29, respectively). The women included in Ota et al.'s (57) study took almost twice as many steps per day as the women in Cirak et al.'s (58), although the samples were very similar demographically. This result suggests that the pedometer may underestimate steps taken or the PPAQ overestimates physical activity for more sedentary and low active participants.

# Correlations Between the ActiGraph and Self-Report Methods Other Than the PPAQ

Since physical activity levels may not be stable during pregnancy, Evenson et al. (59) created a physical activity questionnaire (PIN3Q) to account for potential variation over short periods of time. The authors found low to moderate Spearman correlation coefficients between this questionnaire and the Freedson, Swartz, and Troiano (60) accelerometer cut points for moderate, vigorous, and moderate to vigorous physical activity (Table 2.3; Appendix). Similar to what was seen in other studies utilizing the PPAQ and some of these activity cut points (Table 2.2; Appendix), the lowest correlations were calculated between the PIN3Q and Freedson's cut points, with slightly higher correlations between the questionnaires and Swartz's cut points (59).

Bell et al. (61) compared the amount of physical activity performed by overweight and obese women in their first trimester using various physical activity assessments. Subjects wore

the uni-axial ActiGraph for three days, then completed the Australian Women's Activity Survey (AWAS) and the Recent Physical Activity Questionnaire (RPAQ) (61). The AWAS was specifically designed for women and children, but neither questionnaire was designed specifically for pregnant women. Compared to the ActiGraph, both the AWAS and RPAQ overestimated time in moderate to vigorous physical activity (35, 128, and 81 minutes per day, respectively), and total active time (165, 419, and 243 minutes per day, respectively). However, the AWAS overestimated (257 minutes per day) and the RPAQ underestimated (94 minutes per day) light physical activity, compared to the ActiGraph (125 minutes per day). It is possible that the ActiGraph may not be as good at measuring low intensity ADLs performed by pregnant women. In contrast, the opposite relationship was found between the International Physical Activity Questionnaire (IPAQ) and ActiGraph. The IPAQ underestimated total, light, and moderate physical activity, compared to the ActiGraph (35). In general, the AWAS and ActiGraph, and IPAQ and ActiGraph had higher Spearman correlations compared to the PPAQ and ActiGraph (Table 2.2 and 2.3). Watson et al. (10) calculated weak ICCs between the Global Physical Activity Questionnaire (GPAQ) and tri-axial ActiGraph (using Freedson's cut points) in the second or third trimester for moderate to vigorous physical activity (ICCs = 0.08 and 0.01, respectively) and sedentary behavior (ICCs = 0.05 and -0.05, respectively).

Women in one of their three trimesters wore the ActiGraph for seven days, then completed the KPAS and PPAQ (54). Much higher Spearman correlations were found between the KPAS and ActiGraph when analyzed with Freedson's, Hendelman's, and Swartz's cut points (SCCs between 0.12 to 0.66) for total activity, compared to the ActiGraph and PPAQ (48, 52, 54). The authors also compared women's responses between the PPAQ and KPAS and found moderate to high agreement with correlations between 0.37 - 0.84. This could be because the

KPAS includes more occupational and household questions than the PPAQ or because for all participants, the PPAQ was administered before the KPAS, therefore the PPAQ could have acted as a memory que and improving the women's recall for the KPAS (54). The KPAS may be better suited to be used with pregnant women, compared to the PPAQ; however more research must be done to support these results. Researchers considering using one of these self-report questionnaires may want to reflect on these relationships to help inform their decision. However, all studies have assessed the correlations between the devices and questionnaires at only one time-point during pregnancy. Different results may be found at different gestational ages. *Correlations Between all Self-Report Methods and Devices Other Than the PPAQ and ActiGraph at one Time Point During Pregnancy* 

Studies that have been designed to examine the relationship between various physical activity self-report methods and devices for assessing physical activity in free-living environments at one time point during pregnancy have found weak correlations. Haakstad et al. (62) determined the ActiReg and Physical Activity Pregnancy Questionnaire (PAPQ) had moderate correlations for high intensity, but weak correlations for low/inactive or moderate intensity (SCCs = 0.58, 0.20, and 0.15, respectively) (62). The ActiReg and Norwegian Mother and Child Cohort Study Questionnaire (MoBaQ) were not significantly correlated for total weekly MET-minutes of TEE, but were significantly correlated for physical activity energy expenditure (r = 0.29) (63). Aittsalo et al. (64) reported a Spearman correlation of 0.16 between the Omron pedometer and the Leisure Time Physical Activity Questionnaire (LTPAQ) for a week of monitoring, while Jiang et al. (65) found moderate correlations between the Omron and a Danish physical activity questionnaire (SCC = 0.45) for four days of free-living (64, 65). Overall studies show there is generally low to moderate correlations between self-report and

device based methods for categorizing and measuring physical activity in a free-living setting during pregnancy. These results are similar to those found in studies with non-pregnant participants as Skender et al.'s review (66) determined Pearson and Spearman correlations of total questionnaire scores against accelerometer measures ranged from r = 0.08 - 0.58.

Smith et al. (67) compared the ability of the SenseWear and Physical Activity Recall (PAR) to correctly categorize pregnant women as exercisers and non-exercisers, based on the standard of participating in at least 30 minutes of moderate to vigorous physical activity, three times per week. The criterion reference for being considered an exerciser was if they met this physical activity guideline via 1) an interview, 2) the PAR, and 3) seven days of heart rate monitoring. The criterion classified 25% of the women as exercisers, while the PAR and SenseWear classified 81% and 100%, respectively, of the women as exercisers (67). Investigators must consider these results when deciding which methods to use in future studies. *Correlations Between all Self-Report Methods and Devices Other Than the PPAQ and ActiGraph Throughout Pregnancy* 

Conflicting results on the correlations between various physical activity devices and self-report methods when measured at multiple time points throughout pregnancy have been reported and it is difficult to compare results as no two studies used the same self-report method or devices. Oostdam et al. (6) and Stein et al. (13) determined that self-report methods measured higher levels of physical activity in minutes per week or kcals, respectively, compared to heart rate monitors and/or accelerometers. Oostdam et al. (6) calculated weak correlations (most SCCs around 0.05 - 0.10) between the ActiTrainer accelerometer and Activity Questionnaire for Adolescents and Adults (AQuAA). Stein et al. (13) found slightly higher correlations between a heart rate monitor, and a Caltrac accelerometer and the PAR (r's range from 0.07 to 0.81) when

these methods were completed in each trimester. Also, according to Rousham et al. (7), the Actiwatch and a self-report interview were only significantly correlated for a week of wear time at 12 weeks gestation, but not at 16, 25, 34, or 38 weeks gestation (r = 0.55).

Alternatively, results from a Yamax pedometer classified more women as sedentary, low active, and somewhat active compared to the Leisure Time Exercise Questionnaire (LTEQ) (12). The only significant correlation found was between steps per day and the number of minutes in moderate physical activity determined by the LTEQ in the second trimester (ICC = 0.35). Poudevigne et al. (5) measured women's physical activity via a diary, seven-day recall interview, and a uni-axial ActiGraph. The authors found women reported higher levels of physical activity via a diary, compared to a seven day recall interview, and that a uni-axial ActiGraph correlated better with the diary than the recall interview for measuring energy expenditure (SCC = 0.50 and 0.23, respectively) (5). Direct comparison of levels of physical activity between the two selfreport methods and the accelerometer is not possible due to the differences in reported units. However, because the correlations between the ActiGraph and diary was better than between the ActiGraph and recall interview, the ActiGraph may also show higher levels of physical activity than the recall interview throughout pregnancy. Again, it is difficult to directly compare previous studies as no two used the same device, self-report method, or measurement time points. In general, no method shows significantly higher correlations at a specific pregnancy time period.

It appears that device based and self-report procedures generally do not have high correlations on the amount of physical activity women participate in throughout pregnancy, but it would be beneficial if similar trends of physical activity as gestation continues was measured by the two methods. Unfortunately, we again, have slightly conflicting results. Physical activity levels decreased throughout pregnancy according to both the device and self-report measures in

Rousham et al.'s (7) and Symons-Downs et al.'s (12) study. However, Stein et al. (13) determined that women participated in the most physical activity at 32 weeks gestation and the least physical activity at 12 weeks postpartum, according to the PAR, but there was not a change in physical activity levels according to the heart rate monitor or accelerometer used. The authors hypothesized that the low level of physical activity at postpartum was due to the amount of time needed to take care of a newborn. This was the only published study found that measured physical activity levels at postpartum with both self-report and device based methods, therefore confirmation with future studies is required.

## Summary

Ideally, self-report methods and physical activity devices would produce high correlations and classify women as participating in similar amounts of sedentary, light, moderate, and vigorous physical activity throughout pregnancy. If these data collection techniques correlate, researchers will be better able to compare results across studies and understand how much physical activity should be recommended to pregnant women. Many different devices and self-report methods have been found to have mostly low to moderate correlations when studied in pregnant populations. This agrees with Evenson et al.'s (47) review in which the authors concluded that studies comparing self-report questionnaires and physical activity devices differ depending on the device and self-report method used, length of time the devices were worn, and cut points used to define physical activity intensity. Although it would be ideal for devices and self-report methods to be able to completely agree on the amount of physical activity in which women participate, there may be a limit to how well they can correlations. Self-report methods measure physical activity behavior, while devices measure human movement. Because these are two different physical activity constructs and each method is not able to capture all forms of

activity (e.g. accelerometers cannot capture activities such as swimming), we may not be able to directly compare the results from the two methods (68, 69).

Also, most studies have examined the correlations between these methods only once during pregnancy. It is important to determine the correlations between self-report and device based methods at multiple time points throughout pregnancy to allow researchers to decide what technique(s) to use if data collection is only possible once during pregnancy.

## Validity of Recall Measures for Maternal Physical Activity

The ideal study design to assess the effects of physical activity during pregnancy would be prospectively from pre- or early pregnancy through postpartum. Unfortunately, this design is expensive and requires a substantial amount of time and effort. It is also difficult to enroll women in studies prior to becoming pregnant, due to the potential complications with conception, or in the first trimester because most women are not informing the public about their pregnancies until closer to the second trimester. Due to these considerations, many studies utilize a retrospective study design to collect information on physical activity during pregnancy by maternal recall through self-report (70, 71). However, there are concerns about the accuracy of physical activity data collected retrospectively as recall bias and/or public opinions could affect women's responses. Alternatively, because women tend to be more conscious of their actions while pregnant, they could have a better memory of their physical activity. In a review by Li et al. (72), women demonstrated high *reliability* for recall of initiation and duration of breastfeeding with an overall kappa coefficient of 0.91, a correlation coefficient of 0.86 for initiation, and correlation coefficient of 0.91 for duration. There is potential for women to have similar, good recall ability about the physical activity they participated in during pregnancy.

Most questionnaires used in pregnancy studies ask women about the average amount of time spent in various activities over the past week or trimester. Researchers would benefit if they could use these questionnaires as an accurate method to collect historical information on women's physical activity months to years after their pregnancy. Most studies that have been conducted on the *reliability* of questionnaires and other self-report methods during pregnancy are only for short periods of time such as one or two weeks. Assessing what period(s) of time a questionnaire is able to *accurately* measure physical activity retrospectively would allow researchers to determine which option is the most appropriate tool.

# Reliability of the PPAQ for Recalling Maternal Physical Activity

The one-week reliability of the PPAQ translated to Turkish, Chinese, and French to measure total activity, various types of intensities (sedentary, light, moderate, and vigorous), and types of questions (household/caregiving, occupational, sports/exercise, transportation) has been assessed in women in their first, second, or third trimester. The French translated PPAQ had high reliability with mostly strong correlations of 0.82 to 0.90 for all categories and intensities, except for the transportation (ICC = 0.59) (53). Similarly, strong ICCs between 0.92 and 0.99 were calculated by Cirak et al. (58) for the reliability of the Turkish translated PPAQ. A second study also assessed the one-two week repeatability of the Turkish translated PPAQ and again, the authors found moderate to strong correlations whether women were in their first, second, or third trimester (ICCs > 0.59) (73). The test-retest reliability of the Vietnamese translated PPAQ had similar, good results (ICCs between 0.87 and 0.94) (57). Alternatively, slightly lower ICCs of 0.28 to 0.77 were found for the one week repeatability of the Chinese translated PPAQ, with the lowest correlation for vigorous intensity activities, and the highest correlation for total activity

(light and above) (56). The authors suggest this is likely because most women did not participate in vigorous intensity activities (56).

Matsuzaki et al. (52) assessed the test-retest reliability of the Japanese translated PPAQ completed one and two weeks after the initial PPAQ. One-week reliability was high, especially for total, sedentary, light, moderate, and household/caregiving activities (ICCs between 0.78 and 0.87), but was slightly lower for occupational, sports/exercise, and transportation activity (ICCs between 0.61 – 0.66). All correlations, with the exception of occupational activity, were lower when calculated for the two-week reliability (ICCs between 0.56 and 0.84) (52). This suggests that women's memory about their physical activity during pregnancy potentially worsens as time progresses, and more information is necessary to assess this questionnaires repeatability for longer time periods.

# Reliability of Other Self-Report Physical Activity Methods for Recalling Maternal Physical Activity

The short-term repeatability of various other self-report physical activity methods has been evaluated in pregnant populations and these studies found similar, high repeatability to the PPAQ. When women completed a three day diary of number of minutes of exercise per day at 14 and 28 weeks gestation, a significant test-retest correlation of 0.61 was calculated (74). However, true repeatability was not assessed as the periods of time the dairies reflected were not the same. The IPAQ demonstrated high repeatability when completed two weeks apart, with ICCs between 0.81 and 0.84 for moderate, vigorous, and moderate to vigorous physical activity (75). Bell et al. (61) calculated a kappa statistic to examine the one-week test-retest performance of the AWAS and RPAQ when completed by women in their first trimester. AWAS reproducibility was moderate for most variables, including sedentary time and time spent in different intensities

(kappa ranging between 0.13 and 0.57), while RPAQ had higher kappas ranging between 0.42 and 0.79 (61). Aittasalo et al. (64) actually found the LTPAQ to have quite low one-week repeatability based on the changes in the mean, typical error, and coefficient of variation. The lowest coefficient of variation was 119% for evaluating the minutes per week spent in light, moderate to vigorous, and total leisure time physical activity (64). The authors suggest this questionnaire to only be used in cross-sectional or case-control designed studies with pregnant samples due to the low repeatability.

Bauer et al. (76) explored the ability of the Modifiable Activity Questionnaire (MAQ) to assess historical physical activity during women's pregnancy, six years prior. During their pregnancies, women completed a two-day physical activity diary at 20 and 32 weeks gestation and 12 weeks postpartum. Six years later, women were asked to recall their physical activity for those same time periods using the MAQ. This is the only published study of long term physical activity recall validity in pregnant women. Significant moderate to high Pearson correlations between 0.57 and 0.86 were calculated for the three time points. This shows that women did not have a difficult time recalling their physical activity while pregnant; however this has not been assessed with the PPAQ or other questionnaires. Overall, it appears women are able to reliably and accurately recall their physical activity during pregnancy using both short and long term recall periods. However, investigators have not compared women's long term self-report physical activity levels directly to their previously reported levels utilizing the same self-report method.

## Long Term Physical Activity Recall in Non-Pregnant Adults

Although no researchers have examined the repeatability of long term self-report physical activity techniques in pregnant women, this issue has been studied in non-pregnant adults. Men

and women are moderately able to recall their moderate, vigorous, and total physical activity from interview administered questionnaires 2-8 years previously, as most Spearman correlation coefficients were between 0.38 and 0.84, apart from moderate physical activity 7-8 years ago (SCC = 0.14) (77). However, Lee et al. (78) found men and women to be poor at recalling physical activity participation from up to ten years previously (SCCs = 0.38). Blair et al. (79) had adults complete the LTPAQ at baseline and between 1-10 years later and correlation results showed low to moderate associations (SCCs between 0.20 and 0.50). In this study, the Spearman correlations for women's total, light, moderate, and vigorous physical activity were similar over the ten years, with the exception of questionnaires completed 8-9 years apart, which had much lower correlations (SCCs between -0.11 and 0.35) (79). The authors suggest this could have been due to the low sample size of 39 women for that recall period. This study shows that women may have relatively good memory of their previous physical activity, however women only completed a recall questionnaire once, rather than multiple times over the ten years (79).

Winters-Hart et al. (80) assessed the validity of the Historical Physical Activity Questionnaire (HPAQ) when completed over 17 years. The authors asked women to complete a questionnaire about their past week physical activity four times between 1982 and 1999. They then completed the HPAQ about those four time periods in 1999. As expected, women's memory of their physical activity declined as time progressed. Spearman rank coefficient correlations of 0.39, 0.45, 0.57, and 0.61 were calculated between the HPAQ and the questionnaires in 1982, 1985, 1995, and 1999, respectively.

## Summary

The physical activity self-report methods that have been assessed in pregnant populations generally have high reliability when completed about the same time period one to two weeks

apart, however, the ability of women to recall their physical activity for a specific time may also decline over time. The long-term validity of most questionnaires, including the PPAQ, has not been assessed, which is an important gap in the literature as many studies ask women to recall their pregnancy physical activity months to years after gestation. Determining the long-term repeatability of these various self-report methods will help future researchers better understand the relationship between physical activity and pregnancy and which tools to use in retrospective studies.

