Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences

School for Information Technology

Master of Statistics

Master's thesis

Relationship between physical activity and gestational outcomes

Adama Kazienga

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :
dr. Ruth NYSEN

SUPERVISOR :
Dr. Dorien LANSSENS

2019
2020

Faculty of Sciences
School for Information Technology
Master of Statistics

Master's thesis

Relationship between physical activity and gestational outcomes

Adama Kazienga
Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :
dr. Ruth NYSEN
SUPERVISOR :
Dr. Dorien LANSSENS

Acknowledgements

Firstly and foremost, I would like to thank the almighty God, who helped, guided and favored me in all my activities. Glory to the almighty God.

Secondly, I would like to express my deepest gratitude to my supervisors, Dr Dorien LANSSENS and Dr Ruth NYSEN, for their motivation, constant support and guidance throughout this research work. You helped me by giving advice and I am very happy of being one of your students.

The next, I would also like to express my gratitude to all my lecturers at Hasselt University. Thank you for sharing your immense knowledge with wonderful experiences. The next, my great thank goes to all my classmates for their kind support throughout the last 2 years.

Lastly, but not the least, my deepest gratitude goes to my beloved family members and relatives for your support, advice, sacrifice and all the challenges you went through to raise me physically and mentally. All my day to day activities is nothing without your good wishes and heartfelt prayers. To all my friends, who provided support throughout these two years, thank you so much.

THANK YOU!!!

Abstract

Background: Gestational Hypertensive Disorders are major public health concern and affect approximately 5 to 10% of pregnant women leading to an increased risk of maternal morbidity and mortality. Previous studies have demonstrated that the physical activity is paramount of benefit in improving the occurrence and development of gestational hypertension as well as pre-eclampsia.

Objectives: The main objective of this study was to assess the relationship between physical activity and gestational outcomes. Moreover, it was also aimed at finding the recommended daily number of steps in lowering the risk of gestational hypertensive disorders.

Methods: The study was carried out under the PREMOM project over the period January 2015 to March 2018 at Ziekenhuis Oost-Limburg (Genk, Belgium). Pregnant women with GHD were enrolled and wore activity tracker, which recorded the daily number of steps and distance and later sent the data to a central dashboard. In addition, baseline characteristics were also recorded at the entry of the study. Two stage model was used in assessing the relationship between physical activity and gestational outcomes. Moreover, patients were classified into groups based on their longitudinal profiles and the group index was used in baseline logits model along with potential covariates in assessing the effect of the group index on the outcome towards finding recommended number of steps.

Results: A total of 283 pregnant women were enrolled in the project. Among them, 163 pregnant women were included in the current study and their daily physical activity and baseline characteristics were collected over the study period. The mean age of the patients was 30.68 (sd=4.68), whereas approximately half had a normal gestational outcome. We found that daily physical activity was not associated with gestational outcomes. However, the body mass index, the admission over the study period and the gestational age at the entry of the study were significantly associated with the outcome of interest. The patients were classified into three physical activity level with respective average number of steps per day of 2222, 3535 and 5890 for low, moderate and high physical activity levels. There was no association between the physical activity level and the outcome of interest, suggesting that no recommendation for daily number of steps was found with the data at hand.

Conclusion: The daily physical activity in the current study was not associated with the gestational outcomes. Likewise, no recommended daily number of steps was found with the data at hand. Therefore, further studies are needed in addressing this unmet need.

Keywords: PREMOM, Gestational Hypertensive Disorders, functional principal component analysis, baseline logits model, steps

Contents

List of Tables

List of Figures

1 Background

1.1 Introduction

Despite laudable efforts and commitment, Gestational Hypertensive Disorders (GHD) remain a public health concern in obstetrics worldwide. It affects approximately 5-10% of pregnancies and therefore, is associated with increased risk of both maternal and perinatal morbidity and mortality [1, 2]. The prevalence of GHD in pregnancy was documented by the Flanders register of prenatal outcome and was 4.6% and distributed as follows, 0.3% deliver before 34 weeks, 0.6% deliver between 34 and 37 weeks, and 3.7% deliver after 37 weeks [1]. The GHD can be defined as a new onset of hypertension after 20 weeks of gestation with elevated blood pressure (Systolic $\geq$ 140 and diastolic $\geq$90 mmHg; or diastolic $\geq$100 mmHg) [3]. Pre-eclampsia is characterized by the development of hypertension and proteinuria(0,3g/24hours) after 20 weeks of gestation and affects approximately 3-5% of pregnancies and can threaten the life of the mother and the fetus [4].

Therapeutic interventions are used in the management of women diagnosed with GHD through conventional care or remote monitoring interventions in obstetrical care. Besides the therapeutic interventions, several studies have proven the positive impact of physical activity on gestational outcomes. A study carried out by [5] found that physical exercise only were effective at lowering the odds of developing gestational hypertension and pre-eclampsia. In the same way, another study has documented that the women who did not exercise were three times more likely to develop gestational hypertensive disorders (OR=2.96, CI=1.29, 6.82) [6]. Similar findings were reported by [7] and revealed that moderate exercise over a period of 10 weeks lowered the diastolic blood pressure among pregnant women at risk of GHD.

Although previous studies have shown a positive effect of physical activity on the gestational outcomes, the most often conventional method for activity data collection used was through the questionnaire on the intensity [5, 6, 7]. However, in clinical trials as well as large observational studies, recent advances of wearable technologies have allowed continuous health monitoring. As an example, activity tracker, heart rate monitor and blood pressure monitor are the three such devices which are widely used [8]. An activity tracker is used in objectively measures of the level and timing of physical activity on a daily basis. Additionally, remote monitoring tools have proven to be effective in the management of the women with GHD in obstetrical care [6, 9]. In addition, the majority of studies utilized single measurement or summary measures(mean, median, maximum) of the exposure, i.e. the physical activity in assessing the association with the outcome of interest [10]. Few studies have investigated the impact of physical activity on gestational outcomes in women with GHD taking into account the longitudinal aspect of the predictors as well as the use of the remote monitoring tools for the data collection [10, 11]. This study aimed at assessing the impact of daily physical activity in women with GHD on gestational outcomes as well as finding the recommended cutoff in lowering the risk of GHD.

1.2 Objectives

In the current study, the main objectives were:

- To study the effect of physical activity on gestational outcomes;
- To identify a cut-off value to be used as the recommended number of steps or distance per day leading to a lower risk of GHD.

2 Methodology

2.1 Study material

2.1.1 Study Design

The Pregnancy Remote Monitoring (PREMOM) study was a longitudinal study aimed at assessing the added value of remote monitoring program of pregnant women diagnosed with GHD [1, 12, 13]. The study population was, therefore, all women diagnosed with GHD and also, who had their prenatal follow-up, and delivered at the outpatient clinic of Ziekenhuis Oost-Limburg. In line with the regulation of ethics in medical research, the study protocol was approved by the local ethics committee, responsible for the site [1]. Also, prior to the start of the study, each participant signed a consent form and then was followed remotely intensively from the study entry until delivery [1].
The participants were asked to wear an activity tracker day and night until delivery or hospital admission. The activity tracker recorded the daily distance and number of the steps of each participant. The data recorded were then transmitted to a web based dashboard developed by the Mobile Health Unit of University of Hasselt [1]. In addition to the daily measurement data collected throughout the study period, baseline characteristics were also recorded at the entry of the study. The study data was unbalanced due to the fact that pregnant women were enrolled at different gestational age i.e. first trimester, second trimester and third trimester, and then was followed until delivery.

