TRCLC 17-04
August 30, 2019

# Monitoring Daily Activities and Linking Physical Activity Levels Attributed to Transportation Mobility Choices and Built Environment

## FINAL REPORT

**Jun-Seok Oh[1], Stephen P. Mattingly[2], Ala Al-Fuqaha[1], Kate Hyun[2], Sangwoo Lee[1], Raed Abdullah Hasan[1], Sina Famili[2], Hafez Irshaid[1], Md Mehedi Hasan[1], and Shirin Kamali Rad[2]**

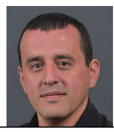
[1] Western Michigan University
[2] University of Texas at Arlington

Transportation Research Center *for Livable Communities*

Western Michigan University | University of Texas at Arlington | Utah State University | Wayne State University | Tennessee State University

<table>
<tr><td>1. Report No.:<br>TRCLC 17-4</td><td>2. Government Accession No.:<br>N/A</td><td colspan="2">3. Recipient's Catalog No.:<br>N/A</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle:<br>Monitoring Daily Activities and Linking Physical Activity Levels Attributed to Transportation Mobility Choices and Built Environment</td><td colspan="2">5. Report Date:<br>August 30, 2019</td></tr>
<tr><td colspan="2">6. Performing Organization Code:<br>N/A</td></tr>
<tr><td colspan="2">7. Author(s):<br>Jun-Seok Oh, Stephen P. Mattingly, Ala I Al-Fuqaha, Kate Hyun, Sangwoo Lee, Raed Abdullah Hasan, Sina Famili, Hafez Irshaid, Md Mehedi Hasan, Shirin Kamali Rad</td><td colspan="2">8. Performing Org. Report No.:<br>N/A</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address:<br>Western Michigan University<br>1903 West Michigan Avenue, Kalamazoo, MI 49008</td><td colspan="2">10. Work Unit No. (TRAIS):<br>N/A</td></tr>
<tr><td colspan="2">11. Contract No.:<br>TRCLC 17-4</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address:<br>Transportation Research Center for Livable Communities (TRCLC)<br>1903 W. Michigan Ave., Kalamazoo, MI 49008-5316</td><td colspan="2">13.Type of Report & Period<br>Final Report<br>[1/1/2018 - 08/30/2019]</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code:<br>N/A</td></tr>
<tr><td colspan="4">15. Supplementary Notes:</td></tr>
<tr><td colspan="4">16. Abstract<br>The relationship between transportation and health may play a significant role in improving the public's well-being due to physical activities and health benefits of active transportation. Travel behavior researchers need to investigate the relationship between transportation mode choices and human health by observing traveler behaviors and their effect on physical activity and public health. This research identifies and categorizes the health outcomes from daily physical activity and daily travel activities by employing wearable devices with sensing and GPS tracking technology. In this study, the research team develops an integrated data collection and processing platform named "PASTA" to monitor the participant's daily travel and physical activities. This platform automates data collection and integrates the big data processing of daily travel GPS trajectories from a mobile application designed by the research team with physical activity data from the Fitbit Charge 2/3. The study collects data from a total of 120 participants from Kalamazoo, MI and Arlington, TX for a 6-month period. The PASTA platform requires many different analyses to observe the transportation impact on health outcomes; these include user activity/trip recognition and transportation mode detection. The study also explores the association between physical activity and an individual's socioeconomic-body composition profile and proposes fusion into an integrated transportation and health impact model (ITHIM). Arlington participants use more active transportation (bicycle and walking) than Kalamazoo participants who use public transit more frequently. For activity/trip recognition, the study observes the highest accuracy for the combined Geohash-GIS approach with dwell time of 5 minutes. The Random Forest model outperforms the other machine learning models in terms of transportation mode detection. The participant's BMI, age, gender and baseline active travel time show a direct effect on the transportation-related physical activity level. This research provides a method to quantify the amount of physical activity and health benefits associated with transportation activity.</td></tr>
<tr><td colspan="2">17. Key Words:<br>Daily Travel Activities, Physical Activities, Transportation Mobility, Mobile Application, Machine Learning</td><td colspan="2">18. Distribution Statement:<br>No restrictions.</td></tr>
<tr><td>19. Security Classification - report<br>Unclassified</td><td>20. Security Classification - page<br>Unclassified</td><td>21. No. of Pages<br>172</td><td>22. Price:<br>N/A</td></tr>
</table>

# About the Authors

| | |
|---|---|
|  | **Jun-Seok Oh**<br>Professor in Civil and Construction Engineering, Western Michigan University<br>Director Research interests in TRCLC: Transportation systems modeling, non-motorized transportation, traffic safety, intelligent transportation systems, traffic simulation |
|  | **Stephen P. Mattingly**<br>Professor in Civil Engineering, University of Texas at Arlington |
|  | **Ala Al-Fuqaha**<br>Professor in Computer Science, Western Michigan University |
|  | **Kate Hyun**<br>Assistant Professor in Civil Engineering, University of Texas at Arlington |
|  | **Sangwoo Lee**<br>Assistant Professor of Exercise Science in Department of Human Performance and Health Education, Western Michigan University |
|  | **Raed Hasan**<br>Ph.D. candidate in Department of Civil and Construction Engineering, Western Michigan University |
|  | **Sina Famili**<br>Ph.D. candidate in Department of Civil Engineering, University of Texas at Arlington |
|  | **Hafez Irshaid**<br>Master’s student in Department of Computer Science (During the research period)<br>Software Engineer at General Motors (Later) |
|  | **Md Mehedi Hasan**<br>Ph.D. candidate in Department of Civil and Construction Engineering, Western Michigan University |
|  | **Shirin Kamali Rad**<br>Master’s student in Department of Civil Engineering, University of Texas at Arlington |