## Conclusions

It is important to assess the effects of physical activity participation during pregnancy on maternal and birth outcomes to recommend the appropriate frequency, intensity, time, and type of physical activity to women. No research has been completed on the reliability of various physical activity devices used in physical activity and pregnancy studies or on the validity of these devices at multiple time points throughout pregnancy. It is suggested that walking speed, ADLs, and tilt angle of the device may affect the reliability and validity of the device. These factors are important to consider as the changing anatomy and physiology and walking speed, and the common activities women participate in throughout gestation may affect the reliability and validity of the devices. Low to moderate correlations between self-report methods and devices has been shown, but this relationship has not been assessed with most methods at multiple time points throughout pregnancy. Research on this topic will help determine if comparison of results utilizing these two different methods is possible and if higher correlations are found during a certain trimester. Finally, if various questionnaires are found to be appropriate for studies examining the effects of women's physical activity during their previous pregnancy, more retrospective, large sample studies can be conducted utilizing this self-report method.

However, the long-term validity of most systems has not been assessed. These gaps must be addressed for more specific conclusions on the effects of physical activity during pregnancy to be made and less expensive, and easier data collection techniques to be used. APPENDIX

Device	1 <sup>st</sup> Trimester	2 <sup>nd</sup> Trimester	3 <sup>rd</sup> Trimester	2 <sup>nd</sup> or 3 <sup>rd</sup> Trimester	2 <sup>nd</sup> and 3 <sup>rd</sup> Trimester and 12 Weeks Postpartum
SenseWear Armband		Smith et al. (2012)		Berntsen et al. (2011)	
Omron Pedometer				Connolly et al. (2011)	
ActiGraph Accelerometer	Kinnunen et al. (2011)	Harrison et al. (2011)		Connolly et al. (2011)	
Yamax Pedometer	Kinnunen et al. (2011)	Harrison et al. (2011)			
GENEA Accelerometer			Van Hees et al. (2011)		
Caltrac Accelerometer					Stein et al. (2003)
New Lifestyles Accelerometer				Connolly et al. (2011)	
Total	2	3	1	4	1

Table 2.1. Summary table of studies examining the validity of physical activity devices when worn during pregnancy and postpartum.

Cut Points (min/day)	Fı	reedson <sup>1,2</sup>	S	Swartz <sup>1,2</sup>	He	ndelman <sup>1,2</sup>	Matthew <sup>3</sup>
Total	0.08	-0.05 - 0.04*	0.32	0.20 - 0.36*	0.43	0.13 - 0.35*	0.50
Sedentary	0.12	0.001 - 0.20*	-0.17	-0.21 - 0.03*	-0.34	-0.300.12*	-0.17
Light	-0.08	-0.01 - 0.09*	0.10	0.12 - 0.34*	0.22	0.05 - 0.28*	0.46
Moderate	0.20	-0.110.03*	0.42	0.12 - 0.30*	0.49	0.15 - 0.38*	0.40
Vigorous	0.34	-0.250.18*	0.27	-0.230.18*	0.25	-0.21 - 0.09*	0.25

Table 2.2. Spearman correlations between the Pregnancy Physical Activity Questionnaire (PPAQ) and ActiGraph data analyzed with various intensity cut points.

*Note.* 1: Chasan-Taber et al. 2004; 2: Matsuzaki et al. 2014; 3: Chandonnet et al. 2012. \*Range for the first, second, and third trimester

Cut Points (min/day)	Freedson <sup>1,2,3</sup>			Swartz <sup>3</sup>	Troiano <sup>3</sup>	
Survey	AWAS	RPAQ	IPAQ		PIN3Q	
Total Activity	0.36	0.53	0.18	-	-	-
Sedentary	0.50	0.30	-	-	-	-
Light	0.29	0.41	0.07	-	-	-
Moderate	0.03	0.06	0.54	0.07	0.24	0.06
Vigorous	0.43	-0.03	-	0.38	0.44	0.40
MVPA	0.03	0.06	-	0.17	0.32	0.16

Table 2.3. Spearman correlations between various physical activity questionnaires and ActiGraph activity cut points during pregnancy.

*Note.* 1: Bell et al., 2013; 2: Harrison et al., 2011 3: Evenson et al., 2012. AWAS: Australian Women's Activity Survey, RPAQ: Recent Physical Activity Questionnaire, IPAQ: International Physical Activity Questionnaire, PIN3Q: Pregnancy Infection and Nutrition (PIN3) Study Questionnaire, MVPA: moderate to vigorous physical activity.

REFERENCES

# REFERENCES

- 1. ACOG. Technical Bulletin. Exercise During Pregnancy and the Postnatal Period. [Internet]. *Wash DC*. 1985;
- 2. Symon-Downs DS, Chasan-Taber L, Evenson KR, Leiferman J, Yeo S. Physical Activity and Pregnancy: Past and Present Evidence and Future Recommendations. *Res Q Exerc Sport*. 2012;83(4):485–502.
- 3. ACOG. ACOG Committee Opinion No. 650: Physical Activity and Exercise During Pregnancy and the Postpartum Period. *Obstet Gynecol.* 2015;126(6):e135-142.
- 4. Jackson S. *Research Methods and Statistics: A Critical Thinking Approach*. 5th ed. Belmont (CA): Wadsworth Cengage Learning; 2016. 508 p.
- 5. Poudevigne MS, O'Connor PJ. Physical activity and mood during pregnancy. *Med Sci Sports Exerc*. 2005;37(8):1374–80.
- 6. Oostdam N, van Mechelen W, van Poppel M. Validation and responsiveness of the AQuAA for measuring physical activity in overweight and obese pregnant women. *J Sci Med Sport*. 2013;16(5):412–6.
- 7. Rousham EK, Clarke PE, Gross H. Significant changes in physical activity among pregnant women in the UK as assessed by accelerometry and self-reported activity. *Eur J Clin Nutr*. 2006;60(3):393–400.
- 8. McParlin C, Robson SC, Tennant PW, et al. Objectively measured physical activity during pregnancy: a study in obese and overweight women. *BMC Pregnancy Childbirth*. 2010;10:76.
- 9. DiNallo JM, Williams NI, Downs DS, Masurier GCL. Walking for Health in Pregnancy. *Res Q Exerc Sport*. 2008;79(1):28–35.
- Watson ED, Micklesfield LK, van Poppel MNM, Norris SA, Sattler MC, Dietz P. Validity and responsiveness of the Global Physical Activity Questionnaire (GPAQ) in assessing physical activity during pregnancy [Internet]. *PLoS ONE*. 2017;12(5) available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5446115/. doi:10.1371/journal.pone.0177996.
- 11. DiNallo JM, Downs DS, Le Masurier G. Objectively assessing treadmill walking during the second and third pregnancy trimesters. *J Phys Act Health*. 2012;9(1):21–8.
- 12. Symons Downs D, LeMasurier GC, DiNallo JM. Baby steps: pedometer-determined and self-reported leisure-time exercise behaviors of pregnant women. *J Phys Act Health*. 2009;6(1):63–72.

- 13. Stein AD, Rivera JM, Pivarnik JM. Measuring energy expenditure in habitually active and sedentary pregnant women. *Med Sci Sports Exerc*. 2003;35(8):1441–6.
- 14. DiNallo JM, Williams NI, Downs DS, Masurier GCL. Walking for Health in Pregnancy. *Res Q Exerc Sport*. 2008;79(1):28–35.
- 15. Brazeau A-S, Karelis AD, Mignault D, Lacroix M-J, Prud'homme D, Rabasa-Lhoret R. Test–retest reliability of a portable monitor to assess energy expenditure. *Appl Physiol Nutr Metab*. 2011;36(3):339–43.
- 16. De Cocker KA, De Meyer J, De Bourdeaudhuij IM, Cardon GM. Non-traditional wearing positions of pedometers: validity and reliability of the Omron HJ-203-ED pedometer under controlled and free-living conditions. *J Sci Med Sport*. 2012;15(5):418–24.
- 17. Reeve MD, Pumpa KL, Ball N. Accuracy of the SenseWear Armband Mini and the BodyMedia FIT in resistance training. *J Sci Med Sport Belconnen*. 2014;17(6):630–4.
- 18. Brazeau A-S, Beaudoin N, Bélisle V, Messier V, Karelis AD, Rabasa-Lhoret R. Validation and reliability of two activity monitors for energy expenditure assessment. *J Sci Med Sport*. 2016;19(1):46–50.
- 19. Borodulin K, Evenson KR, Wen F, Herring AH, Benson A. Physical Activity Patterns during Pregnancy. *Med Sci Sports Exerc*. 2008;40(11):1901–8.
- 20. Esliger DW, Tremblay MS. Technical Reliability Assessment of Three Accelerometer Models in a Mechanical Setup: *Med Sci Sports Exerc*. 2006;38(12):2173–81.
- 21. Santos-Lozano A, Marín PJ, Torres-Luque G, Ruiz JR, Lucía A, Garatachea N. Technical variability of the GT3X accelerometer. *Med Eng Phys.* 2012;34(6):787–90.
- 22. Ozemek C, Kirschner MM, Wilkerson BS, Byun W, Kaminsky LA. Intermonitor reliability of the GT3X+ accelerometer at hip, wrist and ankle sites during activities of daily living. *Physiol Meas.* 2014;35(2):129–38.
- 23. Santos-Lozano A, Torres-Luque G, Marín PJ, Ruiz JR, Lucia A, Garatachea N. Intermonitor variability of GT3X accelerometer. *Int J Sports Med.* 2012;33(12):994–9.
- 24. Hancock GR, Mueller RO, Stapleton LM. *The Reviewer's Guide to Quantitative Methods in the Social Sciences*. Routledge; 2010. 449 p.
- 25. Berntsen S, Stafne SN, Mørkved S. Physical activity monitor for recording energy expenditure in pregnancy. *Acta Obstet Gynecol Scand*. 2011;90(8):903–7.
- 26. Smith KM, Lanningham-Foster LM, Welk GJ, Campbell CG. Validity of the SenseWear® Armband to predict energy expenditure in pregnant women. *Med Sci Sports Exerc*. 2012;44(10):2001–8.

- 27. Papazoglou D, Augello G, Tagliaferri M, et al. Evaluation of a multisensor armband in estimating energy expenditure in obese individuals. *Obes Silver Spring Md*. 2006;14(12):2217–23.
- 28. Van Remoortel H, Giavedoni S, Raste Y, et al. Validity of activity monitors in health and chronic disease: a systematic review. *Int J Behav Nutr Phys Act*. 2012;9:84.
- 29. Ainsworth BE. Issues in the Assessment of Physical Activity in Women. *Res Q Exerc Sport*. 2000;71(sup2):37–42.
- 30. Connolly CP, Coe DP, Kendrick JM, Bassett DR, Thompson DL. Accuracy of physical activity monitors in pregnant women. *Med Sci Sports Exerc*. 2011;43(6):1100–5.
- Lee JA, Williams SM, Brown DD, Laurson KR. Concurrent validation of the Actigraph gt3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. *J Sports Sci.* 2015;33(10):991– 1000.
- 32. Crouter SE, Schneider PL, Karabulut M, Bassett DR. Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. *Med Sci Sports Exerc*. 2003;35(8):1455–60.
- 33. Giannakidou DM, Kambas A, Ageloussis N, et al. The validity of two Omron pedometers during treadmill walking is speed dependent. *Eur J Appl Physiol*. 2012;112(1):49–57.
- 34. Kinnunen TI, Tennant PWG, McParlin C, Poston L, Robson SC, Bell R. Agreement between pedometer and accelerometer in measuring physical activity in overweight and obese pregnant women. *BMC Public Health*. 2011;11:501.
- 35. Harrison CL, Thompson RG, Teede HJ, Lombard CB. Measuring physical activity during pregnancy. *Int J Behav Nutr Phys Act.* 2011;8:19.
- 36. Bassett DR, John D. Use of pedometers and accelerometers in clinical populations: validity and reliability issues. *Phys Ther Rev.* 2010;15(3):135–42.
- Bassett DR, Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, King GA. Validity of four motion sensors in measuring moderate intensity physical activity. *Med Sci Sports Exerc*. 2000;32(9 Suppl):S471-480.
- 38. Melanson EL, Knoll JR, Bell ML, et al. Commercially available pedometers: considerations for accurate step counting. *Prev Med*. 2004;39(2):361–8.
- 39. Shepherd EF, Toloza E, McClung CD, Schmalzried TP. Step activity monitor: increased accuracy in quantifying ambulatory activity. *J Orthop Res Off Publ Orthop Res Soc*. 1999;17(5):703–8.
- 40. Crouter SE, Schneider PL, Bassett DR. Spring-levered versus piezo-electric pedometer accuracy in overweight and obese adults. *Med Sci Sports Exerc*. 2005;37(10):1673–9.

- van Hees VT, Renström F, Wright A, et al. Estimation of Daily Energy Expenditure in Pregnant and Non-Pregnant Women Using a Wrist-Worn Tri-Axial Accelerometer [Internet]. *PLoS ONE*. 2011;6(7) available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3146494/. doi:10.1371/journal.pone.0022922.
- 42. Crouter SE, Churilla JR, Bassett DR. Estimating energy expenditure using accelerometers. *Eur J Appl Physiol*. 2006;98(6):601–12.
- 43. Berntsen S, Hageberg R, Aandstad A, et al. Validity of physical activity monitors in adults participating in free-living activities. *Br J Sports Med.* 2010;44(9):657–64.
- 44. Welk GJ, Blair SN, Wood K, Jones S, Thompson RW. A comparative evaluation of three accelerometry-based physical activity monitors. *Med Sci Sports Exerc*. 2000;32(9 Suppl):S489-497.
- 45. Swartz AM, Strath SJ, Bassett DR, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc*. 2000;32(9 Suppl):S450-456.
- 46. Feito Y, Bassett DR, Tyo B, Thompson DL. Effects of body mass index and tilt angle on output of two wearable activity monitors. *Med Sci Sports Exerc*. 2011;43(5):861–6.
- 47. Evenson KR, Herring AH, Wen F. Self-reported and Objectively Measured Physical Activity Among a Cohort of Postpartum Women: The PIN Postpartum Study. *J Phys Act Health*. 2012;9(1):5–20.
- 48. Chasan-Taber L, Schmidt MD, Roberts DE, Hosmer D, Markenson G, Freedson PS. Development and validation of a Pregnancy Physical Activity Questionnaire. *Med Sci Sports Exerc*. 2004;36(10):1750–60.
- 49. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc*. 1998;30(5):777–781.
- 50. Hendelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc*. 2000;32(9 Suppl):S442-449.
- 51. Matthew CE. Calibration of accelerometer output for adults. *Med Sci Sports Exerc*. 2005;37(11 Suppl):S512-522.
- 52. Matsuzaki M, Haruna M, Nakayama K, et al. Adapting the Pregnancy Physical Activity Questionnaire for Japanese Pregnant Women. *J Obstet Gynecol Neonatal Nurs*. 2014;43(1):107–16.
- 53. Chandonnet N, Saey D, Alméras N, Marc I. French Pregnancy Physical Activity Questionnaire Compared with an Accelerometer Cut Point to Classify Physical Activity among Pregnant Obese Women [Internet]. *PLoS ONE*. 2012 [cited 2017 Apr 19];7(6)

available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3372468/. doi:10.1371/journal.pone.0038818.

- 54. Schmidt M, Ps F, P P, D R, B S, L C-T. Validation of the Kaiser Physical Activity Survey in pregnant women. *Med Sci Sports Exerc*. 2006;38(1):42–50.
- 55. Brett KE, Wilson S, Ferraro ZM, Adamo KB. Self-report Pregnancy Physical Activity Questionnaire overestimates physical activity. *Can J Public Health Rev Can Sante Publique*. 2015;106(5):e297-302.
- Xiang M, Konishi M, Hu H, et al. Reliability and Validity of a Chinese-Translated Version of a Pregnancy Physical Activity Questionnaire. *Matern Child Health J*. 2016;20(9):1940– 7.
- 57. Ota E, Haruna M, Yanai H, et al. Reliability and Validity of the Vietnamese Version of the Pregnancy Physical Activity Questionnaire (ppaq). *Southeast Asian J Trop Med Public Health Bangk*. 2008;39(3):562–70.
- 58. Çırak Y, Yılmaz GD, Demir YP, Dalkılınç M, Yaman S. Pregnancy physical activity questionnaire (PPAQ): reliability and validity of Turkish version. *J Phys Ther Sci.* 2015;27(12):3703–9.
- 59. Evenson KR, Wen F. National trends in self-reported physical activity and sedentary behaviors among pregnant women: NHANES 1999–2006. *Prev Med.* 2010;50(3):123–8.
- 60. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2008;40(1):181–8.
- 61. Bell R, Tennant PWG, McParlin C, et al. Measuring physical activity in pregnancy: a comparison of accelerometry and self-completion questionnaires in overweight and obese women. *Eur J Obstet Gynecol Reprod Biol*. 2013;170(1):90–5.
- 62. Haakstad LAH, Gundersen I, Bø K. Self-reporting compared to motion monitor in the measurement of physical activity during pregnancy. *Acta Obstet Gynecol Scand*. 2010;89(6):749–56.
- 63. Brantsaeter AL, Owe KM, Haugen M, Alexander J, Meltzer HM, Longnecker MP. Validation of self-reported recreational exercise in pregnant women in the Norwegian Mother and Child Cohort Study. *Scand J Med Sci Sports*. 2010;20(1):e48–55.
- 64. Aittasalo M, Pasanen M, Fogelholm M, Ojala K. Validity and Repeatability of a Short Pregnancy Leisure Time Physical Activity Questionnaire. *J Phys Act Health*. 2010;7(1):109–18.
- Jiang H, He G, Li M, et al. Reliability and Validity of a Physical Activity Scale Among Urban Pregnant Women in Eastern China. *Asia Pac J Public Health*. 2015;27(2):NP1208– 16.