2.1.2 Data description

The data used in the current study was from the PREMOM study carried out in Ziekenhuis Oost-Limburg (Genk, Belgium) over the period from January 2015 to March 2018 [13]. The dataset contains baseline data as well as time varying covariates, namely the distance and the number of steps measured on a daily basis using an activity tracker. A total of 283 women completed the study, of which, 163 (57.6%) had physical activity data as well as the baseline characteristics. The outcome of interest was the gestational outcomes, classified as no GHD, gestational hypertension and pre-eclampsia. The gestational hypertension was diagnosed by having a high blood pressure at delivery while the pre-eclampsia was evaluated by having both high blood pressure and proteinuria at delivery [4]. The maternal age in years was assessed at the study entry. In addition, the height (cm) and the weight (kg) were measured before the pregnancy without shoes, and heavy clothing and the body mass index (kg/m^2) was calculated based on these two measures. The gestational age in weeks at enrolment was established by fetal ultrasound examination. Information about Gravidity Parity Abortion (GPA) was obtained at enrolment, whereas the medication intake and the admission were obtained during the follow up of participants over the study period.

The gravidity of the participant, expressed the number of time a woman has been pregnant regardless of the pregnancy outcome. In addition, the number of pregnancies reaching viable gestational age beyond 24 weeks, called the parity and abortion was the number of miscarriages or abortion before 24 weeks of gestation experienced by each participant [14]. The information related to whether the participants have been admitted or took drugs for medical reasons over the study period was captured respectively by the variables admission and intake. The variables used in the analysis are displayed in Table 1.

Table 1: Variables description

Variable	Type	Unit	Description
Ntnumber			Identification number of each participant
Age	continuous	years	Age of the participant
Weight	continuous	kg	Weight of the participant before pregnancy
Length	continuous	m	Length of the participant
Body Mass Index	continuous	kg/m^2	Body Mass Index before pregnancy
Gestational age		weeks	Gestational age at inclusion
Gravidity	categorical	5 levels	Number of pregnancies including the current
Parity	categorical	4 levels	Number of deliveries after 24 weeks of gestation
Abortion	categorical	5 levels	Number of abortions/miscarriages before 24 weeks of gestation
Distance	continuous		Daily distance performed by the participant
Steps	continuous		Number of daily steps
Intake	categorical	2 levels	Medication intake during pregnancy
Admission	categorical	2 levels	Admission

2.1.3 Data management

As before mentioned in the study design, the daily measurement of the distance and number of steps were captured by the activity tracker and sent to the central dashboard [1]. However, there were some duplicates record due to a connectivity issue to the central dashboard, or the data were sent simultaneously with the blood pressure data which were sent twice a day. Therefore, one record was kept daily per participant in the analysis set considering the maximum value of distance and number of steps in case of duplicates record. Moreover, the Gravidity Parity and Abortion (GPA) variable were recorded. The gravidity variable, initially with five levels, was recoded into two levels, mainly Primigravidity (Gravidity $= 1$) and Multigravidity (Gravidity > 1). In the same way, the parity variable was grouped into Nulliparous (Parity $= 0$) and Multiparous (Parity ≥ 1). The Body Mass Index (BMI) was grouped into six levels, according to the World Health Organization [15]. These categories were respectively underweight, normal weight, overweight and obesity from grade 1 to grade 3. In addition, the gestational age at inclusion was also categorized into first trimester (0 - 12 weeks), second trimester (13 - 26 weeks) and third trimester (27 - 42 weeks) [16].

Table 2 indicates the way the variables were recorded.

Table 2: Variables recoding description

Variables	Recoding rule used
Gravidity	
Primigravidity	Gravidity = 1
Multigravidity	Gravidity > 1
Parity	
Nulliparous	Parity = 0
Multiparous	Parity ≥ 1
Abortion	
Yes	Abortion ≥ 1
No	Abortion = 0
Body Mass Index	
Underweight	BMI < 18.5
Normal weight	18.5 ≤ BMI ≤25
Overweight	25 ≤ BMI ≤ 29.9
Obesity Grade 1	30 ≤ BMI ≤ 34.9
Obesity Grade 2	35 ≤ BMI ≤ 39.9
Obesity Grade 3	BMI ≥ 40
Gestational age	
First trimester	0 - 12 weeks
Second trimester	13 - 26 weeks
Third trimester	27 - 42 weeks

2.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was used in getting an insight in the study data structure in making decisions for the subsequent steps of the analysis regarding the model selection. Graphical techniques as well as summary measures, were used for the EDA. Means and proportions were used as summary measures for quantitative and qualitative variables, respectively, while individual profile was used as graphical techniques for repeated measures variables.

2.3 Statistical methods for repeated predictor and scalar response

The primary objective of the current study was to assess the relationship between the physical activity measured repeatedly and the gestational outcomes adjusting for the baseline characteristics. Therefore, an appropriate statistical method should be chosen while taking into account the longitudinal features of the main predictor trajectory over the study period. The distance and the number of steps parameters were measured daily and over time from enrolment until delivery. The standard generalized linear mixed model cannot be used since it assumes correlated outcome data while in the current study, the outcome was non time-varying, and the predictors were time-varying [10].

The single measurement using summary measures of the repeated predictors, the functional principal component analysis as well as the two stage model are three statistical methods which can be used in fitting such kind of data and will be described and discussed in the subsequent sections.

2.3.1 Single measurement

The single measurement approach consists of computing at the first stage, the subject specific summary measure of the time-varying covariate using either the mean, the median, the geometric mean or the maximum [10]. Then, in the next step, a baseline logits model can be fitted using the categorical outcome and the subject specific averages of the repeated predictors while adjusting with the time invariant baseline characteristics of the pregnant women at the entry of the study. The model based on the mean as summary measure can be formulated as follows:
Let J denotes the number of categories for the outcome Y
Let $\{\pi_1, \pi_2, \pi_3, ...\pi_j\}$ denotes the response probabilities satisfying $\sum_{j=1}^{J} \pi_{ij} = 1$
The baseline category logit model is as follows:

$$log(\frac{\pi_{ij}}{\pi_{iJ}}) = \alpha_j + \beta_j X_i + \gamma_j Z_i, j = 1, 2, 3, ...J - 1 \quad (1)$$

Where $X_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$, the subject specific average number of steps, $Z_i = \{Z_1, Z_2, Z_3......Z_p\}$, the baseline characteristics parameters. The model has $J-1$ equations with separate parameters for each [17].
Although assessing the relationship between the physical activity and the gestational outcomes for women with GHD is of paramount importance in the management of these patients, this approach leads to biased results when the data are unbalanced and also a loss of information since the longitudinal features of the repeated predictor was not taking into account [10, 11].