- 66. Skender S, Ose J, Chang-Claude J, et al. Accelerometry and physical activity questionnaires a systematic review [Internet]. *BMC Public Health*. 2016;16 available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4910242/. doi:10.1186/s12889-016-3172-0.
- 67. Smith KM, Foster RC, Campbell CG. Accuracy of physical activity assessment during pregnancy: an observational study. *BMC Pregnancy Childbirth*. 2011;11:86.
- 68. Fulton JE, Carlson SA, Ainsworth BE, et al. Strategic Priorities for Physical Activity Surveillance in the United States. *Med Sci Sports Exerc*. 2016;48(10):2057–69.
- 69. Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act*. 2008;5:56.
- 70. Rudra CB, Williams MA, Lee I-M, Miller RS, Sorensen TK. Perceived exertion in physical activity and risk of gestational diabetes mellitus. *Epidemiol Camb Mass.* 2006;17(1):31–7.
- 71. Dempsey JC, Butler CL, Sorensen TK, et al. A case-control study of maternal recreational physical activity and risk of gestational diabetes mellitus. *Diabetes Res Clin Pract*. 2004;66(2):203–15.
- 72. Li R, Scanlon KS, Serdula MK. The validity and reliability of maternal recall of breastfeeding practice. *Nutr Rev.* 2005;63(4):103–10.
- 73. Tosun OC, Solmaz U, Ekin A, et al. The Turkish version of the pregnancy physical activity questionnaire: cross-cultural adaptation, reliability, and validity. *J Phys Ther Sci.* 2015;27(10):3215–21.
- 74. Lindseth G, Vari P. Measuring Physical Activity During Pregnancy. *West J Nurs Res.* 2005;27(6):722–34.
- 75. Sanda B, Vistad I, Haakstad LAH, et al. Reliability and concurrent validity of the International Physical Activity Questionnaire short form among pregnant women [Internet]. *BMC Sports Sci Med Rehabil.* 2017 [cited 2017 Jul 27 ];9 available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5351171/. doi:10.1186/s13102-017-0070-4.
- 76. Bauer PW, Pivarnik JM, Feltz DL, Paneth N, Womack CJ. Validation of an historical physical activity recall tool in postpartum women. *J Phys Act Health*. 2010;7(5):658–61.
- 77. Slattery ML, Jacobs DR. Assessment of ability to recall physical activity of several years ago. *Ann Epidemiol*. 1995;5(4):292–6.
- 78. Lee MM, Whittemore AS, Jung DL. Reliability of recalled physical activity, cigarette smoking, and alcohol consumption. *Ann Epidemiol*. 1992;2(5):705–14.
- 79. Blair SN, Dowda M, Pate RR, et al. Reliability of Long-term Recall of Participation in Physical Activity by Middle-aged Men and Women. *Am J Epidemiol*. 1991;133(3):266–75.

80. Winters-Hart CS, Brach JS, Storti KL, Trauth JM, Kriska AM. Validity of a questionnaire to assess historical physical activity in older women. *Med Sci Sports Exerc*. 2004;36(12):2082–7.

# CHAPTER THREE: PHYSICAL ACTIVITY DEVICE RELIABLITY AND VALIDTY DURING PREGNANCY AND POSTPARTUM

#### Abstract

Current physical activity (PA) recommendations for women experiencing a normal pregnancy reflect recent research showing numerous health benefits for mother and offspring. However, few studies have evaluated PA devices' reliability and validity during pregnancy, as anatomical and physiological changes throughout gestation could affect an instrument's accuracy. **PURPOSE:** To determine the reliability and validity of PA devices worn on the hip, ankle, and triceps during pregnancy and post-partum. **METHODS:** Thirty-three women performed six activities of daily living and one treadmill walk at approximately 21 and 32 weeks of pregnancy, and 12 weeks post-partum. There were two visits at each time-period, one week apart. Energy expenditure  $(VO_2)$  was measured by indirect calorimetry (IC; criterion measure), while PA was quantified by accelerometers and pedometers placed at the right hip and ankle and left triceps. Interclass reliability and monitor validity compared to IC in relative ( $ml \cdot kg^{-1} \cdot min^{-1}$ ) terms was calculated via Pearson correlation (PC). Both multi and single-trial intraclass reliabilities (ICC) were estimated from analysis of variance (ANOVA) to assess monitor reliability at each time-period. Standard errors of measurement (SEM) were calculated in relative terms for each time-period. **RESULTS:** The reliability of the devices was moderate/strong as 66% of the PCs were between 0.6–1.0. Multi-trial ICCs were largely in the moderate/strong range as 38% of the ICCs were between 0.6–0.79 and 50% were between 0.8–1.0. The SEMs for each device between visits ranged from 7–23% of the mean values. Comparison between IC and devices showed 40 and 46% of the validity coefficients were between 0.4–0.59 and 0.6–0.79,

respectively. **CONCLUSION:** PA devices show moderate/strong reliability and moderate validity for measuring PA during pregnancy and post-partum.

This article was published in *Medicine and Science in Sports and Exercise* on March 1, 2018. Reference:

Conway MR, Marshall MR, Schalff RA, Pfeiffer KA, Pivarnik JM. Physical activity device reliability and validity during pregnancy and postpartum. *Med Sci Sports Exerc*. 2018;50(3):617-623.

# Introduction

The United States Department of Health and Human Services (DHHS) recommends pregnant women participate in 150 minutes of moderate intensity aerobic activity per week, and the American College of Obstetricians and Gynecologists (ACOG) recommends 20 to 30 minutes of moderate exercise per day, most or all days of the week throughout pregnancy (1,2). However, studies have shown that only 13.8% of pregnant women meet the DHHS physical activity (PA) guidelines (3). Benefits from PA participation during pregnancy include a reduced risk of preeclampsia, gestational diabetes, excessive weight gain, and cesarean delivery, while no negative effects on the maternal-fetal dyad have been established (1,4,5,6,7). Although PA has been shown to benefit most women, additional research is needed to decide the optimal intensity and frequency of PA for women to participate in throughout pregnancy and the effects of this PA on maternal and birth outcomes (1). To achieve this and improve the PA guidelines for pregnant women, reliable and valid data collection techniques are required.

The validity of various devices, including the SenseWear armband (SenseWear), Omron pedometer (Omron), and ActiGraph accelerometer (ActiGraph), has been examined in pregnant women, but these studies only assessed their validity at one or two time points during pregnancy or utilized a second PA measurement device as the criterion measure, rather than using indirect calorimetry (8,9,10,11). It is important to consider how the reliability and validity of these devices can be affected by anatomical and physiological changes that occur in women during pregnancy. For example, Crouter et al. (12) determined that a pedometer's validity was lower in an obese population due to the pedometer tilt angle at the hip. However, any PA device's tilt angle would likely change throughout pregnancy when worn at the hip.

In addition to limited validity research, we found no published studies that examined the reliability of PA monitors when worn during pregnancy. This is an important omission as poor reliability significantly affects a device's potential validity. In non-pregnant individuals, low reliability and validity results have been described when devices are worn while subjects walk at slow speeds (13,14). This finding has relevance to pregnant women who have been shown to lower their PA intensity as gestation progresses (15). There is also a lack of published research examining the reliability and validity of these devices prospectively from pregnancy through post-partum. This is significant as researchers often compare post-partum PA levels to those during pregnancy.

Limited research indicates more information is needed on the reliability and validity of device measured PA assessment during multiple time points during pregnancy and post-partum. Therefore, the purpose of this study was to determine the reliability and validity of three popular device measured PA monitors: SenseWear, Omron, and ActiGraph, when worn in two consecutive weeks during the second and third trimesters, and 12 weeks post-partum.

## Methods

#### Study Population and Recruitment

The sample consisted of 33 women (aged  $29.6 \pm 3.5$  years) recruited and enrolled prior to 20 weeks gestation from obstetrical care clinics, local health clubs, and word of mouth. Inclusion

criteria included maternal age  $\geq$  18 years, non-smoker, ability to read and speak English, and pregnancy considered low-risk by the health care provider. Women provided written informed consent to complete six laboratory visits and allowed us to gather obstetric and neonatal records from the delivery hospital including gestational week of delivery, birth weight, one and five minute Apgar scores, mode of delivery, and gestational weight gain. The study protocol was approved by the University's Institutional Review Board.

## Equipment

Laboratory visits occurred in the Human Energy Research Laboratory (HERL) at approximately 21 weeks and 32 weeks of pregnancy, and 12 weeks post-partum. These time points were chosen because they are close to the middle of the second and third trimesters, and 12 weeks post-partum is needed for women to return to their prepartum physiological state. There were two visits at each time period, one week apart. Height and weight were measured on a calibrated wall stadiometer and electronic scale to the nearest cm and 0.1 kg, respectively.

The criterion instrument used for energy expenditure estimation was the Oxycon Mobile portable metabolic analyzer (CareFusion, Hoechberg, Germany), which measures breath-bybreath expired gases, allowing calculation of oxygen consumption (VO<sub>2</sub>) and carbon dioxide production (VCO<sub>2</sub>). This system has been shown to be reliable and valid at various exercise intensities and is an appropriate criterion for the testing done in this study (16). Other devices worn included the ActiGraph (ActiGraph, LLC, Fort Walton Beach, FL, USA; Model: GT3X+) and Omron (Omron Healthcare, Inc., Bannockburn, IN, USA; Model: Hj-720it), which were placed on the anterior axillary line of the right hip and secured with an elastic belt. A second ActiGraph was placed on the lateral side of the right ankle. Finally, the SenseWear (BodyMedia, Inc., Pittsburgh, PA, USA; Model: MF-SW) was worn over the left triceps.

# Laboratory Tasks

During each laboratory visit, participants completed seven different activity tasks for five minutes each, for a total of 35 minutes per session. These tasks were a mixture of activities of daily living and locomotor activities and were always completed in the same order; from the estimated lowest to highest intensity. With the exception of the treadmill walk, the women were instructed to complete the tasks at any speed or intensity they preferred. Less than 5% of participants took more than one minute rest between tasks. The tasks included:

- Laundry: Filled empty laundry basket with four towels at a table, walked to a second table three meters away, folded the towels into thirds. Repeated process.
- Dusting: Dusted counters and shelves. Subject had the choice to move objects.
- Sweeping: With a broom, subject swept confetti from one cone to another, three meters apart. Repeated process.
- Child Care: Subject picked up toys from the ground as a research assistant tossed them within cones three meters away.
- Hallway Walking: Two cones were placed 31 meters apart. The subject walked from one cone to another and back at a self-selected speed. Repeated process.
- Treadmill Walking: Walked on treadmill at three miles per hour. Option was given to hold railings.
- Aerobics: Pregnancy Aerobic Warm-Up video was completed. Examples of tasks include squats, lunges, and stretches.

## Data Collection and Reduction

Minutes 2.5 - 4 of each task were used to estimate Oxycon (ml·kg<sup>-1</sup>·min<sup>-1</sup>) and ActiGraph steady state data (counts per minute). Oxycon-measured breath-by-breath expired respiratory gases were collected and 30 second time periods were averaged in relative terms  $(ml \cdot kg^{-1} \cdot min^{-1})$ . ActiGraph raw data were collected at a sampling rate of 30 Hz and data were reintegrated to 1 – sec epochs. Omron steps were calculated manually from the difference of steps between the end and start of each five minute task, which were recorded on a data sheet during each lab visit. SenseWear kilocalorie and step data were analyzed for the entire 35 minute visit at each time point and not for individual tasks due to proprietary limitations of the SenseWear program.

#### Statistical Analysis

Reliability and validity analyses, as illustrated in the following text, were completed for all laboratory visits and devices (visits 1 and 2; visits 3 and 4; visits 5 and 6). However, for ease of understanding, reliability and validity analyses will be explained using the Omron data collected during the first and second laboratory visits. Also, reliability and validity analyses are presented for the entire 35 minute visit at each time point as we were most interested in results of the PA measurement devices worn for a variety of activities that women might be completing in a day.

#### *Reliability*

Total steps taken in visit one were compared to the total steps taken in visit two as determined by the Omron. Interclass reliability (Rxx) was calculated via Pearson correlation. Both multi and single trial intraclass reliabilities were estimated from analysis of variance (ANOVA) where Rxx = (MSs - MSe)/MSs for two trials and (MSs - MSe)/(MSs + MSe) for single trial. MSs was the mean square for subjects, MSe was the mean square for error, and Rxx was the reliability coefficient for the measures. Standard errors of measurement (SEM) were
calculated as SEM = Sx \* SQRT(1-Rxx) where Sx was the SD of the measures. SEM values were calculated in relative (%) terms for the total visits.

### Validity

Each individual's total steps taken for the seven tasks were averaged between visits one and two. Total step values were then compared to the criterion measure Oxycon VO<sub>2</sub> results (which were calculated the same way as described for the Omron results) in relative (ml·kg<sup>-</sup>  $^{1}\cdot$ min<sup>-1</sup>) terms using Pearson correlation (Rxy).

### Results

Participants were an average of 167.1 cm tall, and 72.8, 78.3, and 69.6 kg at 21 and 32 weeks gestation and 12 weeks post-partum, respectively. Due to missed appointments or device malfunction, a sample size between 23 - 27 was used for the reliability analysis for each device for visits one through four, but a sample size of 19 - 25 was used for visits five and six. For the validity analysis of each device, a sample size of 32 - 33 was used for visits one and two, 30 - 31 was used for visits three and four, and 26 - 28 was used for visits five and six. Table 3.1 (Appendix) shows the means and standard deviations of each device's recordings at each time point.

#### *Reliability*

Interclass correlation reliability results for each device are presented in Table 3.2 (Appendix). The reliability of the devices was moderate to strong as 66% (n = 12/18) of the Pearson correlations were between 0.6 – 1.0 (17). Only two of six correlation coefficients examining the relationship between visit five and six were above 0.6.

Multi-trial intraclass correlation coefficients are presented in Table 3.3 (Appendix). Values were generally higher in comparison to the interclass correlations and largely in the moderate to strong range as 38% (n = 7/18) of the ICCs were between 0.6 - 0.79 and 50% (n = 9/18) were between 0.8 - 1.0 (88% are above 0.6) (17). Devices had the lowest reliability when worn at visits five and six as 83% (n = 5/6) of the six correlations were below 0.8. As expected, single trial reliability coefficients were slightly lower for all devices and study time points compared to multi-trial values. Forty-four percent (n = 8/18) of the single trial coefficients were between 0.6 - 0.79, however, only 27% (n = 5/18) were between 0.8 - 1.0 (Table 3.4; Appendix). The hip ActiGraph and both the steps and kcals calculated by the SenseWear had the lowest single trial reliability coefficients for visits five and six, while the Omron had the lowest coefficients at visit three and four. The ankle ActiGraph had the lowest single trial reliability for both visits one and two. SEMs represent how repeated measures on the same instrument varies from the theoretical "true" value. Table 3.5 (Appendix) depicts the SEMs for each device between visits, which ranged from 7 - 23% of the mean values. For four of the six devices, SEMs were highest when worn at visits five and six.

### Validity

For the validity analysis, each device was compared to the criterion Oxycon VO<sub>2</sub> results in relative (ml·kg<sup>-1</sup>·min<sup>-1</sup>) terms. Comparison between relative VO<sub>2</sub> and devices showed 40% (n = 6/15) and 46% (n = 7/15) of the validity coefficients were between 0.4 – 0.59 and 0.6 – 0.79, respectively (Table 3.6; Appendix).

#### Discussion

The purpose of this study was to determine the reliability and validity of three popular PA measurement devices worn at multiple time points during pregnancy and post-partum using a variety of lifestyle and locomotor activities. It has been well over a decade since reliability and validity of PA measurement devices have been evaluated systematically during pregnancy, and the technology has changed significantly since then (18). It was not our purpose to predict a particular outcome measure (VO<sub>2</sub>, counts, steps, etc.) but rather, determine the relationship between different devices measuring a conglomeration of different movements that are similar to activities of daily living.