2.3.2 Functional principal component analysis

The functional principal component analysis (FPCA) can be considered as a major tool for dimension reduction within the functional data analysis framework aimed at summarizing the infinite dimensional random trajectories considered to be an element of infinite dimension space to a finite number of functional principal component scores and is based on first and second order moment [18].
Lets consider $X_i(t)$ a realization of a time varying variable $X(t)$ for the i^{th} subject, $1 \leq i \leq n$ at each time point $\in$ T. Then, we assume also that the measurement between the different subjects are independent and also that $X_i(t)$ have smooth trajectories over time t. The FPCA is applied in decomposing the pattern of temporal variation in the repeated predictor. The basis features of the FPCA is that the time varying trait of the i^{th} subject admit the Karhunen-Loeve expansion [19]:

$$X_i = \mu_t + \sum_{k=1}^{K} \epsilon_{ik}\phi_k(t)$$

where $\mu_t = EX(t)$ and ϵ_{ik} are uncorrelated random variable with mean zero and variance λ_k satisfying $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ....\lambda_K$. The ϵ_{ik} are the k^{th} functional principal component scores of the i^{th} subject associated with the eigenfunction ϕ_k for all $k \geq 1$.
It can be noticed that after obtaining the individual scores, they can be entered as response or predictors in further statistical analysis as well as using for clustering.

The choice of the number of components to be used was based on the fraction of variance explained through the use of the scree-plot.
However, the standard functional principal component analysis cannot be applied in our case as we note that the set of time point $t_{i1}, t_{i2}, ...t_{iN}$ differ accross the study subject. Therefore, a functional principal component analysis was performed, taking into account this features of the data at hand, based on principal component analysis through conditional expectation (PACE) [20]. The principal analysis of random trajectories algorithm is used as the basis method in PACE for obtaining the Karhunen-Loeve representation of smooth functional data, where measurements are contaminated with additional measurement error. Therefore, in a first step, all available measurements are pooled into one scatter plot and used as a one dimensional smooth to obtain the estimate of $\hat{\mu}(t)$ of the overall mean function $\mu(t)$. Then, when the mean function and the eigenfunctions have been obtained, the next steps is completing the empirical Karhunen-Loeve representation, thus the dimension reduction is the estimate of the functional principal component scores [21].
The model using the functional principal component scores as summary measures of repeated predictors is similar to Equation 1 in Section 2.3.1, where $X_i = \{X_1, X_2, X_3....X_s\}$, are the first functional principal component scores with high fraction of variance explained of the data.
Although the functional principal components scores summarize both the overall measurement level as well as the measurements changes over time, its interpretation remain difficult [10].

2.3.3 Two stage model

An alternative way of modelling the data at hand is to use the two stage model approach by summarizing the measurements from an individual and then use the summary statistics through standard statistical methods while accounting for the between subject variability. The variability of the estimates at the first stage need to be take into account in the second stage i.e. the measurement error in providing valid inference for the second stage model parameter estimates [22]. In others words, it is assumed at the first stage, that a parametric curve, i.e. linear and quadratic trend fit the observations for each subject [23].
At stage 1, it is assumed that Y_i satisfies the polynomial regression:

$$Y_{ij} = \beta_{0i} + \beta_{1i} time_{ij} + \beta_{2i} time_{ij}^2 + \epsilon_{ij}$$

where Y_{ij} is the number of steps of i^{th} pregnant women measured at time $time_j$, β_{0i} is the subject specific intercept whereas β_{1i} and β_{2i} were the linear and quadratic subject specific slopes, ϵ_{ij} is the error term assumed to be independent and normally distributed. In the second stage, the individual specific intercept and slopes obtained from the first stage are used as predictors along with the relevant baseline characteristics collected at the entry of the study through a baseline logits model with gestational outcomes as response of interest. The second stage model formulation is similar to Equation 1 in Section 2.3.1,where $X_i = \{X_1, X_2, X_3\}$, individual specific slopes (linear and quadratic trend) and intercept.
Out of the three methods described above, the two stage model was chosen in assessing the relationship between the physical activity and the gestational outcomes. This can be justified by the fact that the two stage model takes into account the longitudinal nature of the physical activity parameters measured on daily basis over the study period and the interpretation of the parameters estimates is straightforward [10, 23]. Additionally, the study data was unbalanced and therefore, using single measurement such as the mean of the repeated predictor was not appropriate because it leads to loss of information and even biased results [21].

Although the functional principal analysis methods account for the longitudinal nature of the physical activity measures, its interpretation of the effect of functional principal scores on the outcome is not straightforward [10].
The baseline characteristics as well as the predicted intercept and slopes of the number of steps were included in the model as predictors. In addition, the participants' weight were measured weekly over the study period, and we planned also to include it in the model in capturing the association between the weight gain during the pregnancy and the outcome. Therefore, the subject specific intercept and slopes were also obtained from an ordinary regression, and only the slope was only included in the model. Moreover, an interaction between the body mass index and the daily steps slope was also considered. The likelihood ratio test was used for the statistical inference and the deviance and the Pearson statistics were used as goodness-of-fit statistics that provide a global model check [17]

2.4 Recommended number of steps

The second objective of the current study was to find a cut-off point as recommended daily number of steps in lowering the risk of GHD at delivery for enrolled patients. As aforementioned in Section 2.3.1, that the data at hand was unbalanced, using the subject specific average of the daily number of steps and then categorized using a cut-off point leads to bias and should be avoided in case of unbalanced data [24]. Previous studies have shown that the majority of guidelines of physical activity during pregnancy suggests moderate intensity. Additionally, the World Health Organization (WHO) guidelines on physical activity recommends that adult age between 18 to 64 years should engage in at least 150 minutes of moderate intensity aerobic activity throughout the week. In addition, the frequency of activity per week reported by different countries guidelines of physical activity during pregnancy ranged from 3 to 5 days per week and from 15 to 30 min as duration [25]. Also, the number of steps per minutes have been estimated by previous studies reported by Catrine [26], found that 100 steps per minutes was acceptable heuristic cadence threshold for moderate activity walking in 21 to 40 years olds. Another alternative naive method of clustering the patients was to consider these findings of previous studies, and try to classify patients into active or inactive based on their rate of achieving the chosen cut-off point. However, this will lead to severe bias, since at the later stage of the pregnancy, the number of daily steps tends to decrease as compared to early pregnancy. An appropriate method that take into account the longitudinal features of the daily measurement of the number of steps is the clustering method. Therefore hierarchical clustering method using agglomerative algorithm was used, in which objects are initially assigned to their own cluster and then pair of the cluster repeatedly merged until the whole tree is formed [27].
The clustering is performed in two stage. In the first stage, a polynomial regression was fitted in order to obtain the subject-specific intercept and slopes (linear and quadratic trend). Then, in the second stage, these estimates obtained from the first stage are used for the clustering and the cluster index obtained will be used for the second part, in modelling the association between the cluster index and the outcome of interest while adjusting for the baseline characteristics using a baseline logits model as described in Equation 1 of Section 2.3.1, where X_i is the cluster index.

2.5 Software

The SAS version 9.4 and R Studio version 3.6.1 were used for the data analysis. A pvalue of less than 5% was considered statistically significant.

3 Result

3.1 Exploratory Data Analysis

3.1.1 Baseline characteristics

A total of 283 patients completed the study until delivery. Among these, 163 (57.59%) subjects had baseline characteristics and time-varying covariates, mainly the distance and the number of steps measured on a daily basis and was used for the final analyses. Table 1 presents the baseline characteristics of the enrolled patients. It can be observed from the table, that the mean age was 30.68 (4.26) and half of the patients had a normal gestational outcome. There were no underweight patients in the current study, but the number of patients with normal weight was 71 (43.56%). In addition, the majority of participants were enrolled in the current study during their second and third trimester of pregnancy with 88 (53.99%) and 55 (33.71%) patients respectively in the second and third trimester. Moreover, over the study period, 46 (28.22%) patients were admitted to the hospital whereas 54 (33.13%) have taken medication. Looking at the gravidity, 39.26 % were primigravidae whereas just above half (50.42%) were nulliparous.

Table 3: Baseline characteristics

Parameters	Value
	n=163
Age in years, mean (SD)	30.68 (4.26)
Gestational outcomes, n (%)	
Normal	83 (50.92)
Gestational hypertension	54 (33.13)
Pre-eclamsia	26 (15.95)
Body mass index , n (%)	
Underweight	0 (0)
Normal weight	71 (43.56)
Overweight	43 (26.33)
Obesity grade 1	24 (14.72)
Obesity grade 2	15 (9.20)
Obesity grade 3	10 (6.13)
Gestational age in weeks, n (%)	
First trimester	20 (12.27)
Second trimester	88 (53.99)
Third trimester	55 (33.74)
Medication intake, n (%)	
Yes	54 (33.13)
No	109 (66.87)
Admission since monitoring, n (%)	
Yes	46 (28.22)
No	117 (71.78)
Gravidity, n (%)	
Primigravidae	64 (39.26)
Multigravidae	99 (60.74)
Parity, n (%)	
Nulliparous	83 (50.92)
Multiparous	80 (49.08)

3.1.2 Individual profiles

The individual profiles on the daily measurement of the number of steps were used in identifying the general pattern of each patient throughout the study period. Figure 1 illustrates the individual profiles of the patients over the study period. It can be noticed from the graph that there were much variability between and within individuals. In addition, it shows fluctuations in terms of the overtime evolution of the number of steps over the study period suggesting that the daily number of steps was not constant or decreasing as time increases. The data is highly unbalanced and should be taken into account in the model choice and the daily fluctuations also suggests that only a linear time effect might not be sufficient and an adjustment with quadratic time effect is needed.

Figure 1 illustrates the individual profiles of the daily number of steps of randomly selected patients, of which the time indicates the day in the study.

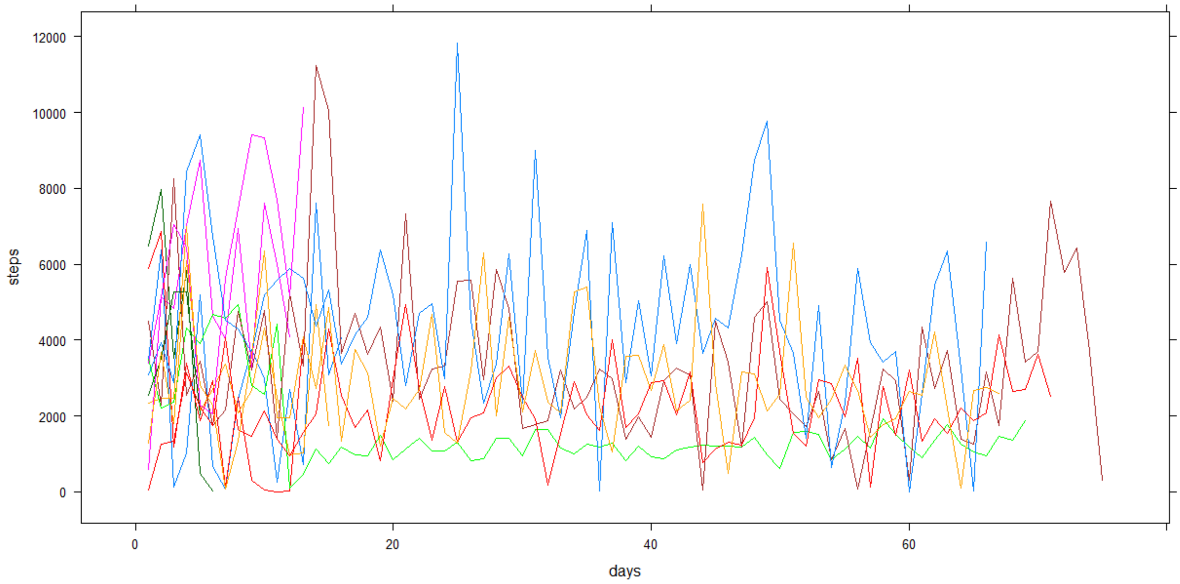

Figure 1: Individual profile of the number of steps over time

3.2 Two stage model

3.2.1 First stage

In the first stage, we have checked whether the linear time effect is sufficient to capture the variability of the daily number of steps in the current study. This first attempt shows that the linear time effect only does not capture very well the variability of the daily number of steps. Therefore, an adjustment was performed by adding a quadratic time effect and we obtained the fit plot of all the patients except the patients with only one measurement (5 in total). The majority of the fit plots checked and were like Figure 2. Therefore, we can conclude that adjusting the linear trend by a quadratic trend is more suitable because it capture most of variability in the data. The subject specific intercept as well as the linear and quadratic time effect were extracted in order to be used as covariates of interest in the second stage.