### *Reliability*

Compared to validity, fewer studies have been conducted on the reliability of the PA devices evaluated in this investigation, and we found no published reports of their reliability when worn during pregnancy. Thus, direct comparison with previous studies performed throughout gestation is not possible. In the present study, the Omron had a large range of ICCs across visit time points (r = 0.4 - 1.0). The tri-axial ActiGraph, worn at the ankle and hip during our study, produced overall high ICCs (r = 0.6 - 0.9). The ActiGraph worn at the ankle had almost as high of reliability as indirect calorimetry as determined by low SEMs (average percent of the mean = 9.8 (ActiGraph) and 7.7 (relative VO<sub>2</sub>)). The SenseWear consistently had lower reliability at visits five and six (32 weeks gestation), as shown by the single and multi-trial reliability coefficients and SEMs. However, it produced high ICCs (r = 0.8 - 0.9) at the other visit time points. This implies that the SenseWear may not be the ideal choice of device if data collection can only occur in women's third trimester.

Overall, results from the current study showed that three commonly used PA measurement devices have similar reliabilities at all study time points during pregnancy and post-partum when completing various lifestyle and locomotor activities. When the activity and intensity is the same for all women, differences in reliability is largely a function of biological variability or device measurement error. Our study participants were instructed to complete the tasks at any intensity they preferred, with the exception of walking on the treadmill where speed

was set at 4.8 kph (3 mph). Although this research focused on the reliability of the devices for a variety of activities, we calculated SEM for the activity that was standardized for everyone (treadmill walking). It would be expected that SEM would be lowest for the treadmill walking, compared to the activities overall, and less influenced by differences in the participant' effort from visit to visit. This was indeed the case, except for the ActiGraph worn at the hip, which had the lowest SEM percent mean for walking in the hallway; however there was only a 2.5% difference between the treadmill and hallway walk tasks (13.2 and 10.7%, respectively). The average SEM percent mean for the devices was 8.3, 10.8, 20.3, 33.3, 52.1, 74.7, and 96.4% for the treadmill, hallway walk, aerobic video, child care, laundry, sweeping, and dusting, respectively. Also, although it had a large range of ICCs, the Omron had the highest reliability of the devices for the treadmill as determined by its SEM (percent of the mean 4.6%). The significance of these findings is that although the devices appear to be reliable for a set intensity, even a very reliable device may be affected by a women's chosen activity intensity on a given day.

In addition, it is not surprising that overall, intraclass reliabilities were determined to be higher than interclass. At moderate activity intensity, variability of counts, steps, or kcals among women was fairly small. Thus, the limited range of responses tends to result in lower interclass reliability determined via Pearson correlations.

Single trial reliability coefficients were lower than the intraclass coefficients for all devices and time points (Table 3.4; Appendix). The ankle ActiGraph had the lowest single trial reliability at 21 weeks gestation, the Omron at 32 weeks gestation, and the hip ActiGraph and SenseWear kcals and steps at 12 weeks post-partum. This is important for researchers to consider if they are only able to collect data on participants one time, during one trimester. However, 72%

(n = 13/18) of the coefficients were above 0.6 representing that these devices have moderate to strong single trial reliability overall and data could potentially only be collected once at each time point for each participant (17).

The devices appear to be most reliable when worn during the third trimester, when the women's adnominal area is most solid. We hypothesize this to be because the devices would be less likely to move around the belt when pressed against a firmer abdominal area late in pregnancy compared to when the women have much less weight gain in the second trimester and post-partum. This would allow very little variation between the two visits, and therefore high reliability.

### Validity

The SenseWear appears to be a valid instrument when compared to indirect calorimetry in both pregnant and non-pregnant subjects (11,19). However, similar to reliability results, some studies have found the SenseWear to be less valid at slower, compared to faster, walking speeds (13,20). Due to limitations of the SenseWear program, individual tasks could not be compared between indirect calorimetry and the SenseWear in our study. Overall, the SenseWear was the least valid device when kilocalorie data were compared to relative VO<sub>2</sub> (average r = 0.39), but had comparable validity to the ActiGraph placed at the ankle and hip when step data were used (r = 0.66).

Omron validity has been shown to be high when worn by pregnant and non-pregnant populations when compared to manual counting as a criterion measure (8,21). One study showed lower correlations for slower treadmill walking speeds when the Omron was worn in the pants pocket, but not if it was worn as a necklace or in a carrier bag, when compared to manual counting (14). In the current study, the Omron was compared to indirect calorimetry, not manual

counting, but showed correlations between r = 0.40 - 0.59 for the total visit, which was similar to the other devices examined.

We were not able to locate any published studies examining the validity of the ActiGraph compared to indirect calorimetry when worn by pregnant women. However, level of agreement for step counting between the uni-axial ActiGraph and pedometers or manual counting has been shown to be moderate to strong when worn during pregnancy (8,9,10). This agrees with the current study as both the ActiGraph and Omron produced similar correlation results compared to indirect calorimetry.

Crouter et al. (22) examined the relationships between a variety of uni-axial ActiGraph regression equations and energy expenditure determined by indirect calorimetry when non-pregnant subjects completed an assortment of activities. The authors concluded that the equations are only valid for the activities and populations for which they were developed and do not work well for a wide range of intensities. A pregnancy-specific calibration equation relating ActiGraph counts per min to energy expenditure has not been published and this must be considered when interpreting ActiGraph results.

The validity of the SenseWear and ActiGraph has been examined when non-pregnant participants completed household or lifestyle activities, when compared to indirect calorimetry. During 120 minutes of household and sport activities, the SenseWear had a strong relationship to indirect calorimetry (ICC = 0.73), while the correlation between the uni-axial ActiGraph and indirect calorimetry was 0.55 (19). These results agree with validity results reported by others between the uni-axial ActiGraph and indirect calorimetry during lifestyle activities (23,24). Our study shows that the tri-axial accelerometer showed similar validity to the other PA devices examined when compared to the criterion indirect calorimetry. Overall, it is difficult to make

direct comparisons to previously published studies as a variety of data analyses have been used, and no studies have been published comparing indirect calorimetry results to ActiGraphs when worn during pregnancy.

A previous study stated that the accuracy of a device is negatively affected by increasing tilt angle (12). It is assumed that the devices worn at the hip's tilt angle would be most affected at the 32 week time period, compared to the other two data collection points. Although hip circumference was not measured in this study, the potential change in tilt angle did not appear to affect the validity of the devices worn at the hip as similar validity coefficients were calculated during pregnancy as post-partum.

Our study is unique in that multiple PA measurement devices were worn by women performing various lifestyle and locomotor activities, at multiple time points during pregnancy and post-partum. Indirect calorimetry was also utilized as the criterion measure, rather than a second PA device. No previously published studies have examined the reliability of the devices when worn during pregnancy, and only a few have studied the validity. However, there were limitations to this study. Although participants completed a variety of everyday activities, they were performed in a lab environment. While this improves study internal validity, caution must be taken when comparing to free-living results. Also, due to proprietary limitations with the SenseWear program, results of individual tasks could not be compared directly to indirect calorimetry results. Finally, activities performed by our study participants were mostly performed at a moderate intensity, so we are not sure how well the devices would perform for women completing more vigorous PA during pregnancy. However, from a public health perspective, researchers are more likely to have an interest in measuring moderate PA when

performed during pregnancy as most women are not meeting the DHHS and ACOG PA recommendations.

### Conclusion

Overall, the PA measurement devices examined in our study showed moderate to strong reliability and validity during pregnancy and post-partum. Intraclass reliability showed very similar values among the devices. The SenseWear had slightly higher interclass reliability during pregnancy, but much lower when worn post-partum. However, it is important to note that the SenseWear is no longer available for purchase. The hip and ankle ActiGraphs had consistently moderate to strong reliability at all test time points and the Omron had slightly lower reliability at 32 weeks gestation compared to 21 weeks gestation and 12 weeks post-partum. The ActiGraphs also had slightly higher validity results than the other devices. Taken together, these results support the use of any of these devices when conducting research on moderate PA during pregnancy and we believe that pregnant women's PA levels can be compared across studies that have utilized the different devices evaluated here. However, the reliability and potentially validity of the devices may be affected if higher activity intensity occurs. In the future, for researchers interested in validating PA questionnaires, devices evaluated in the present study appeared to provide similar validity in our study conditions.

APPENDIX

	RVO <sub>2</sub> (ml·kg <sup>-1</sup> ·min <sup>-1</sup> )	ActiGraph Ankle (counts·sec <sup>-1</sup> )	ActiGraph Hip (counts·sec <sup>-1</sup> )	Omron (steps·min <sup>-1</sup> )	SenseWear (Kcals·min <sup>-1</sup> )	SenseWear (steps·min <sup>-1</sup> )
Visit 1 21 Weeks	10.7 (1.8)	49.5 (12.9)	27.4 (7.4)	39.3 (8.0)	4.6 (1.3)	47.7 (17.0)
Visit 2 21 Weeks	10.3 (1.7)	51.3 (11.3)	26.5 (7.7)	37.8 (8.5)	4.4 (1.1)	42.3 (14.8)
Visit 3 32 Weeks	9.9 (2.0)	46.0 (13.0)	23.1 (6.9)	38.3 (6.6)	4.6 (1.3)	41.8 (15.5)
Visit 4 32 Weeks	10.2 (1.9)	50.7 (16.1)	22.0 (7.2)	38.9 (9.5)	4.5 (1.2)	39.5 (17.8)
Visit 5 PP	10.9 (1.5)	48.2 (16.8)	26.9 (8.6)	39.7(15.9)	4.6 (1.4)	43.1 (20.2)
Visit 6 PP	11.0 (1.7)	49.6 (16.8)	27.2 (8.0)	41.0 (7.6)	4.3 (1.3)	44.7 (14.3)

Table 3.1. Means and standard deviations (SD) at each time point, for each device.

Note. PP: Post-partum, RVO<sub>2</sub>: relative oxygen consumption

	RVO <sub>2</sub>	ActiGraph Ankle	ActiGraph Hip	Omron	SenseWear Kcals	SenseWear Steps
0.0 - 0.19						
0.2 - 0.39				3-4	5-6	
0.4 - 0.59	5-6	1-2	5-6			5-6
0.6 - 0.79	1-2 3-4	5-6	1-2 3-4	1-2 5-6		
0.8 - 1.0		3-4			1-2 3-4	1-2 3-4

Table 3.2. Interclass reliability coefficients (via Pearson correlation) for the entire 35 minute visit, at each time point, for each device.

*Note.* 1-2: Visit 1 and 2, 3-4: Visit 3 and 4, 5-6: Visit 5 and 6. RVO<sub>2</sub>: relative oxygen consumption, (n = 23 (V1-2), 24 (V3-4), 22 (V5-6), ActiGraph Ankle (n = 22 (V1-2), 23 (V3-4), 19 (V5-6), ActiGraph Hip (n = 22 (V1-2), 23 (V3-4), 18 (V5-6), Omron (n = 27 (V1-2), 26 (V3-4), 25 (V5-6), SenseWear (n = 26 (V1-2), 27 (V3-4), 21 (V5-6).

	RVO <sub>2</sub>	ActiGraph Ankle	ActiGraph Hip	Omron	SenseWear Kcals	SenseWear Steps
0.0 - 0.19						
0.2 - 0.39						
0.4 - 0.59				3-4	5-6	
0.6 - 0.79	1-2 5-6	1-2	1-2 5-6	5-6		5-6
0.8 - 1.0	3-4	3-4 5-6	3-4	1-2	1-2 3-4	1-2 3-4

Table 3.3. Multi-trial intraclass reliability coefficients (via ANOVA) for the entire 35 minute visit, at each time point, for each device.

*Note.* 1-2: Visit 1 and 2, 3-4: Visit 3 and 4, 5-6: Visit 5 and 6. RVO<sub>2</sub>: relative oxygen consumption, (n = 23 (V1-2), 24 (V3-4), 22 (V5-6), ActiGraph Ankle (n = 22 (V1-2), 23 (V3-4), 19 (V5-6), ActiGraph Hip (n = 22 (V1-2), 23 (V3-4), 18 (V5-6), Omron (n = 27 (V1-2), 26 (V3-4), 25 (V5-6), SenseWear (n = 26 (V1-2), 27 (V3-4), 21 (V5-6).

	RVO <sub>2</sub>	ActiGraph Ankle	ActiGraph Hip	Omron	SenseWear Kcals	SenseWear Steps
0.0 - 0.19						
0.2 - 0.39				3-4	5-6	
0.4 - 0.59	5-6		5-6			5-6
0.6 - 0.79	1-2	1-2	1-2	1-2		
	3-4	5-6	3-4	5-6		
0.8 - 1.0		3-4			1-2	1-2
0.0 - 1.0		5-4			3-4	3-4

Table 3.4. Single trial reliability coefficients for the entire 35 minute visit, at each time point, for each device.

*Note.* 1-2: Visit 1 and 2, 3-4: Visit 3 and 4, 5-6: Visit 5 and 6. RVO<sub>2</sub>: relative oxygen consumption, (n = 23 (V1-2), 24 (V3-4), 22 (V5-6), ActiGraph Ankel (n = 22 (V1-2), 23 (V3-4), 19 (V5-6), ActiGraph Hip (n = 22 (V1-2), 23 (V3-4), 18 (V5-6), Omron (n = 27 (V1-2), 26 (V3-4), 25 (V5-6), SenseWear (n = 26 (V1-2), 27 (V3-4), 21 (V5-6).

	RVO2 (ml·kg <sup>-1</sup> ·min <sup>-1</sup> )	ActiGraph Ankle (counts·sec <sup>-1</sup> )	ActiGraph Hip (counts·sec <sup>-1</sup> )	Omron (steps·min <sup>-1</sup> )	SenseWear (Kcals·min <sup>-1</sup> )	SenseWear (steps·min <sup>-1</sup> )
Visit 1-2 21 Weeks	0.89 (8.4)	6.1 (12.4)	3.6 (13.7)	3.1 (8.2)	0.34 (7.4)	4.0 (8.8)
Visit 3-4 32 Weeks	0.73 (7.2)	4.5 (9.7)	2.6 (12.2)	5.7 (14.9)	0.38 (8.1)	4.6 (11.3)
Visit 5-6 PP	0.84 (7.5)	3.5 (7.3)	4.4 (16.5)	5.8 (14.4)	1.05 (23.3)	9.1 (20.7)

Table 3.5. Standard error of measurement expressed in units and percent of the mean units (%) for the entire 35 min visit, at each time point, for each device.

*Note.* PP: Post-partum, RVO<sub>2</sub>: relative oxygen consumption, (n = 23 (V1-2), 24 (V3-4), 22 (V5-6), ActiGraph Ankle (n = 22 (V1-2), 23 (V3-4), 19 (V5-6), ActiGraph Hip (n = 22 (V1-2), 23 (V3-4), 18 (V5-6), Omron (n = 27 (V1-2), 26 (V3-4), 25 (V5-6), SenseWear (n = 26 (V1-2), 27 (V3-4), 21 (V5-6).

	ActiGraph Ankle	ActiGraph Hip	Omron	SenseWear Kcals	SenseWear Steps
0.0 - 0.19				5-6	
0.2 - 0.39					
0.4 - 0.59		1-2	1-2 3-4	1-2 3-4	
0.6 - 0.79	1-2 3-4	3-4			1-2 3-4
0.8 - 1.0		5-6			

Table 3.6. Validity coefficients (via Pearson correlation) for the entire visit, for each time point, for each device when compared to relative VO<sub>2</sub>.

*Note.* 1-2: Visit 1 and 2, 3-4: Visit 3 and 4, 5-6: Visit 5 and 6. ActiGraph Ankle (n = 33 (V1-2), 31 (V3-4), 28 (V5-6)), ActiGraph Hip (n = 32 (V1-2), 31 (V3-4), 28 (V5-6)), Omron (n = 33 (V1-2), 31 (V3-4), 29 (V5-6), SenseWear (n = 33 (V1-2), 30 (V3-4), 27 (V5-6).

REFERENCES

# REFERENCES

- 1. ACOG Committee Opinion No. 650: Physical activity and exercise during pregnancy and the postpartum period. *Obstet Gynecol.* 2015;126(6):e135-142.
- 2. United States Department of Health and Human Services. 2008 Physical Activity *Guidelines for Americans*. Washington D.C.: U.S. Department of Health and Human Services; 2008. 76 p. Available from: U.S. GPO, Washington.
- Evenson KR, Wen F. National trends in self-reported physical activity and sedentary behaviors among pregnant women: NHANES 1999–2006. *Prev Med.* 2010;Mar;50(3):123–8.
- Aune D, Saugstad OD, Henriksen T, Tonstad S. Physical activity and the risk of preeclampsia: a systematic review and meta-analysis. *Epidemiology*. 2014;May;25(3):331–43.
- 5. Aune D, Sen A, Henriksen T, Saugstad O, Tonstad S. Physical activity and the risk of gestational diabetes mellitus: a systematic review and dose–response meta-analysis of epidemiological studies. *Eur J Epidemiol.* 2016;Oct;31:967.
- 6. Da Silva SG, Ricardo LI, Evenson KR, Hallal PC. Leisure-time physical activity in pregnancy and maternal-child health: a systematic review and meta-analysis of randomized controlled trials and cohort studies. *Sports Med.* 2017;Feb;47(2):295-317.
- 7. Domenjoz I, Kayser B, Boulvain M. Effect of physical activity during pregnancy on mode of delivery. *Am J Obstet Gynecol.* 2014;Oct;211(4):401.e1-11.
- 8. Connolly CP, Coe DP, Kendrick JM, Bassett DR Jr, Thompson DL. Accuracy of physical activity monitors in pregnant women. *Med Sci Sports Exerc.* 2011;Jun;43(6):1100–1105.
- 9. Harrison CL, Thompson RG, Teede HJ, Lombard CB. Measuring physical activity during pregnancy. *Int J Behav Nutr Phys Act.* 2011;Mar;21(8):19.
- 10. Kinnunen TI, Tennant PW, McParlin C, Poston L, Robson SC, Bell R. Agreement between pedometer and accelerometer in measuring physical activity in overweight and obese pregnant women. *BMC Public Health.* 2011;Jun27;11:501.
- 11. Smith KM, Lanningham-Foster LM, Welk GJ, Campbell CG. Validity of the SenseWear® Armband to predict energy expenditure in pregnant women. *Med Sci Sports Exerc*. 2012;Oct;44(10):2001–8.