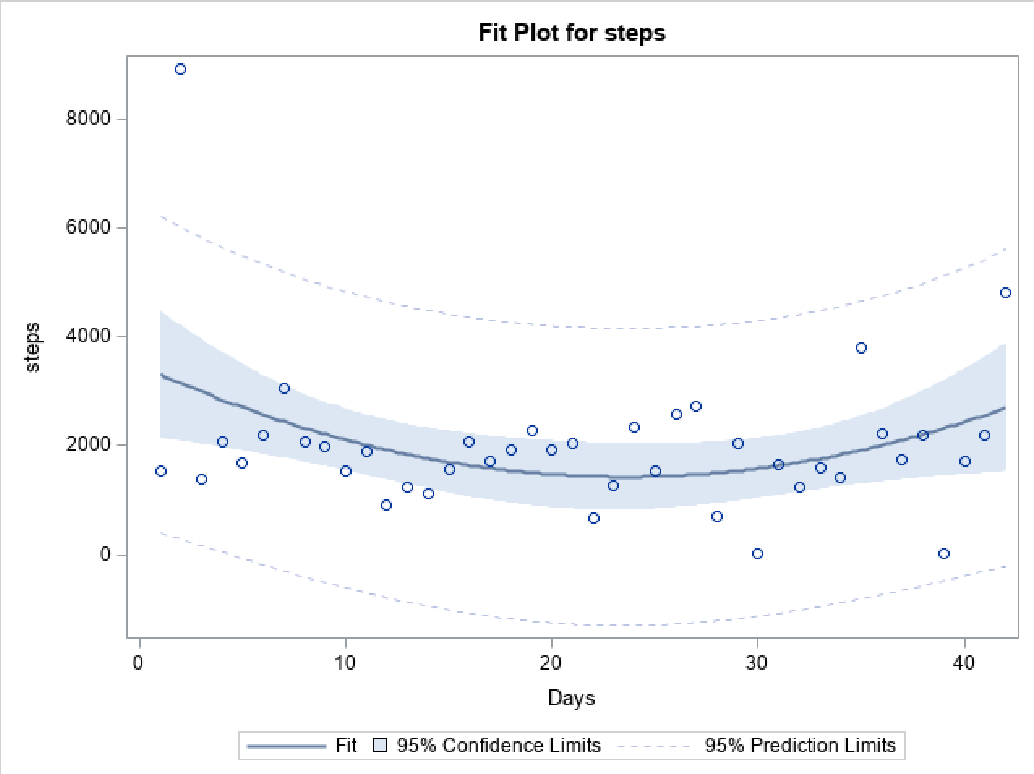

Figure 2: First stage fit plot of daily number of steps

3.2.2 Second stage

The parameter estimates from the first stage were used as well as the baseline characteristics of the enrolled patients. Also, it was adjusted by the measurement error obtained from the first stage in order to provide valid inference for the second stage model parameter estimates. Prior to the parameter estimates interpretation, the deviance and the pearson statistics tests was used to test the overall goodness of fit of the model. The p-value for the deviance and pearson statistics were respectively 0.980 and 0.924, suggesting that the model fits well the data. The results of the second stage are shown in Table 4. Holding the other variables constant, the log odds of obesity grade 1 patients relative to normal weight was 1.315 higher for having gestational hypertension to normal gestational outcome. In others words, obesity grade 1 patients were more likely to have gestational hypertension outcome as compared to normal weight patients (p=0.0458). Likewise, the multiparous patients, were more likely to have normal gestational outcome as compared to nulliparous (coef =-0.9495, p=0.0271). The pregnant women enrolled during the second and third trimester were more likely to have normal gestational outcome as compared to the reference group, i.e. the patients enrolled during the first trimester holding other variables fixed. As expected the patients who were admitted over the study period were more likely to have gestational outcome to gestational hypertension as compared to not admitted patients (p=0.0463).

Holding other variables constant, the obesity grade 3 patients were more likely to have pre-eclampsia to normal gestational outcome as compared to normal weight patients (coef=1.948, p=0.0469). In addition, as expected the log odds of the patients who were admitted over the study period was 1.94 higher for having pre-eclampsia to normal gestational outcome as compared to not admitted patients.
Table 5 displays the Type 3 analyses of effect based on the Wald test. It can be seen that the variables intercept and slopes of the daily number of steps, the patient age, and the parity are not statistically significant. However, the BMI, the gestational age at enrolment and the admission over the study period were associated with the gestational outcomes. Therefore, using the data at hand, there was no evidence that the physical activity was associated with gestational outcones.

Table 4: Parameters estimates two stage model

Parameter	DF	Estimate	SE	Wald ChiSq	Pvalue
Gestational hypertension					
Intercept	1	1.7087	0.9414	3.2943	0.0695
Number of steps					
Intercept steps	1	-0.6296	0.5878	1.1469	0.2842
slope linear trend steps	1	-0.3176	1.6663	0.0363	0.8488
slope quadratic trend steps	1	-0.3192	1.3029	0.06	0.8065
BMI					
Normal		Reference			
Overweight	1	0.3916	0.4878	0.6445	0.4221
Obesity grade 1	1	1.3218	0.6715	3.8747	0.049
Obesity grade 2	1	-1.4071	0.8746	2.588	0.1077
Obesity grade 3	1	-1.9511	1.2607	2.3953	0.1217
Gestational age at enrolment					
First trimester		Reference			
Second trimester	1	-1.6277	0.6641	6.007	0.0142
Third trimester	1	-2.0755	0.7805	7.0716	0.0078
Admission over study period					
No		Reference			
Yes	1	-1.5746	0.6549	5.7812	0.0162
Age	1	0.0886	0.2131	0.1728	0.6776
Parity					
Nulliparous		Reference			
Multiparous	1	-0.8923	0.4478	3.9701	0.0463
Pre-eclampsia					
Intercept	1	-6.4756	2.1683	8.919	0.0028
Number of steps					
Intercept steps	1	1.2374	0.9543	1.681	0.1948
Slope linear trend steps	1	3.1399	2.7572	1.2968	0.2548
Slope quadratic trend steps	1	2.1264	2.3139	0.8444	0.3581
BMI					
Normal		Reference			
Overweight	1	1.0271	0.8468	1.4711	0.2252
Obesity grade 1	1	1.7571	0.944	3.4641	0.0627
Obesity grade 2	1	0.2449	1.3172	0.0346	0.8525
Obesity grade 3	1	3.1551	1.182	7.1253	0.0076
Gestational age at enrolment					
First trimester		Reference			
Second trimester	1	1.9398	1.5127	1.6445	0.1997
Third trimester	1	1.5086	1.6299	0.8567	0.3547
Admission					
No		Reference			
Yes	1	3.2405	0.795	16.6134	<.0001
Age	1	-0.0468	0.3325	0.0198	0.8881
Parity					
Nulliparous		Reference			
Multiparous	1	0.0898	0.6886	0.017	0.8963

Table 5: Type 3 Effect two stage model

Effect	DF	Wald Chisq	Pvalue
Intercept steps	2	3.2978	0.1923
Slope linear trend steps	2	1.4965	0.4732
Slope quadratic trend steps	2	1.0926	0.5791
Body Mass Index	8	19.4561	0.0126
Gestational age at enrolment	4	10.7455	0.0296
Admission	2	25.5269	<.0001
Age	2	0.2181	0.8967
Parity	2	4.1026	0.1286

3.3 Recommended number of steps

3.3.1 Clustering of patients

The longitudinal profiles of the study patients were used through hierarchical clustering techniques in grouping patients based on the similarities of their over time evolution. Then, the cluster index obtained was included as covariates in a baseline logits model in assessing the association between the cluster index and the gestational outcomes adjusting with the potential covariates collected at the entry of the study. Table 6 displays the distribution of the cluster analysis. It can be seen that three clusters were identified and classified as high, moderate and low physical activity levels. The number of patients was respectively 19 (12.34%), 109 (70.78 %) and 26 (16.88 %) for low, moderate and high physical activity level. In terms of the average number of steps per day, it was respectively 2222, 3535 and 5890.

Table 6: Distribution average daily number of steps per physical activity level

Parameters	n (%)	Average number of steps per day
Low level	19 (12.34)	2222
Moderate level	109 (70.78)	3535
High level	26 (16.88)	5890

3.3.2 Modeling recommended number of steps

The cluster index obtained from the clustering method was used as well as the potential covariates collected at the entry in the study in assessing the relationship between the cluster index and outcome of interest. The deviance and the pearson statistic tests were used in checking the overall goodness-of-fit of the fitted model. The p-value for the deviance and pearson statistics were respectively 0.989 and 0.1125, suggesting that the model fits well the data. The results of the fitted model are shown in Table 7. It can be seen that there was no association between the gestational outcome and the physical activity levels of the enrolled patients, keeping other variables constant. Also, the patients enrolled during their second and third trimester were likely to have normal gestational outcome to gestational hypertension as compared to the reference group enrolled during their first trimester (p=0.0105, p=0.0039 respectively). Moreover, obesity grade 1 patients were more likely to have gestational hypertension outcome as compared to normal weight patients (p=0.0351), holding the other covariates fixed. Also, the multiparous patients were more likely to have normal gestational outcome to gestational hypertension as compared to the nulliparous patients, keeping the others predictors constant (p=0.0386).

In addition, the log odds of obesity grade 3 was 2.81 (p=0.016) higher of having pre-eclampsia to normal gestational outcome as compared to normal weight patients, holding the others variables fixed. Also, as expected the patients who were admitted were likely to have pre-eclampsia as compared to the reference group i.e. the patients who were not admitted over the study period (coef=3.51, $p < 0.0001$).
Table 8 displays the Type 3 analyses of effect based on the Wald test. It can be observed that the BMI, the gestational age at the entry of the study as well as the admission over the study period were significantly associated with the outcome of interest in the current study. In contrast, the physical activity level, the patient age and the parity are not associated with the gestational outcomes.

Table 7: Parameters estimates recommended number of steps

Parameter	DF	Estimate	SE	Wald ChiSq	Pvalue
Gestational hypertension					
Intercept	1	1.6495	0.9108	3.2798	0.0701
Physical activity level					
Low		Reference			
Moderate	1	-0.5465	0.7989	0.4679	0.494
High	1	-0.0078	0.6572	0.0001	0.9905
BMI					
Normal		Reference			
Overweight	1	0.4314	0.4769	0.8181	0.3657
Obesity grade 1	1	1.3776	0.6538	4.4392	0.0351
Obesity grade 2	1	-1.3088	0.846	2.3934	0.1218
Obesity grade 3	1	-1.9316	1.2416	2.4204	0.1198
Gestational age					
First trimester		Reference			
Second trimester	1	-1.6105	0.6293	6.5487	0.0105
Third trimester	1	-2.0079	0.6948	8.3506	0.0039
Admission					
No		Reference			
Yes	1	-1.4744	0.6309	5.4613	0.0194
Age	1	0.0908	0.2127	0.1823	0.6694
Parity					
Nulliparous		Reference			
Multiparous	1	-0.9007	0.4355	4.2774	0.0386
Pre-eclampsia					
Intercept	1	-3.8438	1.6166	5.6537	0.0174
Physical activity level					
Low		Reference			
Moderate	1	0.894	1.1036	0.6562	0.4179
High	1	-1.0852	0.9125	1.4143	0.2343
BMI					
Normal		Reference			
Overweight	1	0.5305	0.839	0.3999	0.5272
Obesity grade 1	1	1.4601	0.8735	2.794	0.0946
Obesity grade 2	1	0.4432	1.2823	0.1195	0.7296
Obesity grade 3	1	2.8113	1.1136	6.3731	0.0116
Gestational age					
First trimester		Reference			
Second trimester	1	0.9973	1.2455	0.6411	0.4233
Third trimester	1	-0.0934	1.2985	0.0052	0.9426
Admission					
No		Reference			
Yes	1	3.5169	0.8607	16.6978	$<.0001$
Age	1	-0.0777	0.3176	0.0598	0.8068
Parity					
Nulliparous		Reference			
Multiparous	1	-0.1767	0.642	0.0758	0.7831

Table 8: Type 3 Effect recommended number of steps

Effect	DF	Wald ChiSq	Pvalue
Physical activity level	4	5.7827	0.216
BMI	8	18.9388	0.0152
Gestational age	4	12.4311	0.0144
Admission	2	25.0895	$<.0001$
Age	2	0.2861	0.8667
Parity	2	4.2776	0.1178

4 Discussion

This study examined the association between the physical activity throughout the pregnancy and the gestational outcomes. The result revealed that, there was no association between the physical activity and the outcome of interest whereas significant effect were found between the response and potential covariates such as the BMI, the gestational age at enrolment and admission status. In addition, no association was found between the cluster index and the gestational outcomes, indicating that no recommended daily number of steps leading to lower risk of GHD was found with the data at hand.

Daily walking of the study patients throughout the study period was not associated with the gestational outcomes. In the contrast, a study carried out by [28] has found that regular maternal walking over the pregnancy period may be a preventive tool for pre-eclampsia as well as reduces the incidence of gestational hypertension. Our findings were not in line with the results of this previous study because the study design, the study population as well as the potential covariates of interest collected were different. In the current study, the design was observational whereas, it was interventional in the study carried out by [28], in which patients were allocated to either control group (usual care) and study group (adherence program of regular walking for 5 days per week, for 30 min). Also, women with different gravidity levels as well as from different gestational age were enrolled in the current study while only women in first trimester and overweight primigravidae were enrolled in the study carried out by [28]. In terms of the potential covariates collected throughout the study period, the physical activity parameters were collected daily in our study whereas in the comparative study, only the group index was used to assess the effect of the walking on the outcome of interest. Likewise, pregnant women engaged in any regular physical activity as compared to inactive women, experienced a 35% reduced risk of pre-eclampsia [29]. Also, the difference in the conclusion between this study and our study may be attributed to the fact that activity tracker was used in our study to measure accurately the physical activity parameters whereas in the study carried out by [29], a questionnaire was used and the participant were asked to report the frequency as well as the type of physical activity, they are performing during the pregnancy such that walking, bicycling or swimming.

Looking at the variable BMI, as expected, it was found to be significantly associated with the gestational outcomes and this was in line with the results of previous studies. To illustrates, a study carried out by [30] found that obese women have a two-fold to three fold higher risk of developing gestational hypertensive disorders. Likewise, an increased incidence of pre-eclampsia as well as gestational hypertension was documented by [31]. In the same way, the results of a study carried out by [32], revealed that there is a signficant relationship between BMI and the risk of pre-eclampsia. Therefore, there is a need for women of normal prepregnancy BMI to follow advice of prenatal care providers in preventing excessive gestational weight gain that can caused adverse gestational outcomes.

It was also found that the gestational age at enrolment was also associated with the outcome of interest. However, the finding was counter-intuitive, because the patients enrolled in their second or third trimester were more likely to have normal gestational outcome to gestational hypertension as compared to the reference group, i.e. the first trimester patients. This may be attributed to the fact that, although, these patients were enrolled during their second and third trimester, they had performed daily physical activity prior to their enrolment.

Also, only few number of patients were enrolled during the first trimester of pregnancy in the current study.

Admission over the study period was significantly associated with the gestational outcomes. As expected, the patient who were admitted over the study period were more likely to have pre-eclampsia to normal gestational outcome as compared to not admitted patients. This can be explained by the fact that, they were admitted because they have shown some signs of pre-eclampsia. There is an urgent need to provide useful information about GHD that enables women to spot signs and symptoms of these diseases. As a result, this leads to earlier diagnoses and timely management of GHD in reducing the risk of adverse gestational outcomes [13].

Also, there was no association between the physical activity level identified through the clustering method towards findings recommended number of steps and the gestational outcomes, suggesting that no recommended daily number of steps leading to a lower risk of GHD was found with the data at hand. The study patients were divided into three physical activity levels over the study period, based on their longitudinal profiles using appropriate statistical method with respectively average daily number of steps of 2222, 3535 and 5890 for low, moderate and high physical activity level. In the litterature, a study carried out by [33], recommended 6000 steps a day in lowering the risk of gestational diabetes mellitus. However, in the current study, out of the 163 enrolled patients, only 26 (16.88 %) of the patients were in the high physical activity level group with 5890 as average daily number of steps, suggesting further studies are needed.