- Crouter SE, Schneider PL, Bassett DR Jr. Spring-levered versus piezo-electric pedometer accuracy in overweight and obese adults. *Med Sci Sports Exerc.* 2005;Oct;37(10):1673– 9.
- Brazeau AS, Beaudoin N, Belisle V, Messier V, Karelis AD, Rabasa-Lhoret R. Validation and reliability of two activity monitors for energy expenditure assessment. J Sci Med Sport. 2016;Jan;19(1):46-50.
- 14. De Cocker KA, De Meyer J, De Bourdeaudhuij IM, Cardon GM. Non-traditional wearing positions of pedometers: validity and reliability of the Omron HJ-203-ED pedometer under controlled and free-living conditions. *J Sci Med Sport*. 2012;Sep;15(5):418–24.
- 15. Borodulin KM, Evenson KR, Wen F, Herring AH, Benson AM. Physical activity patterns during pregnancy. *Med Sci Sports Exerc*. 2008;Nov;40(11):1901–8.
- 16. Carter J, Jeukendrup AE. Validity and reliability of three commercially available breathby-breath respiratory systems. *Eur J Appl Physiol.* 2002:Mar;86(5):435-441.
- 17. Jackson SL. *Research Methods and Statistics: A Critical Thinking Approach.* 5<sup>th</sup> ed. Belmont (CA): Wadsworth Cengage Learning; 2016. 508 p.
- 18. Stein AD, Rivera JM, Pivarnik JM. Measuring energy expenditure in habitually active and sedentary pregnant women. *Med Sci Sports Exerc.* 2003;Aug;35(8):1441-6.
- 19. Berntsen S, Hageberg R, Aandstad A et al. Validity of physical activity monitors in adults participating in free-living activities. *Br J Sports Med.* 2010;Jul;44(9):657-64.
- 20. Machač S, Procházka M, Radvanský J, Slabý K. Validation of physical activity monitors in individuals with diabetes: energy expenditure estimation by the multisensor SenseWear Armband Pro3 and the step counter Omron HJ-720 against indirect calorimetry during walking. *Diabetes Technol Ther.* 2013;May;15(5):413–18.
- Lee JA, Williams SM, Brown DD, Laurson KR. Concurrent validation of the Actigraph GT3X+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. *J Sports Sci.* 2015;33(10): 991–1000.
- 22. Crouter SE, Churilla JR, Bassett DR Jr. Estimating energy expenditure using accelerometers. *Eur J Appl Physiol*. 2006;Dec;98(6):601–12.
- 23. Swartz AM, Strath SJ, Bassett DR Jr, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc*. 2000;Sep;32(9 Suppl):S450-6.

24. Welk GJ, Blair SN, Wood K, Jones S, Thompson RW. A comparative evaluation of three accelerometry-based physical activity monitors. *Med Sci Sports Exerc*. 2000;Sep;32(9 Suppl):S489-97.

# CHAPTER FOUR: COMPARISON OF THE PREGNANCY PHYSICAL ACTIVITY QUESTIONNAIRE AND ACCELEROMETERS WORN DURING PREGNANCY AND POSTPARTUM

# Abstract

The Pregnancy Physical Activity Questionnaire (PPAQ) is a commonly used tool to assess pregnant women's current physical activity (PA) levels. However, few studies have evaluated the correlations between the PPAQ and PA measurement devices during free living conditions at multiple time points throughout pregnancy. **PURPOSE:** The purpose of this study was to compare the PPAQ and device based PA assessment across phases of pregnancy and postpartum. METHODS: PA behaviors of 47 women were quantified by the PPAQ and accelerometers worn at the right hip and ankle, at approximately 21 and 32 weeks of pregnancy, and 12 weeks postpartum. Women were evaluated at least eight hours per day for at least three week days and one weekend day. Percent time spent in light, moderate, and vigorous PA were compared between the PPAQ and accelerometers using a two-way repeated measures analysis of variance (ANOVA) and Spearman correlation coefficients. **RESULTS:** There was a significant interaction between the PPAQ and hip AciGraph and physical activity intensities (p < 0.05) at 21 and 32 weeks gestation and 12 weeks postpartum. At all three time points, the PPAQ underestimated percent time spent in light physical activity and overestimated percent time spent in moderate physical activity as compared to the hip ActiGraph. However, the women participated in very little vigorous physical activity according to both methods. Overall, correlations were low to moderate with values ranging from 0.01 to 0.50 for all three time points. **CONCLUSION:** Discrepancies between PA measurement modalities may be due to recall bias where the women underestimated the amount of light and overestimated their time in moderate

PA. Alternatively, accelerometers may have overestimated light and underestimated moderate

PA. Our findings are similar to results found with non-pregnant populations (Troiano et al., 2008). Researchers should consider these results when utilizing the PPAQ or accelerometers to collect PA data throughout pregnancy and postpartum as different conclusions could be made depending on the method used.

## Introduction

The American College of Obstetricians and Gynecologists (ACOG) recommends that in the absence of obstetric or medical contraindications (e.g., persistent second or third trimester bleeding, anemia, or poorly controlled type 1 diabetes), women participate in 150 minutes of moderate physical activity per week and continue this physical activity into the postpartum period. ACOG has also stated that physical *inactivity* during pregnancy is associated with complications such as maternal obesity, preeclampsia, and gestational diabetes (1). However, it is important that we continue to improve our ability to determine the optimum amount of exercise that women should perform during pregnancy and understand the relationship between physical activity and maternal and birth outcomes.

Historically, obstetricians and gynecologists were concerned that women who were physically active during pregnancy were risking complications such as miscarriages, overheating the fetus, poor fetal growth, and high fetal heart rate (2, 3). For women with normal, noncomplicated pregnancies, most of these concerns have been alleviated using information gathered from measurement techniques such as self-report questionnaires and physical activity devices like the Pregnancy Physical Activity Questionnaire (PPAQ) and ActiGraph accelerometer (ActiGraph), respectively. However, there is still much to learn about relationship between physical activity and maternal and birth outcomes, especially in women of various races and across races and/or varying socioeconomic status (e.g., infants of African American mother are 50% more likely to be born premature than are infants of white mothers) (4). If researchers better understand the information provided by various measurement techniques, future studies can allow us to learn more details on the effects of physical activity during pregnancy so women and children have the best health outcomes possible.

It is important that various data collection techniques, such as recall instruments like the PPAQ and activity monitors like the ActiGraph, generally agree on the amount of physical activity women participate in, so results can be compared from one study to another. Device measured physical activity allows for detailed information on the frequency, time, and intensity of physical activity to be collected without the issue of recall bias, but devices are often expensive and have high participant burden. National surveillance systems and large epidemiologic studies often rely upon self-report measures as they are cost effective and easy to complete, but they do have the potential for participant recall and response bias. If these selfreport methods and physical activity devices showed high correlations and classified women as participating in similar amounts of light, moderate, and vigorous physical activity throughout pregnancy in free-living environments, researchers would be better able to compare results across studies and understand how much physical activity should be recommended as the optimal amount for pregnant women. Previous research has shown mostly low to moderate correlations between the PPAQ and ActiGraph when studied in pregnant populations (5–7). However, researchers have typically examined this relationship between these methods at one timepoint during pregnancy. It has been suggested that slower walking speeds, increased tilt angle, and participating in activities of daily living (ADLs) potentially affect the measurement of the ActiGraph. This is important to consider as women's walking speed decreases, and the tilt angle of devices worn at the hip likely increases throughout gestation, and women accumulate most of their physical activity via ADLs (8-13). Also, perceptions of physical activity intensity may change throughout gestation, which may affect the women's responses to the PPAQ (14). These factors may affect the correlations between the ActiGraph and PPAQ, depending on the trimester.

The purpose of the current study was to compare the percent time spent in light,

moderate, and vigorous physical activity between the PPAQ and ActiGraph placed on the right hip (hip ActiGraph) worn in free-living conditions for one week at 21 and 32 weeks gestation and 12 weeks postpartum. We also evaluated the ActiGraph placed at the ankle but unfortunately, no intensity cut points have been established for this location. Therefore, to assess the correlations between the PPAQ and ActiGraph placed on the right ankle (ankle ActiGraph), we used total metabolic equivalents (METs) and counts per minute (CPM), respectively, for one week of free-living at 21 and 32 weeks gestation and 12 weeks postpartum.

### Methods

### Study Sample and Recruitment

The sample consisted of 47 women (aged 29.7  $\pm$  3.5 years) recruited and enrolled prior to 20 weeks gestation from obstetrical care clinics, local health clubs, and word of mouth. Inclusion criteria included maternal age  $\geq$  18 years, non-smoker, ability to read and speak English, and pregnancy considered low-risk by the health care provider. Women provided written informed consent and the study was approved by the University's Institutional Review Board.

### Equipment and Data Collection

Data collection occurred for one week at approximately 21 weeks and 32 weeks of pregnancy, and 12 weeks postpartum. These time points were chosen because they are close to the middle of the second and third trimesters, and 12 weeks postpartum is needed for women to return to their prepartum physiological state. Participants wore two ActiGraph accelerometers (ActiGraph, LLC, Fort Walton Beach, FL, USA: Model: GT3X+), which were placed on the anterior axillary line of the right hip and the lateral side of the right and secured with an elastic belt. They were asked to remove the monitors for sleeping, bathing, or swimming.

Women returned the monitors to the investigators at the conclusion of the seven days and completed the PPAQ in the laboratory based on the previous week's physical activity at each time point when the ActiGraphs were worn. Women were asked to select the option that best approximated the amount of time spent in each activity per day or week.

### Data Reduction

ActiGraph raw data were collected at a sampling rate of 30 Hz and data were reintegrated to 1 – sec epochs. A week was considered valid if the ActiGraphs were worn for at least three week days and one weekend day for eight hours a day. We defined non-wear time for the accelerometry counts as no epoch counts detected over a period of greater than 60 continuous minutes. Physical activity via the hip ActiGraph was calculated as minutes per day (using count thresholds) spent in various intensities (e.g., light, moderate, vigorous, and very vigorous) using Freedson 2011 cut points (15). These cut points were chosen because the PPAQ was originally validated using the Freedson 1998 cut points and they are also commonly used, making comparison to other studies simpler (6). Because there was very little vigorous and very vigorous physical activity calculated, we combined these intensities into one and labeled it as "vigorous." We have reported the average percent time per day spent in these various intensities, using threshold cut points. Physical activity via the ankle ActiGraph was reported as the average CPM per week because cut points have not been published for an ActiGraph worn at the ankle.

For each activity included in the PPAQ, six time range options are listed for participants to choose from (e.g., Question 4: time spent preparing meals: none, less than <sup>1</sup>/<sub>2</sub> hour per day, <sup>1</sup>/<sub>2</sub> to almost 1 hour per day, 1 to almost 2 hours per day...). The average of the time range option answered by each participant was used for calculating the amount of time spent participating in that activity (e.g., if participant chose "1 to almost 2 hours per day," 90 minutes was used as the

amount of time spent partaking in that activity). Average time spent in each activity was then multiplied by its intensity as listed by the Compendium of Physical Activities (16) to calculate energy expenditure in MET-minutes per week. For comparison to the hip ActiGraph, each activity was then classified by intensity: light (<2.9 METs), moderate (3.0 - 5.9 METs), or vigorous ( $\geq 6.0$  METs). The number of MET-minutes per week was calculated for each intensity and percent time per week spent in these intensities was reported. For comparison to the ankle ActiGraph, total MET-minutes per week for each participant, at each time point was calculated. *Statistical Analysis* 

Statistical analysis was conducted using SPSS Data Analysis version 24 (SPSS Inc, Chicago, IL). A two-way repeated measured analysis of variance (ANOVA) was conducted to examine the relationship among the hip ActiGraph and PPAQ and physical activity intensity at each gestational age. Eight of the 18 variables included in the ANOVA were not normally distributed, however, ANOVAs have been found to be robust to violations of normality (17, 18). Significance was set at an alpha level of p < 0.05. Spearman correlation coefficients were calculated between the PPAQ and hip ActiGraph for each intensity and time point and also between the PPAQ total MET-minutes per week and ankle ActiGraph CPM at each time point. **Results** 

Means and standard error of the mean (SEM) for the percent time spent in each intensity for the PPAQ and hip ActiGraph at all time points are shown in Figures 4.1, 4.2, and 4.3 (Appendix). The mean (SEM) for the total MET minutes per week measured by the PPAQ were 19913.8 (1403.9), 14850.0 (886.5), and 19817.1 (887.4) at 21 and 32 weeks gestation and 12 weeks postpartum, respectively. The average CPM per week measured by the ankle ActiGraph

were 708.9 (41.0), 626.9 (32.3), and 767.5 (55.8) at 21 and 32 weeks gestation and 12 weeks postpartum, respectively.

There was a significant interaction between the PPAQ and hip AciGraph and physical activity intensities (p < 0.05) at 21 and 32 weeks gestation and 12 weeks postpartum. At all three time points, the PPAQ underestimated percent time spent in light physical activity and overestimated percent time spent in moderate physical activity as compared to the hip ActiGraph. However, the women participated in very little vigorous physical activity according to both methods.

Spearman correlation coefficients between the PPAQ and hip ActiGraph are represented in Table 4.1 (Appendix). Overall, correlations were low to moderate with values ranging from 0.01 to 0.50 for all three time points. On average, correlations were highest between the PPAQ and hip ActiGraph at 32 weeks gestation (ranging from 0.34 to 0.40) and for light intensity (ranging from 0.20 to 0.50). Spearman correlation coefficients between the PPAQ and ankle ActiGraph were 0.39, 0.14, and 0.08 at 21 and 32 weeks gestation and 12 weeks postpartum, respectively.

## Discussion

The purpose of this study was to compare PPAQ and ActiGraph accelerometer measured physical activity at three time points throughout pregnancy and postpartum. Our data indicate the PPAQ underestimates percent time spent in light physical activity and overestimates percent time spent in moderate physical activity as compared to the ActiGraph worn at the hip at all three time points. Regardless of which instrument was used, very little vigorous physical activity was measured. There were low to moderate correlations calculated between the PPAQ and both the hip and ankle ActiGraphs at 21 and 32 weeks gestation and 12 weeks postpartum. There are

three potential reasons for these discrepancies: 1) women may have difficulties accurately recalling or quantifying the duration of their activities via questionnaire during their pregnancy, 2) the accelerometer device may be underestimating women's true physical activity, and/or 3) there may be a limit to how well physical activity questionnaires and devices are related.

The results of this study agree with a recent review by Evenson et al. (19) which summarized 12 studies published prior to 2011 conducted with pregnant participants comparing nine different self-report physical activity questionnaires to accelerometers and pedometers. Most of these studies assessed the physical activity measurements for one week of free-living, but at only one time during pregnancy. The results of the comparisons ranged from "poor to substantial agreement," depending on the device used, location of the device, length of time the device was worn, and cut points used to define physical activity intensity. Overall, in the studies examining the relationship between the PPAQ and hip ActiGraph during pregnancy, low to moderate Spearman correlations (ranging from -0.30 to 0.50) have been found, regardless of the cut points used to categorize physical activity intensity (5-7). However, these studies used uniaxial ActiGraphs and pooled women from the three trimesters, rather than examining the relationship at each trimester individually. The present study utilized tri-axial ActiGraph accelerometers and examined the relationship between the PPAQ and ActiGraphs during the second and third trimesters and 12 weeks postpartum for each participant but found a similar range of low to moderate correlations, similar to previously published studies. Thus, it appears that gestational age does not affect the relationship between the questionnaire and device and women may consistently overestimate their light and moderate physical activity, as compared to the ActiGraph, or the ActiGraph may underestimate pregnant women's physical activity as compared to the PPAQ.

These results also agree with those found by Triano et al. (20) when using the National Health and Nutrition Examination Survey (NHANES) data. The authors found that fewer than 5% of adults met the physical activity recommendations of 30 or more minutes of moderate or greater intensity activity on 5 of 7 days a week according the ActiGraph data, but 51% of adults met the 150 minutes per week of moderate or greater intensity activity according to the self-report data. Troiano et al. (20) also suggest that participants are greatly overestimating their physical activity by misclassifying sedentary or light activity as moderate, or accelerometers are not able to capture all of a person's physical activity and therefore underestimating physical activity. Our population appears to be a more active group as 100% of the women at each time point met the Department of Health and Human Services (DHHS) physical activity as measured by the hip ActiGraph. Therefore, even women who are highly active may still misclassify their activity and/or the ActiGraph is still not able to capture all of the activity in which they participated.