The physical activity data were measured throughout the study period using the activity tracker, and this can be considered as a strength as compared to self-reported questionnaires which are subject to large measurement errors and also uncontrollable bias [8]. However, since the design of the study was observational and no causal inference can be made [34]. Therefore, further studies are needed, of which potential eligible women will be randomized either in usual care or intervention group. In the usual care, daily measurement would be captured from the entry until delivery whereas the intervention group will involved in a walking adherence. This regular walking adherence can use the recommendations of the WHO, i.e. 5 days per week , for 30 minutes [25]. Also, in the current study, few patients were enrolled during their first trimester of pregnancy, there is a need for further studies to enrol the patients during their first trimester in order to study the effect of early pregnancy physical activity on the gestational outcomes. Also, the two stage model used as statistical method in the analysis of the data may be limited in case of elevated number of few measurements for the patient such as patients enrolled at their last week of pregnancy.

5 Conclusion

In this current study, we found that taking into account the longitudinal features of the physical activity, there was no significant effect between the physical activity and the gestational outcomes. In the contrast, significant effect were found between the body mass index, the gestational age at enrolment and the admission status over the study period. Moreover, the cluster analysis identified three physical activity levels and revealed also that there was no association between the physical activity level over the study period and gestational outcome. Therefore, the recommended number of daily steps that leading to lower risk of GHD at delivery was not found with the data at hand. Therefore, further research is needed in addressing this important unmet need towards improvement of the management of GHD patients during pregnancy.

References

[1] Dorien Lanssens, Thijs Vandenberk, Christophe JP Smeets, Hélène De Cannière, Sharona Vonck, Jade Claessens, Yenthel Heyrman, Dominique Vandijck, Valerie Storms, Inge M Thijs, et al. Prenatal remote monitoring of women with gestational hypertensive diseases: cost analysis. *Journal of medical Internet research*, 20(3):e102, 2018.

[2] Romy Gaillard, Rachel Bakker, Sten P Willemsen, Albert Hofman, Eric AP Steegers, and Vincent WV Jaddoe. Blood pressure tracking during pregnancy and the risk of gestational hypertensive disorders: the generation r study. *European heart journal*, 32(24):3088–3097, 2011.

[3] Alessia Mammaro, Sabina Carrara, Alessandro Cavaliere, Santina Ermito, Angela Dinatale, Elisa Maria Pappalardo, Mariapia Militello, and Rosa Pedata. Hypertensive disorders of pregnancy. *Journal of prenatal medicine*, 3(1):1, 2009.

[4] Ben WJ Mol, Claire T Roberts, Shakila Thangaratinam, Laura A Magee, Christianne JM De Groot, and G Justus Hofmeyr. Pre-eclampsia. *The Lancet*, 387(10022):999–1011, 2016.

[5] Margie H Davenport, Stephanie-May Ruchat, Veronica J Poitras, Alejandra Jaramillo Garcia, Casey E Gray, Nick Barrowman, Rachel J Skow, Victoria L Meah, Laurel Riske, Frances Sobierajski, et al. Prenatal exercise for the prevention of gestational diabetes mellitus and hypertensive disorders of pregnancy: a systematic review and meta-analysis. *British journal of sports medicine*, 52(21):1367–1375, 2018.

[6] Ruben Barakat, Mireia Pelaez, Yaiza Cordero, Maria Perales, Carmina Lopez, Javier Coteron, and Michelle F Mottola. Exercise during pregnancy protects against hypertension and macrosomia: randomized clinical trial. *American journal of obstetrics and gynecology*, 214(5):649–e1, 2016.

[7] S Yeo, NM Steele, MC Chang, SM Leclaire, DL Ronis, and R Hayashi. Effect of exercise on blood pressure in pregnant women with a high risk of gestational hypertensive disorders. *The Journal of reproductive medicine*, 45(4):293–298, 2000.

[8] Jiawei Bai, Yifei Sun, Jennifer A Schrack, Ciprian M Crainiceanu, and Mei-Cheng Wang. A two-stage model for wearable device data. *Biometrics*, 74(2):744–752, 2018.

[9] Sarah M Nusser, Stephen S Intille, and Ranjan Maitra. Emerging technologies and next-generation intensive longitudinal data collection. *Models for intensive longitudinal data*, pages 254–277, 2006.

[10] Yin-Hsiu Chen, Kelly K Ferguson, John D Meeker, Thomas F McElrath, and Bhramar Mukherjee. Statistical methods for modeling repeated measures of maternal environmental exposure biomarkers during pregnancy in association with preterm birth. *Environmental Health*, 14(1):9, 2015.

[11] Marieke Welten, Marlou LA de Kroon, Carry M Renders, Ewout W Steyerberg, Hein Raat, Jos WR Twisk, and Martijn W Heymans. Repeatedly measured predictors: a comparison of methods for prediction modeling. *Diagnostic and prognostic research*, 2(1):5, 2018.

[12] Dorien Lanssens, Sharona Vonck, Valerie Storms, Inge M Thijs, Lars Grieten, and Wilfried Gyselaers. The impact of a remote monitoring program on the prenatal follow-up of women

with gestational hypertensive disorders. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 223:72–78, 2018.

[13] Dorien Lanssens, Thijs Vandenberk, Christophe JP Smeets, Hélène De Cannière, Geert Molenberghs, Anne Van Moerbeke, Anne van den Hoogen, Tiziana Robijns, Sharona Vonck, Anneleen Staelens, et al. Remote monitoring of hypertension diseases in pregnancy: a pilot study. *JMIR mHealth and uHealth*, 5(3):e25, 2017.

[14] Gravidity and parity. https://en.wikipedia.org/wiki/Gravidity_and_parity. Accessed: 2020-03-24.

[15] Ellis Voerman, Susana Santos, Hazel Inskip, Pilar Amiano, Henrique Barros, Marie-Aline Charles, Leda Chatzi, George P Chrousos, Eva Corpeleijn, Sarah Crozier, et al. Association of gestational weight gain with adverse maternal and infant outcomes. *Jama*, 321(17):1702–1715, 2019.

[16] Trimesters pregnancies. https://www.24baby.nl/zwangerschapskalender/zwangerschapstrimesters/. Accessed: 2020-03-24.

[17] Alan Agresti. *Foundations of linear and generalized linear models*. John Wiley & Sons, 2015.

[18] JO Ramsay, Giles Hooker, and Spencer Graves. Introduction to functional data analysis. In *Functional data analysis with R and MATLAB*, pages 1–19. Springer, 2009.

[19] Kyunghee Han, Pantelis Z Hadjipantelis, Jane-Ling Wang, Michael S Kramer, Seungmi Yang, Richard M Martin, and Hans-Georg Müller. Functional principal component analysis for identifying multivariate patterns and archetypes of growth, and their association with long-term cognitive development. *PloS one*, 13(11), 2018.

[20] Piotr Kokoszka and Matthew Reimherr. *Introduction to functional data analysis*. CRC Press, 2017.

[21] Garrett Fitzmaurice, Marie Davidian, Geert Verbeke, and Geert Molenberghs. *Longitudinal data analysis*. CRC press, 2008.

[22] Adam A Szpiro and Christopher J Paciorek. Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics*, 24(8):501–517, 2013.