From an objective point of view, it would be helpful to researchers if self-report and device physical activity measurements produced high correlations on the amount of physical activity women participate in throughout pregnancy. However, this may be unlikely due to the techniques measuring two different physical activity constructs: self-report methods are measuring human behavior and devices are measuring human movement (21). Measuring physical activity is complicated and no single technique will likely be able to measure all aspects of physical activity. For example, Saris et al. (22) explains how physical activity can be expressed in many ways such as "energy, in the amount of work (watts), time period of activity, units of movement (counts), or even as a percentage based on the score of a questionnaire." Each

of these represents physical activity in a different and specific way and researchers must consider which exact construct they want to focus on when deciding how to collect their physical activity data (23). However, it may be best if researchers used multiple data collection techniques (devices and self-report) in each of their studies when possible (22–25). This will likely provide a more true approximation of participants' physical activity and understanding of how physical activity affects maternal and birth outcomes (22).

Researchers have suggested that accelerometer devices' output may be affected by changes in tilt angle. If true, this has significance for pregnant women as the tilt angle of a monitor worn at the hip is likely to increase throughout gestation. A monitor worn at the ankle may prevent these alterations. Unfortunately, cut points have not been created for an ActiGraph worn at the ankle and we therefore cannot calculate intensity using data from an ankle ActiGraph. However, we were able to compare CPM per week from the ankle ActiGraph and MET minutes per week from the PPAQ and found low to moderate correlations, similar to our results between the hip ActiGraph and PPAQ. Yet, this must be reexamined once cut points for ActiGraphs worn at the ankle are established and published.

Our study is unique in that ActiGraph accelerometers were worn for a week of free-living and the physical activity questionnaire was completed by women twice during pregnancy and once postpartum. A majority of previously published studies assessed their relationship for measuring physical activity for one week of free-living, but at only one time during pregnancy. We also asked women to wear the devices as two different locations: the hip and ankle, which has not been published previously. However, there were limitations to this study. Although participants wore an ActiGraph placed on the ankle, there are no published cut points for this wear location, therefore we were not able to quantify physical activity intensity using data from

this device. We also used the MET values provided by the Compendium of Physical Activities, which were not pregnancy specific as there are currently no MET values published for pregnant women. We would benefit from additional MET values for pregnant women participating in a variety of activities.

Finally, current federal pregnancy and exercise guidelines were developed from the belief that pregnant women would likely develop similar long term chronic disease prevention benefits to their non-pregnant counterparts. They were not evidence based with respect to pregnancy outcomes and/or complications. If it is not appropriate to compare self-report and device based physical activity data, perhaps additional recommendations for pregnant women could be developed and published. The current physical activity recommendations are based upon data collected via questionnaires, and we can therefore compare results from studies utilizing questionnaires to those recommendations. If and when pregnancy physical activity guidelines are developed for pregnancy related issues, developers should be cognizant of the fact that measurement modality will present the same shortcomings as exist with the nonpregnant population. If other recommendations were created from data collected via physical activity devices then it might be appropriate to compare results from studies using devices to those recommendations, rather than the ones based on questionnaire data (21). If this could be accomplished, future researchers might be better able to understand the effects of women meeting the physical activity recommendations on maternal and fetal health outcomes, regardless of the measurement technique used.

### Conclusion

Overall, the PPAQ and ActiGraphs worn at the hip and ankle showed low to moderate correlations during pregnancy and postpartum. There were slightly higher correlations between

the hip ActiGraph and PPAQ at 32 weeks gestation, however, the average correlations were very similar at all pregnancy time points. The ankle ActiGraph and PPAQ had the highest correlation at 21 weeks gestation, when the PPAQ reported the lowest amount of activity of the three time points. In general, it does not appear that any time point during pregnancy produces much higher correlations between the PPAQ and ActiGraph. Our findings and those of Troiano et al. (20) and Evenson et al. (19) agree that there are large discrepancies between device based and self-report physical activity measurements in both non-pregnant and pregnant populations. In general, device based methods appear to underestimate physical activity compared to self-report measurements or self-report techniques overestimate activity compared to physical activity devices. Because it appears self-report and device methods have low correlations when measuring pregnant women's physical activity, researchers must be mindful of this when selecting what instrument(s) to use for data collection, how the data will be interpreted, and what publications to compare the results to (21). We would also recommend that supplementary physical activity recommendations be published based on device measured data and/or researchers focus on creating a questionnaire that appropriately measures physical activity but also agrees with physical activity devices.

APPENDIX



Figure 4.1. Percent time (standard error of the mean) spent in light, moderate, and vigorous physical activity at 21 weeks gestation. Light: < 2.9 METs, Moderate: 3.0 - 5.9 METs, Vigorous:  $\geq 6.0$ . n = 36



Figure 4.2. Percent time (standard error of the mean) spent in light, moderate, and vigorous physical activity at 32 weeks gestation. Light: < 2.9 METs, Moderate: 3.0 - 5.9 METs, Vigorous:  $\geq 6.0$ . n = 35



Figure 4.3. Percent time (standard error of the mean) spent in light, moderate, and vigorous physical activity at 12 weeks postpartum. Light: < 2.9 METs, Moderate: 3.0 - 5.9 METs, Vigorous:  $\geq 6.0$ . n = 30
	21 Weeks Gestation	32 Weeks Gestation	12 Weeks Postpartum
Light	0.50	0.40	0.20
Moderate	0.46	0.34	0.01
Vigorous	0.05	0.35	0.39

Table 4.1. Spearman correlation coefficients between the Pregnancy Physical Activity Questionnaire (PPAQ) and ActiGraph worn at the hip at three time points during pregnancy and postpartum.

REFERENCES

# REFERENCES

- 1. ACOG. ACOG Committee Opinion No. 650: Physical Activity and Exercise During Pregnancy and the Postpartum Period. *Obstet Gynecol*. 2015;126(6):e135-142.
- 2. McMurray RG, Mottola MF, Wolfe LA, Artal R, Millar L, Pivarnik JM. Recent advances in understanding maternal and fetal responses to exercise. *Med Sci Sports Exerc*. 1993;25(12):1305–21.
- 3. Papiernik E, Kaminski M. Multifactorial study of the risk of prematurity at 32 weeks of gestation. I. A study of the frequency of 30 predictive characteristics. *J Perinat Med*. 1974;2(1):30–6.
- 4. Preterm Birth | Maternal and Infant Health | Reproductive Health | CDC2017; [cited 2018 Apr 14 ] Available from: https://www.cdc.gov/reproductivehealth/maternalinfanthealth/pretermbirth.htm.
- 5. Chandonnet N, Saey D, Alméras N, Marc I. French Pregnancy Physical Activity Questionnaire Compared with an Accelerometer Cut Point to Classify Physical Activity among Pregnant Obese Women [Internet]. *PLoS ONE*. 2012 [cited 2017 Apr 19];7(6) available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3372468/. doi:10.1371/journal.pone.0038818.
- 6. Chasan-Taber L, Schmidt MD, Roberts DE, Hosmer D, Markenson G, Freedson PS. Development and validation of a Pregnancy Physical Activity Questionnaire. *Med Sci Sports Exerc*. 2004;36(10):1750–60.
- 7. Matsuzaki M, Haruna M, Nakayama K, et al. Adapting the Pregnancy Physical Activity Questionnaire for Japanese Pregnant Women. *J Obstet Gynecol Neonatal Nurs*. 2014;43(1):107–16.
- 8. Ainsworth BE. Issues in the Assessment of Physical Activity in Women. *Res Q Exerc Sport*. 2000;71(sup2):37–42.
- 9. Borodulin K, Evenson KR, Wen F, Herring AH, Benson A. Physical Activity Patterns during Pregnancy. *Med Sci Sports Exerc*. 2008;40(11):1901–8.
- Brazeau A-S, Karelis AD, Mignault D, Lacroix M-J, Prud'homme D, Rabasa-Lhoret R. Test–retest reliability of a portable monitor to assess energy expenditure. *Appl Physiol Nutr Metab.* 2011;36(3):339–43.
- 11. Crouter SE, Schneider PL, Bassett DR. Spring-levered versus piezo-electric pedometer accuracy in overweight and obese adults. *Med Sci Sports Exerc*. 2005;37(10):1673–9.

- 12. Machač S, Procházka M, Radvanský J, Slabý K. Validation of physical activity monitors in individuals with diabetes: energy expenditure estimation by the multisensor SenseWear Armband Pro3 and the step counter Omron HJ-720 against indirect calorimetry during walking. *Diabetes Technol Ther*. 2013;15(5):413–8.
- 13. Van Remoortel H, Giavedoni S, Raste Y, et al. Validity of activity monitors in health and chronic disease: a systematic review. *Int J Behav Nutr Phys Act*. 2012;9:84.
- 14. Marshall MR, Pivarnik JM. Perceived Exertion of Physical Activity During Pregnancy. J *Phys Act Health.* 2015;12(7):1039–43.
- 15. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. *J Sci Med Sport*. 2011;14(5):411–6.
- 16. Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc*. 2011;43(8):1575–81.
- 17. Lunney GH. Using Analysis of Variance with a Dichotomous Dependent Variable: An Empirical Study. *J Educ Meas*. 1970;7(4):263–9.
- 18. Field A. Discovering Statistics Using SPSS. SAGE Publications; 2009. 857 p.
- 19. Evenson KR, Herring AH, Wen F. Self-reported and Objectively Measured Physical Activity Among a Cohort of Postpartum Women: The PIN Postpartum Study. *J Phys Act Health*. 2012;9(1):5–20.
- 20. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2008;40(1):181–8.
- 21. Fulton JE, Carlson SA, Ainsworth BE, et al. Strategic Priorities for Physical Activity Surveillance in the United States. *Med Sci Sports Exerc*. 2016;48(10):2057–69.
- 22. Saris WHM. Habitual physical activity in children: methodology and findings in health and disease. *Med Sci Sports Exerc*. 1986;18(3):253–63.
- 23. Sylvia LG, Bernstein EE, Hubbard JL, Keating L, Anderson EJ. A Practical Guide to Measuring Physical Activity. *J Acad Nutr Diet*. 2014;114(2):199–208.
- 24. Freedson PS. Field Monitoring of Physical Activity in Children. *Pediatr Exerc Sci.* 1989;1(2):8–18.
- 25. Melanson EL, Freedson PS, Blair S. Physical activity assessment: A review of methods. *Crit Rev Food Sci Nutr.* 1996;36(5):385–96.

# CHAPTER FIVE: VALIDITY OF THE PREGNANCY PHYSICAL ACTIVITY QUESTIONNAIRE FOR MATERNAL PHYSICAL ACTIVITY RECALL

## Abstract

The concern with any self-report or recall physical activity method is the accuracy of the responses provided by participants. It is important that the long-term validity of the Pregnancy Physical Activity Questionnaire (PPAQ) is assessed at multiple time points during pregnancy and postpartum if it is to be used as a recall tool in future studies. PURPOSE: The purpose of this study was to test the historical recall validity of women completing the PPAQ about their physical activity at three time points during pregnancy and postpartum and then five months to eight years after giving birth. METHODS: Between 2010 and 2018, 48 women completed the PPAQ at 21 and 32 weeks gestation and 12 weeks postpartum about their previous week's physical activity. These same women were emailed three separate PPAOs between five months and eight years after originally completing the questionnaires to recall their physical activity during those same time periods. Of these 48 women, 40 completed the follow up historical recall questionnaires. Total number of metabolic (MET) minutes per week and percent time spent in light, moderate, and vigorous activity were compared between the original and recall PPAQ values using paired sample t-tests or Wilcoxon Rank tests and Spearman correlation coefficients (SCC). The participants were then separated into two groups via a median split: those who originally completed the PPAQs  $\geq$  five years ago and  $\leq$  five years ago. The paired sample t-tests and SCCs were again conducted. RESULTS: Women tended to underestimate their total METminutes per week and percent time spent in moderate activity by 3000-4000 MET minutes per week and 6%, respectively, and overestimate the percent time in light activity by 4-6%, when comparing recall to original values. Women reported spending little time in vigorous intensity activity at both time points (2-4%). There were lower SCCs for women who were recalling their

physical activity  $\geq$  5 years postpartum compared to women who were recalling their physical activity  $\leq$  five years postpartum for most time points and intensities. **CONCLUSION:** Because of the public health impact of physical activity on long-term health outcomes, historical physical activity questionnaires would be useful, and it is important we continue to assess the long-term validity of self-report methods such as the PPAQ. It would be best for future researchers to ask women to recall their physical activity during pregnancy using the PPAQ less than five years postpartum, and it can also be assumed that they will have likely underestimated their total and moderate physical activity.

## Introduction

An ideal study design to assess the effects of physical activity during pregnancy would be prospectively from pre- or early pregnancy through postpartum. Unfortunately, this design can be expensive and requires a substantial amount of time and effort. It is also difficult to enroll women in studies prior to them becoming pregnant, due to the potential complications with conception, or early in the first trimester because most women do not inform the public about their pregnancies until closer to the second trimester. Due to these considerations, many researchers utilize a retrospective study design to collect information on physical activity during pregnancy by maternal recall through self-report (1, 2). However, there are concerns about the accuracy of physical activity data collected retrospectively as recall bias and/or public opinions could affect women's responses. Alternatively, because many women may tend to be more conscious of their actions while pregnant, their ability to recall events such as previous exercise and recreational physical activity could be enhanced. For example, in a review by Li et al. (3), women demonstrated high reliability for recall of initiation and duration of breastfeeding with an overall kappa coefficient of 0.91, a correlation coefficient of 0.86 for initiation, and correlation coefficient of 0.91 for duration. There is potential for women to have similar, good recall ability about the physical activity they participated in during pregnancy.

A popular self-report physical activity instrument is the Pregnancy Physical Activity Questionnaire (PPAQ). Although the historical recall *validity* of this questionnaire has not been examined in published studies, it has been found to have good short-term reliability (one or two weeks) with mostly moderate to strong intraclass correlation coefficients above 0.60 for total activity and minutes in various types of intensities (sedentary, light, moderate, and vigorous) for women in their first, second, or third trimesters (4–7). Matsuzaki et al. (8) assessed the test-retest

reliability of the Japanese translated PPAQ completed one and two weeks after the initial PPAQ. One-week reliability was high, especially for time spent in total, sedentary, light, moderate, and household/caregiving activities (ICCs between 0.78 and 0.87). All correlations, with the exception of occupational activity, were lower when calculated for two-week reliability (ICCs between 0.56 and 0.84) (8). This suggests that women's memory about their physical activity behaviors during pregnancy potentially worsens as time progresses, even across relatively short time spans. However, the Modifiable Activity Questionnaire (MAQ) was found to have moderate to strong validity correlations with a physical activity diary completed by women six years previously at three times points throughout pregnancy and postpartum (9).

It is important that the long-term validity of the PPAQ is assessed at multiple time points during pregnancy and postpartum if it is to be used as a recall tool in future studies. Therefore, the purpose of this study was to test the historical recall validity of women completing the PPAQ about their physical activity at 21 and 32 weeks gestation and 12 weeks postpartum, two months to eight years after giving birth. It was hypothesized that the recall validity would be low (r < 0.30) for all time points (21 weeks and 32 weeks gestation, 12 weeks postpartum).

## Methods

#### Study Sample and Recruitment

Between 2010 and 2017, 48 women were recruited and enrolled in a study examining various physical activity measurement techniques during pregnancy and postpartum (referred to as The Mama Study). Inclusion criteria included: enrolled prior to 20 weeks gestation, maternal age  $\geq$  18 years, non-smoker, ability to read and speak English, and pregnancy considered low-risk by the health care provider. One portion of The Mama Study required women to complete a PPAQ about their physical activity in the past week when they were ~21 weeks gestation, ~32

weeks gestation, and ~12 weeks postpartum. These same 48 women were contacted in 2018 (two months to eight years after their original enrollment in The Mama Study) to participate in this follow up study and 40 of these 48 women completed the recall survey (83%). Women provided informed consent approved by the University's Institutional Review Board.

### Data Collection

Women were emailed three Qualtrics survey links to three separate PPAQs to recall their physical activity during their 21<sup>st</sup> and 32<sup>nd</sup> week gestation and 12<sup>th</sup> week postpartum, when originally enrolled in The Mama Study. They were asked to select the option that best approximated the amount of time spent in each activity per day or week. Women were compensated with a \$20 gift card for the completion of all three questionnaires.

## Data Reduction

For each activity included in the PPAQ, six time range options were listed from which participants could choose (e.g., Question 4: time spent preparing meals: none, less than  $\frac{1}{2}$  hour per day,  $\frac{1}{2}$  to almost 1 hour per day, 1 to almost 2 hours per day...). The average of the time range option answered by each participant was used to for calculating the amount of time spent participating in that activity (e.g., if participant chose "1 to almost 2 hours per day," 90 minutes was used as the amount of time spent partaking in that activity). Average time spent in each activity was then multiplied by its intensity as listed by the Compendium of Physical Activities (10) to calculate energy expenditure in MET-minutes per week. Each activity was then classified by intensity: light (<2.9 METs), moderate (3.0 - 5.9 METs), or vigorous ( $\geq 6.0$  METs). Total number of MET-minutes per week was calculated for each intensity and percent time per week spent in these intensities was reported for both the original and recall questionnaires. Total MET-

minutes per week and percent time spent in light, moderate, and vigorous activity were then compared between original and recall PPAQ values.