[23] Geert Verbeke. Linear mixed models for longitudinal data. In *Linear mixed models in practice*, pages 63–153. Springer, 1997.

[24] Garrett M Fitzmaurice, Nan M Laird, and James H Ware. *Applied longitudinal analysis*, volume 998. John Wiley & Sons, 2012.

[25] Kelly R Evenson, Ruben Barakat, Wendy J Brown, Patricia Dargent-Molina, Megumi Haruna, Ellen M Mikkelsen, Michelle F Mottola, Katrine M Owe, Emily K Rousham, and SeonAe Yeo. Guidelines for physical activity during pregnancy: comparisons from around the world. *American journal of lifestyle medicine*, 8(2):102–121, 2014.

[26] Catrine Tudor-Locke, Elroy J Aguiar, Ho Han, Scott W Ducharme, John M Schuna, Tiago V Barreira, Christopher C Moore, Michael A Busa, Jongil Lim, John R Sirard, et al. Walking cadence (steps/min) and intensity in 21–40 year olds: Cadence-adults. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1):1–11, 2019.

[27] Osama Abu Abbas. Comparisons between data clustering algorithms. *International Arab Journal of Information Technology (IAJIT)*, 5(3), 2008.

[28] Nahla W Shady, Hany F Sallam, Shymaa S Ali, and Ahmed M Abbas. The effect of regular daily walking on adverse pregnancy outcomes among overweight primigravidas: a prospective cohort study. *Proceedings in Obstetrics and Gynecology*, 7(3):1–9, 2017.

[29] Tanya K Sorensen, Michelle A Williams, I-Min Lee, Edward E Dashow, Mary Lou Thompson, and David A Luthy. Recreational physical activity during pregnancy and risk of preeclampsia. *Hypertension*, 41(6):1273–1280, 2003.

[30] Ravi Thadhani, Meir J Stampfer, David J Hunter, Joann E Manson, Caren G Solomon, and Gary C Curhan. High body mass index and hypercholesterolemia: risk of hypertensive disorders of pregnancy. *Obstetrics & Gynecology*, 94(4):543–550, 1999.

[31] David G Lombardi, John R Barton, John M O'Brien, Niki K Istwan, and Baha M Sibai. Does an obese prepregnancy body mass index influence outcome in pregnancies complicated by mild gestational hypertension remote from term? *American journal of obstetrics and gynecology*, 192(5):1472–1474, 2005.

[32] Morteza Motedayen, Mohammad Rafiei, Mostafa Rezaei Tavirani, Kourosh Sayehmiri, and Majid Dousti. The relationship between body mass index and preeclampsia: A systematic review and meta-analysis. *International Journal of Reproductive BioMedicine*, 17(7):463, 2019.

[33] Ayako Hayashi, Hidenori Oguchi, Yumi Kozawa, Yukiko Ban, Junji Shinoda, and Nobuhiko Suganuma. Daily walking is effective for the management of pregnant women with gestational diabetes mellitus. *Journal of Obstetrics and Gynaecology Research*, 44(9):1731–1738, 2018.

[34] Pedro F Saint-Maurice, Richard P Troiano, David R Bassett, Barry I Graubard, Susan A Carlson, Eric J Shiroma, Janet E Fulton, and Charles E Matthews. Association of daily step count and step intensity with mortality among us adults. *Jama*, 323(12):1151–1160, 2020.

[35] Thijs Vandenberk, Jelle Stans, Christophe Mortelmans, Ruth Van Haelst, Gertjan Van Schelvergem, Caroline Pelckmans, Christophe JP Smeets, Dorien Lanssens, Hélène De Cannière, Valerie Storms, et al. Clinical validation of heart rate apps: mixed-methods evaluation study. *JMIR mHealth and uHealth*, 5(8):e129, 2017.

[36] Dorien Lanssens, Thijs Vandenberk, Inge M Thijs, Lars Grieten, and Wilfried Gyselaers. Effectiveness of telemonitoring in obstetrics: scoping review. *Journal of medical Internet research*, 19(9):e327, 2017.

[37] Wilfried Gyselaers, Sharona Vonck, Anneleen Simone Staelens, Dorien Lanssens, Kathleen Tomsin, Jolien Oben, Pauline Dreesen, and Liesbeth Bruckers. Body fluid volume homeostasis is abnormal in pregnancies complicated with hypertension and/or poor fetal growth. *PloS one*, 13(11), 2018.

R and SAS Codes

```
 ######## Important codes only #######

######### Plotting individual profiles for randomly selected patients#########
dataset<-read.table('C:/Uhasselt/MscThesis/predictor.csv', sep=',', header=TRUE)
attach(dataset)
library(tidyverse)
######### Order observations per ID and measurement date#########
dataset<-dataset[order(ID, measured_at),]
dataset1<-dataset %>%
  group_by(ID) %>%
  mutate(Days = 1:n())
######### Select randomly 12 patients and then plot their profiles #########
dataset2<-dataset1[dataset1$ID %in%sample(unique(dataset1$ID), 12, replace = FALSE),]
xyplot(steps~Days,data =dataset2,groups=ID,type='l')


/*********************************************************************************/
/*SAS Code first objective: Relationship physical activity and gestational outcomes*/

/*import the dataset*/
proc import datafile = 'C:\Uhasselt\MscThesis\predictors.csv'
 out = predictors
 dbms = CSV
 ;
run;
/*sorting the dataset*/
proc sort data=dataquad;
by id days;
run;
/* obtain fits plots for first stage, two stage model*/
ods graphics on;
proc glm data=predictors ;
model steps = days days*days;
by id;
run;
ods graphics off;

/*get the predicted intercept, slopes for teh daily number of steps per subject*/
data predict;
set predict;
days2=days*days;
run;

proc reg data=predict outest=stage1 edf noprint ;
model steps=days days2;
by id;
```

```
run;
quit;

/*fit the baseline logits model for the first objective*/
proc import datafile = 'C:\Uhasselt\MscThesis\dataset1.csv'
 out = dataset1
 dbms = CSV
 ;
run;
proc logistic data = dataset1;
class  catbmi(ref = "2") catga(ref = "1")  admission (ref = "0")
catparity (ref = "0") / param = ref;
model outcome(ref = "0") = interstep daystep  days2step mestep catbmi
catga  admission  age catparity  / link = glogit aggregate scale=none;
run;

/*******************************************************/
/*SAS Code second objective: Recommende number of steps*/

/* hierarchical clustering*/
ods graphics on;
proc cluster data=dataset1 method=ward ccc pseudo out=tree
plots=den(height=rsq);
var interstep daystep  days2step mestep;
id id;
run;
ods graphics off;

proc tree data=tree out=New nclusters=3 noprint;
height _rsq_;
copy interstep daystep  days2step mestep ;
id id;
run;

proc import datafile = 'C:\Uhasselt\MscThesis\dataset2.csv'
 out = dataset2
 dbms = CSV
 ;
run;
proc logistic data = dataset2;
class  catbmi(ref = "2") catga(ref = "1") admission (ref = "0") catparity (ref = "0")
catabort(ref = "0") cluster(ref = "3")/ param = ref;
model outcome(ref = "0") = cluster catbmi  catga  admission age
catparity/ link = glogit aggregate scale=none;
run;
```