### Statistical Analysis

Statistical analysis was conducted using SPSS Data Analysis version 24 (SPSS Inc, Chicago, IL). Paired samples t-tests were conducted between the original and recall PPAQ responses for total MET-minutes per week and percent time in light, moderate, and vigorous intensity activity for those variables that were normally distributed. Nine of the 24 variables were not normally distributed and therefore, Wilcoxon Sign Rank tests were used to compare the original and long-term recall PPAQ responses. Significance was set at an alpha level of p < 0.05. Spearman rank order correlations coefficients (SCC) were also used to compare these same variables. The participants were then separated into two groups via a median split: those who originally enrolled in The Mama Study five or more years ago and those who enrolled less than five years ago. This was done to see if time was an effect modifier in this relationship. The paired samples t-tests or Wilcoxon Sign Rank tests, and SCC were again conducted with the participants grouped. Correlations between the women who were enrolled five or more years ago and the women enrolled less than five years ago were then compared using Fisher Z tests and significance was again set at an alpha level of p < 0.05.

#### Results

Participants were an average of 167.1 cm tall, and 72.8, 78.3, and 69.6 kg at 21 and 32 weeks gestation and 12 weeks postpartum, respectively. Due to missed appointments or loss from follow up, a sample size of 40 (5+ years: 19, < 5 years: 21), 38 (5+ years: 18, < 5 years: 20), and 36 (5+ years: 18, < 5 years: 18) was used for the analyses for 21 and 32 weeks gestation and 12 weeks postpartum timepoints, respectively.

Means, standard deviations, and results from paired samples t-tests and Wilcoxon Sign Rank tests of total MET-minutes per week and percent time spent in light, moderate, and vigorous physical activity from the original and recall PPAQs at all time points are shown in Table 5.1 (Appendix). There was a significant difference between a majority of the original and recall values for total MET-minutes per week, light, moderate, and vigorous intensity activity at 21 weeks gestation (9 of the 12 values) and 32 weeks gestation (9 of the 12 values) according to the t-tests and Wilcoxon Sign Rank tests. Of the 12 tests between the original and recall PPAQ values at 12 weeks postpartum, only three were significantly different (total MET-minutes per week for all participants, total MET-minutes per week for participants who originally enrolled in The Mama Study less than five years ago, and moderate intensity activity for all participants). Original values were higher than recall values for total MET-minutes per week and moderate intensity activity, but were lower for light and a majority of the vigorous intensity activity, at all study time points.

Spearman correlations between original and recall PPAQs are represented in Table 5.2 (Appendix). Overall, correlations were moderate to strong with values ranging from 0.12 to 0.88 for all three time points. On average, correlations were highest at 21 weeks gestation (ranging from 0.42 to 0.88) and for light intensity (ranging from 0.32 to 0.72). Correlations for participants who enrolled in The Mama Study less than five years ago were higher for all variables, at all time points, except for vigorous intensity at 21 weeks gestation and 12 weeks postpartum. However, none of these correlations were significantly different from one another according to the Fisher Z test results.

## Discussion

The major finding from this study is women tend to underestimate their total METminutes per week and percent time spent in moderate intensity activity and overestimate the percent time spent in light intensity activity when asked to recall their physical activity during pregnancy between two months and eight years postpartum via the PPAQ, when compared to their responses collected during pregnancy. Women reported spending little time in vigorous intensity activity at both the original time point and when recalling their activity. Overall, these relationships did not change depending on how recently or long ago the women were pregnant. However, although they were not significantly different from one another, there were lower Spearman correlations for women who were recalling their physical activity five or more years postpartum compared to women who were recalling their physical activity less than five years postpartum for all time points and intensities, except vigorous activity at 21 weeks gestation and 12 weeks postpartum.

The concern with any self-report or recall physical activity method is the accuracy of the responses provided by participants. The only published study focusing on the long-term validity of self-report methods during pregnancy is by Bauer et al. (9), but it did not utilize the PPAQ or even compare the same self-report methods to each other. The authors found that women were able to report similar amounts of physical activity via the MAQ to what was recorded via a two-day diary, six years prior, when they were approximately 21 and 32 weeks gestation and 12 weeks postpartum (the same time points used in this current study). Bauer et al. (9) reported slightly higher correlations (Pearson correlations between 0.57 and 0.86) than our study (SCCs between 0.12 and 0.88), but both studies showed that when compared to their previously reported physical activity, women are fairly accurate at recalling their physical activity at

multiple time points during pregnancy. This could be because women may tend to be more conscious of their actions while pregnant, and therefore could have a better memory of their physical activity.

When comparing the average differences between original and recall values, the study time point and how long ago or recently the women were a part of The Mama Study does not appear to affect the results. On average, women consistently recalled 3000 – 4000 fewer total MET-minutes per week and 6% less moderate physical activity, but recalled 4-6% and 1-2% more light and vigorous physical activity, respectively, than what was reported at the original time point. It is difficult to directly compare our results to published studies because of differences in units. Some studies have found similar results of pregnant participants recalling less total physical activity than what was reported at the original time point, but most do not have a consistent trend for any of the intensities (5–7, 9, 11). However, these studies only evaluated short term (one to two weeks) recall *reliability*, rather than validity. Winters-Hart et al. (12) and Lee et al. (13) found that when non-pregnant participants recalled total physical activity between 10 and 17 years later, higher amounts were reported compared to original values, which disagrees with the trends found in the current study.

Although only one group of researchers have examined the correlations for long term self-report physical activity techniques in pregnant women, this issue has been studied in non-pregnant adults. Men and women were moderately able to recall their total, moderate, and vigorous physical activity from interview administered questionnaires two to eight years previously (SCCs = 0.38 - 0.84) (14). Alternatively, Lee et al. (13) found men and women to be poor at recalling physical activity participation from up to ten years previously (SCC = 0.38). Winters-Hart et al. (12) assessed the validity of the Historical Physical Activity Questionnaire

(HPAQ) when completed over 17 years. The authors asked women to complete a questionnaire about their past week physical activity four times between 1982 and 1999. They then completed the HPAQ about those four time periods in 1999. As expected, women's memory of their physical activity declined as time progressed. This agrees with our results as women who were recalling their physical activity five or more years ago had lower SCCs compared to those women who were recalling their physical activity less than five years ago. This represents that the ability for women to recall their physical activity may diminish over time, even though pregnancy is a more memorable time in most women's lives.

A limitation to the present study is the relatively homogenous sample of women. Most of the participants were of middle income and Caucasian which could lead to an increased risk of sampling error and selection bias. Future studies should be completed with various ethnic and social-economic status groups. We also only asked women to recall their physical activity once postpartum. It would be ideal if we asked the women to recall their physical activity once every year postpartum using the PPAQ so we could better track the recall trends over time for all participants.

#### Conclusion

Overall, our results show that it does not appear to matter if the women had their child recently or many years ago, the mean differences between original and recall values are similar, and women tend to underestimate their total and moderate physical activity and overestimate their light physical activity. However, according to the correlations, the error of recalling this activity was larger for women enrolled in The Mama Study five or more years ago, compared to women who enrolled in The Mama Study less than five years ago. This is important for future researchers as it is best to ask women to recall their physical activity during pregnancy using the

PPAQ less than five years postpartum, and we can assume that they will have likely underestimated their total physical activity. Because of the public health impact of physical activity on long-term health outcomes, historical physical activity questionnaires would be useful and it is important we continue to assess the long-term validity of self-report methods such as the PPAQ. This would also be helpful to create more specific physical activity recommendations for pregnant women because being able to include women in studies who had their children many months to years previously would allow us to have larger sample sizes and therefore, collect more information. Also, the PPAQ asks women to recall how much time they have spent in various activities in the past week. Although this questionnaire was created based on activities in which many pregnant and postpartum women participate, it may be missing some physical activities they spend a substantial amount of time in and/or questions could be more specific to help women recall activities. For example, one question in the PPAQ is how much time is spent feeding children while standing and another question asks how much time is spent carrying children. Some women would ask to clarify which question breastfeeding should be included in, but most did not. This may be problematic because when later recalling their activity, women may place breastfeeding in a different category than they did at the original enrollment, causing a larger disagreement between original and recall values. Altering the questionnaire to be more detailed or creating a new questionnaire that asks about more specific, common activities, may help women recall and have more valid results. Finally, it would be helpful if the questionnaire accounted for changes over time, especially if researchers are only able to collect data once during pregnancy or postpartum. Many women report feeling sick in their first trimesters and uncomfortable in their third trimesters, therefore, they participate in less physical activity compared to the second trimester. If women only recount their activity in the first or third

trimester, there may be an underestimation to their total physical activity. Alternatively, if women only recount their activity in the second trimester, there may be an overestimation. With all these factors under consideration, additional research is needed to create more detailed selfreport data collection techniques, so we can better understand the effects of physical activity on maternal and birth outcomes. APPENDIX

	21 Weeks Gestation		32 Weeks Gestation		12 Weeks Postpartum	
-	Original	Recall	Original	Recall	Original	Recall
Total MET Min/Wk						
All	18710.6 (9421.1)	14221.5 (4677.7) <sup>+</sup>	15263.6 (5555.1)	12040.4 (4340.6)*	21807.0 (7917.5)	18835.9 (4778.9) <sup>+</sup>
5 + years	18678.6 (9588.0)	$14218.7 \\ (4604.6)^+$	14847.3 (4521.1)	12220.9 (4305.4)*	21855.5 (8270.8)	19635.8 (3147.1)
< 5 years	18739.5 (7609.1)	14224.0 (4856.6)*	15638.4 (6441.5)	11878.0 (4477.3)*	21758.5 (7787.9)	18036.0 (5980.0)*
Light						
All	61.5 (19.5)	67.6 (15.9)*	65.2 (18.8)	71.1 (15.1)*	52.0 (13.6)	56.0 (13.2)
5 + years	58.3 (19.9)	66.4 (13.3)*	67.2 (14.7)	68.9 (12.6)	51.5 (14.2)	54.4 (14.2)
< 5 years	64.5 (19.2)	68.8 (18.2)	63.4 (22.1)	73.1 (17.1)*	52.5 (13.4)	57.6 (12.4)
Moderate						
All	35.3 (18.3)	28.2 (13.8)*	31.4 (16.8)	24.5 (12.8)*	45.5 (13.0)	39.9 (12.2)*
5 + years	38.7 (20.2)	29.7 (12.7)*	29.8 (14.1)	26.0 (11.5)	46.2 (14.5)	40.1 (13.9)
< 5 years	32.3 (16.4)	26.9 (14.9)*	32.8 (19.1)	23.1 (14.0)*	44.8 (11.7)	39.7(10.3)
Vigorous						
All	3.0 (4.1)	3.6 (4.1)	3.3 (5.6)	$4.3(6.2)^+$	2.4 (3.0)	3.9 (10.7)
5 + years	2.9 (4.3)	$3.8(4.4)^+$	2.8 (3.8)	$5.0(6.7)^+$	2.2 (3.0)	5.4 (14.9)
< 5 years	3.1 (4.0)	3.5 (4.0)	3.7 (6.8)	3.6 (5.8)	2.6 (3.1)	2.5 (3.0)

Table 5.1. Means and standard deviations of total metabolic equivalent minutes per week (MET Min/Wk) and percent time spent in light, moderate, and vigorous physical activity from the original Pregnancy Physical Activity Questionnaire (PPAQ) and recall PPAQ at three time points during pregnancy and postpartum and by time intervals.

*Note:* 5+ years: Data from women who originally enrolled in The Mama Study five years or more ago. < 5 years: Data from women who originally enrolled in The Mama Study less than five years ago. \*Original and recall values are significantly different at the 0.05 level according to paired samples t-test. <sup>+</sup>Original and recall values are significantly different at the 0.05 level according to Wilcoxon Sign-Rank test.

	21 Weeks	32 Weeks	12 Weeks
	Gestation	Gestation	Postpartum
	n = 40	n = 38	n = 36
Total MET			
Min/Wk			
All	0.70	0.64	0.60
5+ years	0.63	0.44	0.17
< 5 years	0.75	0.80	0.85
Light			
All	0.72	0.62	0.46
5+ years	0.63	0.36	0.29
< 5 years	0.79	0.84	0.62
Moderate			
All	0.70	0.55	0.32
5+ years	0.50	0.12	0.34
< 5 years	0.88	0.81	0.36
Vigorous			
Âll	0.46	0.53	0.64
5+ years	0.52	0.27	0.66
< 5 years	0.42	0.73	0.66

Table 5.2. Spearman correlation coefficients between the original Pregnancy Physical Activity Questionnaire (PPAQ) and recall PPAQ and by time intervals.

< 5 years 0.42 0.73 0.66</p> *Note:* 5+ years: Data from women who originally enrolled in The Mama Study five years or more ago. < 5 years: Data from women who originally enrolled in The Mama Study less than five years ago. \*5+ years and < 5 years correlations are significantly different at the 0.05 level.</p>

REFERENCES

# REFERENCES

- 1. Rudra CB, Williams MA, Lee I-M, Miller RS, Sorensen TK. Perceived exertion in physical activity and risk of gestational diabetes mellitus. *Epidemiol Camb Mass*. 2006;17(1):31–7.
- 2. Dempsey JC, Butler CL, Sorensen TK, et al. A case-control study of maternal recreational physical activity and risk of gestational diabetes mellitus. *Diabetes Res Clin Pract*. 2004;66(2):203–15.
- 3. Li R, Scanlon KS, Serdula MK. The validity and reliability of maternal recall of breastfeeding practice. *Nutr Rev.* 2005;63(4):103–10.
- 4. Chandonnet N, Saey D, Alméras N, Marc I. French Pregnancy Physical Activity Questionnaire Compared with an Accelerometer Cut Point to Classify Physical Activity among Pregnant Obese Women [Internet]. *PLoS ONE*. 2012 [cited 2017 Apr 19 ];7(6) available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3372468/. doi:10.1371/journal.pone.0038818.
- 5. Çırak Y, Yılmaz GD, Demir YP, Dalkılınç M, Yaman S. Pregnancy physical activity questionnaire (PPAQ): reliability and validity of Turkish version. *J Phys Ther Sci.* 2015;27(12):3703–9.
- 6. Ota E, Haruna M, Yanai H, et al. Reliability and Validity of the Vietnamese Version of the Pregnancy Physical Activity Questionnaire (ppaq). *Southeast Asian J Trop Med Public Health Bangk*. 2008;39(3):562–70.
- Xiang M, Konishi M, Hu H, et al. Reliability and Validity of a Chinese-Translated Version of a Pregnancy Physical Activity Questionnaire. *Matern Child Health J.* 2016;20(9):1940– 7.
- 8. Matsuzaki M, Haruna M, Nakayama K, et al. Adapting the Pregnancy Physical Activity Questionnaire for Japanese Pregnant Women. *J Obstet Gynecol Neonatal Nurs*. 2014;43(1):107–16.
- 9. Bauer PW, Pivarnik JM, Feltz DL, Paneth N, Womack CJ. Validation of an historical physical activity recall tool in postpartum women. *J Phys Act Health*. 2010;7(5):658–61.
- 10. Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc*. 2011;43(8):1575–81.
- 11. Tosun OC, Solmaz U, Ekin A, et al. The Turkish version of the pregnancy physical activity questionnaire: cross-cultural adaptation, reliability, and validity. *J Phys Ther Sci.* 2015;27(10):3215–21.

- 12. Winters-Hart CS, Brach JS, Storti KL, Trauth JM, Kriska AM. Validity of a questionnaire to assess historical physical activity in older women. *Med Sci Sports Exerc*. 2004;36(12):2082–7.
- 13. Lee MM, Whittemore AS, Jung DL. Reliability of recalled physical activity, cigarette smoking, and alcohol consumption. *Ann Epidemiol*. 1992;2(5):705–14.
- 14. Slattery ML, Jacobs DR. Assessment of ability to recall physical activity of several years ago. *Ann Epidemiol*. 1995;5(4):292–6.

#### **CHAPTER SIX: SUMMARY AND CONCLUSIONS**

## **Summary**

The purpose of this dissertation was to examine various physical activity assessment modalities, including devices and self-report questionnaires, when used throughout pregnancy and postpartum. Specifically, to assess the 1) reliability and validity of three physical activity devices, 2) correlation between a self-report questionnaire and a physical activity device worn at the hip and ankle, and 3) validity of women recalling their physical activity, all at 21 and 32 weeks gestation and 12 weeks postpartum. To clarify the optimal intensity and frequency of physical activity for women to participate in throughout and after pregnancy, and the effects of this physical activity on maternal and birth outcomes, reliable and valid assessments which measure similar amounts of physical activity are required.

### Reliability and Validity of Physical Activity Devices

Prior to the first paper of this dissertation, there were no published studies that have examined the reliability of any physical activity devices when worn during pregnancy and postpartum. Poor reliability potentially affects a device's validity and low reliability of devices has been shown when non-pregnant participants walk at slow speeds (1, 2). Because women tend to decrease the intensity as pregnancy progresses, it is important to assess device's reliability at multiple time points during pregnancy and postpartum (3). The results of the first paper of this dissertation implied that three commonly used physical activity measurement devices (ActiGraph accelerometer, Omron pedometer, and SenseWear Armband) have similar reliabilities at 21 and 32 weeks gestation and 12 weeks postpartum when completing various lifestyle and locomotor activities. However, the devices worn at the hip (ActiGraph and Omron) appeared to be the most reliable when worn during the third trimester. We hypothesize this to be because the devices would be less likely to move around the belt when pressed against a firmer abdominal area, coincident with women's stomachs in the third trimester. This would allow for little variation between the two visits and therefore, high reliability.

The validity of physical activity devices has only been assessed at one or two time points during pregnancy and postpartum and/or a second device was used as the criterion, rather than indirect calorimetry (4–8). However, a larger tilt angle of a device, slower walking speeds, and activities of daily living (ADLs) have been shown to decrease the validity of devices when worn by non-pregnant participants (1, 2, 9). This is important for pregnant women as the tilt angle of a device worn at the hip will likely increase, their walking speed decreases, and they participate in mostly ADLs. We determined that the devices again, had similar validities across all study time points, but the ActiGraph worn at the hip and ankle showed the highest values.

Overall, the results of the first portion of this dissertation imply that any of the three physical activity devices studied may be used when conducting research on women's ADLs and locomotor activities during pregnancy and postpartum. We also believe that the results from studies utilizing these devices are comparable because of their similar reliability and validity. However, the laboratory-based activities women participated in when assessing the reliability and validity of the devices were mostly moderate intensity. Few studies have assessed devices' reliability and validity at higher intensities and must be conducted in the future, particularly to evaluate relatively fit pregnant women.

Correlations Between the Pregnancy Physical Activity Questionnaire and ActiGraph Accelerometers

Part two of this dissertation was to examine the association between a popular self-report questionnaire, the Pregnancy Physical Activity Questionnaire (PPAQ), and a physical activity device, the ActiGraph accelerometer, worn at the hip and ankle at three time points during pregnancy and postpartum (21 and 32 weeks gestation, and 12 weeks postpartum). Overall, correlations between the methods were low to moderate (0.01 - 0.50) at all three time points. On average, correlations were highest between the PPAQ and hip ActiGraph at 32 weeks gestation (ranging from 0.34 - 0.40) and for light intensity (ranging from 0.20 - 0.50). The correlation between the PPAQ and ankle ActiGraph was highest at 21 weeks gestation. Also, the PPAQ underestimated percent time spent in light intensity activity and overestimated percent time spend in moderate intensity activity as compared to the hip ActiGraph.

The results of this study agree with those previously published focusing on the agreement between devices and self-report methods when worn and completed by pregnant and non-pregnant participants. Most studies with pregnant participants, have assessed the correlation between self-report and device-based methods at only one time point throughout pregnancy, but have also found low to moderate correlations (10–13). Troiano et al. (14) concluded that non-pregnant adults are greatly overestimating their physical activity when using self-report methods, or accelerometers are not able to capture all a person's physical activity and therefore, underestimating physical activity.

From an objective point of view, it would be helpful to researchers if self-report and device physical activity measurements produced high correlations on the amount of physical activity women participate in throughout pregnancy, so results can be compared from one study to another. However, comparing results of studies utilizing different measurement techniques is difficult because of the lack of criterion (15). We cannot state that either self-report or device-based techniques provide the most correct answer regarding true participants' physical activity is and there may be a limit to how well these two techniques can correlate as self-report

questionnaires are measuring human behavior, while devices are measuring human movement. Measuring physical activity is complicated and no single technique will likely be able to measure all aspects of physical activity. For example, Saris et al. (16) explains how physical activity can be expressed in many ways such as "energy, in the amount of work (watts), time period of activity, units of movement (counts), or even as a percentage based on the score of a questionnaire." Each of these represents physical activity in a different and specific way and researchers must consider which exact construct they want to focus on when deciding how to collect their physical activity data (17). However, it may be best if researchers used multiple data collection techniques (devices and self-report) in each of their studies when possible (15–18). This will likely provide a more true approximation of participants physical activity and understanding of how physical activity affects maternal and birth outcomes (16). At the present time, we believe it inappropriate to compare study results where different physical activity measurement techniques have been used.

The current Department of Health and Human Services (DHHS) physical activity recommendations for pregnant women are the same as for non-pregnant adults (150 minutes per week of moderate intensity activity) (19). However, these specific recommendations were not based on published studies and instead, were developed from assuming pregnant women would derive similar health benefits from physical activity as the non-pregnant population. If we want to use both self-report methods and devices to measure women's physical activity to create pregnancy specific recommendations, we may have to create two separate sets of recommendations: one based on data collected via devices and one based on data collected via self-report methods (20). We could then compare the results from studies to the appropriate recommendations, depending on the data collection technique used. Unfortunately, this has a

large feasibility challenge associated with it, as it is already difficult to disseminate current recommendations to medical professionals and the general public. For the time being, researchers should continue to promote the Department of Health and Human Services (DHHS) and American College of Obstetricians and Gynecologists (ACOG) recommendations and once/if two sets of recommendations are created, promote those only to the physical activity research community. We can then continue to conduct research and eventually form a more detailed set of recommendations for the public utilizing the combined information from self-report and device focused studies to better understand the effects of physical activity on maternal and birth outcomes.

## Historical Recall Validity of the Pregnancy Physical Activity Questionnaire

In the third and final part of this dissertation, we compared the responses from women who completed a PPAQ completed at 21 and 32 weeks gestation and 12 weeks postpartum to what they recalled about those same time periods, two months to eight years postpartum. The major finding from this study was, several months to years after delivery, women tend to underestimate their total MET-minutes per week and percent time spent in moderate intensity activity and overestimate the percent time spent in light intensity activity compared to their responses collected during pregnancy. Women reported spending little time in vigorous intensity activity at both original enrollment and when recalling their activity. On average, these relationships did not change depending on how recently or long ago the women were pregnant. However, the rank-order relationships changed to a greater extent for women who were recalling their physical activity five or more years postpartum.

Because researchers often ask women to recall their physical activity months to years after giving birth, it is important that the long-term validity of self-report physical activity methods be assessed at multiple time points during pregnancy and postpartum, so these selfreport methods can be used in future studies assessing the health impact of participating in physical activity during pregnancy. If we find questionnaires, such as the PPAQ, are valid for longer time periods, researchers will be able to include a larger number of women in their studies and collect more information. According to our results, it would be best to ask women to recall their activity during pregnancy using the PPAQ if they are less than five years postpartum, however we can assume that they will have likely underestimated their total physical activity.

A suggestion to improve the recall validity of the PPAQ and other questionnaires, would be to create more specific questions. For example, there is not a question specifically asking how much time a woman spends breastfeeding or how many hours of sleep she gets consecutively. Although this is a pregnancy, not postpartum, questionnaire, and therefore women may not have any children yet, many of the current questions already focus on children-based activities. If the PPAQ included more specific activities, it may help women remember their activity more accurately and researchers will gain more valid information. A second suggestion for future research is to alter the questionnaire to account for changes over time. Women's activities and the intensity of the activities likely change from the first to third trimester. If a questionnaire could account for these changes, then we are more likely to get a more well-rounded understanding of women's physical activity throughout pregnancy, rather than at just one time point. Finally, it is recommended that this be examined using a larger sample size. The author of this dissertation used a median split to examine how time may have affected the women's recall ability because breaking the sample into more than two groups would have precluded meaningful

statistical analysis. If a larger sample size is used in the future, researchers could assess how well women recall their physical activity every year postpartum.

## Limitations

Unfortunately, the researchers did not collect detailed information on the participants' racial/ethnic group, socio-economic status, marital status, education background, or other similar descriptive characteristics. Because most women were recruited from a college-town in the Mid-Michigan area, most women were white, and likely in the middle to upper socio-economic class, married, and had a college education. Therefore, the sample of women used for this dissertation is racially homogenous and current findings apply primarily to those who identify as the above listed characteristics. This is likely to not affect the results of the first aim of this dissertation (reliability and validity of physical activity devices) but has potential to alter the outcome of the second (correlations between the PPAQ and ActiGraph accelerometer) and third aims (recall validity of the PPAQ).

A second limitation is the ActiGraph cut points and MET values used were not pregnancy specific. The Freedson cut point used to categorize activity collected by the ActiGraph accelerometers were created using healthy adults as participants. Because pregnant women have different anatomies and physiologies and are likely to be less active compared to their non-pregnant selves and the non-pregnant population in general, cut points created with pregnant participants would likely provide more accurate results. We also collected physical activity data for women wearing ActiGraphs on the ankle, but cut points for any population have not been published for ActiGraphs worn at the ankle. We could therefore, not categorize intensity for this data. Finally, the Compendium of Physical Activity (21) MET values used for the PPAQ

activities were not pregnancy specific. Using MET values that were created using pregnant women, would again, likely improve the validity of the results.

## Strengths

These three studies are a unique addition the physical activity during pregnancy literature. This is the first study that has assessed the reliability of physical activity devices when worn during pregnancy and postpartum. This is beneficial as a device's validity is likely to be affected by its reliability and this dissertation provides important information for future researchers when determining which physical activity device to use. A handful of published studies have examined the validity of physical activity devices when worn during pregnancy and the correlations between self-report and device-based methods during pregnancy, but most have assessed only one or two time point during pregnancy. Because it has been suggested that slower walking speeds and increased tilt angle potentially affect the measurement of the ActiGraph and women's walking speed decreases and the tilt angle of devices worn at the hip likely increases throughout gestation, it is important to assess the measurement of the devices and the correlations between the PPAQ and ActiGraph at multiple time points during pregnancy.

This is also only the second study that has assessed the long-term validity of a self-report method during pregnancy. Bauer et al. (22) has examined this but used a two-day diary at baseline and a questionnaire for recalling, rather than utilizing the same method as we did in the third dissertation aim. Asking our participants to complete the PPAQ at both original enrollment and at the recall time point, makes understanding women's ability to recall their physical activity more clear, without having to be concerned about the overall agreement between the two methods.

## Conclusions

This dissertation addresses existing gaps in the literature on physical activity measurements used during pregnancy and postpartum by examining the reliability and validity of three popular physical activity devices, the correlation between a self-report and device-based physical activity data collection techniques, and the recall validity of the PPAQ, all at multiple time points during pregnancy and postpartum. The findings provide some clarity on the relationship between various methods and when these methods may be most appropriate to be used in studies focusing on physical activity during pregnancy. It also brings light to the many gaps in the literature. Several limitations explained above indicate areas for future research, such as creating ActiGraph cut points and MET values focusing on pregnant women. Such additions will provide valuable improvements on the results provided by both self-report and device-based data collection systems.

Information gathered here also suggests that more pregnancy specific physical activity recommendations to be created. The current recommendations of 150 minutes per week of moderate intensity activity or 20-30 minutes per day of moderate activity, most or all days of the week, are based on the assumption that pregnant women and their children will have similar health benefits to non-pregnant people being physically active (19, 23). We need more studies that examine the maternal and birth outcomes when women participate in various durations and intensities of physical activity during pregnancy and the effects of this differing activity during the three trimesters to form more specific recommendations.

We would also recommend that supplementary physical activity recommendations be published based on device measured data and/or researchers focus on creating a questionnaire that appropriately measures physical activity but also agrees with physical activity devices. The

current physical activity recommendations are based off data collected via questionnaires and we can therefore compare results from studies utilizing questionnaires to those recommendations. If we want to continue to measure women's physical activity throughout pregnancy with devices, additional recommendations based on data collected via devices need to be created.

It may also be beneficial if a questionnaire was created that focuses on changes over time. Many women participate in different amounts of physical activity depending on the trimester and how they are feeling that day/week. Because we examined how much time women spend in various intensities with both the PPAQ and the ActiGraph, new questionnaires should be made or current questionnaires could be altered to better mimic their activity in specific trimesters. If researchers ask women to recall their activity over their whole pregnancy, various time points may not be accounted for.

Overall, the results of this dissertation support the continued examination of physical activity measurement techniques during pregnancy and postpartum. This work will help to clarify the optimal frequency, duration, intensity, and type of physical activity for women to participate in throughout and after pregnancy, and the effects of this physical activity on maternal and birth outcomes.

REFERENCES

# REFERENCES

- 1. Brazeau A-S, Beaudoin N, Bélisle V, Messier V, Karelis AD, Rabasa-Lhoret R. Validation and reliability of two activity monitors for energy expenditure assessment. *J Sci Med Sport*. 2016;19(1):46–50.
- 2. De Cocker KA, De Meyer J, De Bourdeaudhuij IM, Cardon GM. Non-traditional wearing positions of pedometers: validity and reliability of the Omron HJ-203-ED pedometer under controlled and free-living conditions. *J Sci Med Sport*. 2012;15(5):418–24.
- 3. Borodulin K, Evenson KR, Wen F, Herring AH, Benson A. Physical Activity Patterns during Pregnancy. *Med Sci Sports Exerc*. 2008;40(11):1901–8.
- 4. Connolly CP, Coe DP, Kendrick JM, Bassett DR, Thompson DL. Accuracy of physical activity monitors in pregnant women. *Med Sci Sports Exerc*. 2011;43(6):1100–5.
- 5. Harrison CL, Thompson RG, Teede HJ, Lombard CB. Measuring physical activity during pregnancy. *Int J Behav Nutr Phys Act.* 2011;8:19.
- 6. Kinnunen TI, Tennant PWG, McParlin C, Poston L, Robson SC, Bell R. Agreement between pedometer and accelerometer in measuring physical activity in overweight and obese pregnant women. *BMC Public Health*. 2011;11:501.
- Smith KM, Lanningham-Foster LM, Welk GJ, Campbell CG. Validity of the SenseWear® Armband to predict energy expenditure in pregnant women. *Med Sci Sports Exerc*. 2012;44(10):2001–8.
- 8. Stein AD, Rivera JM, Pivarnik JM. Measuring energy expenditure in habitually active and sedentary pregnant women. *Med Sci Sports Exerc*. 2003;35(8):1441–6.
- 9. Crouter SE, Schneider PL, Bassett DR. Spring-levered versus piezo-electric pedometer accuracy in overweight and obese adults. *Med Sci Sports Exerc*. 2005;37(10):1673–9.
- Chandonnet N, Saey D, Alméras N, Marc I. French Pregnancy Physical Activity Questionnaire Compared with an Accelerometer Cut Point to Classify Physical Activity among Pregnant Obese Women [Internet]. *PLoS ONE*. 2012 [cited 2017 Apr 19];7(6) available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3372468/. doi:10.1371/journal.pone.0038818.
- 11. Chasan-Taber L, Schmidt MD, Roberts DE, Hosmer D, Markenson G, Freedson PS. Development and validation of a Pregnancy Physical Activity Questionnaire. *Med Sci Sports Exerc*. 2004;36(10):1750–60.

- 12. Matsuzaki M, Haruna M, Nakayama K, et al. Adapting the Pregnancy Physical Activity Questionnaire for Japanese Pregnant Women. *J Obstet Gynecol Neonatal Nurs*. 2014;43(1):107–16.
- 13. Evenson KR, Chasan-Taber L, Symons Downs D, Pearce EE. Review of Self-reported Physical Activity Assessments for Pregnancy: Summary of the Evidence for Validity and Reliability. *Paediatr Perinat Epidemiol.* 2012;26(5):479–94.
- 14. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2008;40(1):181–8.
- 15. Melanson EL, Freedson PS, Blair S. Physical activity assessment: A review of methods. *Crit Rev Food Sci Nutr.* 1996;36(5):385–96.
- 16. Saris WHM. Habitual physical activity in children: methodology and findings in health and disease. *Med Sci Sports Exerc*. 1986;18(3):253–63.
- 17. Sylvia LG, Bernstein EE, Hubbard JL, Keating L, Anderson EJ. A Practical Guide to Measuring Physical Activity. *J Acad Nutr Diet*. 2014;114(2):199–208.
- 18. Freedson PS. Field Monitoring of Physical Activity in Children. *Pediatr Exerc Sci.* 1989;1(2):8–18.
- 19. United States Department of Health and Human Services. 2008 Physical Activity Guidelines for Americans. 2008;
- 20. Fulton JE, Carlson SA, Ainsworth BE, et al. Strategic Priorities for Physical Activity Surveillance in the United States. *Med Sci Sports Exerc*. 2016;48(10):2057–69.
- 21. Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc*. 2011;43(8):1575–81.
- 22. Bauer PW, Pivarnik JM, Feltz DL, Paneth N, Womack CJ. Validation of an historical physical activity recall tool in postpartum women. *J Phys Act Health*. 2010;7(5):658–61.
- 23. ACOG. ACOG Committee Opinion No. 650: Physical Activity and Exercise During Pregnancy and the Postpartum Period. *Obstet Gynecol.* 2015;126(6):e135-142.