# Disclaimer

The contents of this report reflect the views of the authors, who are solely responsible for the facts and the accuracy of the information presented herein. This publication is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. This report does not necessarily reflect the official views or policies of the U.S. government, or the Transportation Research Center for Livable Communities, who assume no liability for the contents or use thereof. This report does not represent standards, specifications, or regulations.

# Acknowledgments

This research was funded by the US Department of Transportation through the Transportation Research Center for Livable Communities (TRCLC), a Tier 1 University Transportation Center at Western Michigan University.

# TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| PASTA | Physical Activity through Smart Travel Activity |
| R&D | Research and Development |
| OECD | Organization for Economic Co-operation and Development |
| TRB | Transportation Research Board |
| ODPHO | Office of Disease Prevention and Health Promotion |
| CDC | The Centers for Disease Control and Prevention |
| DB | Disease Burden |
| GBD | Global Burden of Disease |
| GHGE | Greenhouse Gas Emissions |
| GPS | Global Positioning System |
| HSIRB | Human Subjects Research Institutional Review Board |
| BMI | Body Mass Index |
| BFM | Body Fat Mass |
| PBF | Percent Body Fat |
| DALY | Disability-Adjusted Life Year |
| YLD | Years Lived with Disability |
| YLL | Years Life Lost |
| ITHIM | Integrated Transportation and Health Impact Model |
| PA | Physical Activity |
| RR | Relative Risk |
| RRR | Relative Risk Ratio |
| WHO | World Health Organization |
| MET | Metabolic Equivalents Task |
| APF | Attributable Proportion Fraction |
| CRA | Comparative Risk Assessment |
| PAF | Population Attributable Fraction |
| RF | Random Forest |

| | |
|---|---|
| AdaBoost | Adaptive Boosting |
| SVM | Support Vector Machine |
| ANN | Artificial Neural Network |
| OR | Odds Ratio |
| SEs | Standard Error |
| ML | Multinomial Logit |
| ANOVA | Analysis of Variance |
| AUC | Area Under the Curve |
| ROC | A receiver Operating Characteristic Curve |
| TPR | True Positive Rate<br>Also known as sensitivity, recall or probability of detection |
| FPR | False Positive Rate<br>Also known as the fall-out or probability of false alarm |
| PM | Particulate Matter |
| NOx | Nitrogen Oxides |
| %HRR | Percentage of Heart Rate Reserve |
| $HR_{act}$ | Heart Rate during Activity |
| $HR_{rest}$ | Heart Rate at Rest |
| HRmax | Maximum Heart Rate |
| PAI | Physical Activity intensity |
| API | Application programming interface |

# Executive Summary

Transportation decisions impact human health at least in three ways, such as traffic crashes, environmental impact, and physical activity. While ample efforts to reduce traffic crashes and environmental impacts exist, less attention has been paid to transportation decisions and their impacts on physical fitness. Recent efforts on the relationship between transportation and physical fitness mostly focus on active transportation. The potential benefits of active transportation, including savings in mobility costs, benefits from related businesses, community savings in costs, etc., are directly and indirectly associated with health and environmental benefits. Although predicting how a particular transportation planning decision affects physical fitness remains difficult, the total impacts appear likely to be large. Diseases associated with inadequate physical fitness cause an order of magnitude of illness and deaths. Even modest reductions in these illnesses could provide significant health benefits. Therefore, a strong need for investigating how transportation options affect the physical activity and public health seems clear. This study assesses the factors impacting the amount of physical activity and the proportion of an individual's daily activity attributable to transportation activities.

**Literature Review:**

The literature review investigates health impact and outcome modelling, transportation data collection and user activity recognition, travel mode detection and associated characteristics, and integrated health-transportation impacts. Physical inactivity represents one of the most serious public health concerns facing health organizations worldwide. According to the World Health Organization (WHO), physical inactivity accounts for 6% of global deaths, which ranks it as the fourth leading cause of global mortality. In addition, physical inactivity contributes to 21-25% of breast and colon cancers, 27% of diabetes, and 30% of ischemic heart disease (WHO, 2019). The 2013 world-wide direct and indirect healthcare costs of physical inactivity reached $53.8 billion. Physical activity through active transportation represents a potential solution to physical inactivity and these serious healthcare issues and costs. Developed countries, including the United States of America, continue to investigate the connections between transportation and public health and the environment.

When measuring and monitoring the effects of the transportation system on health, many automatic data collection applications offer more promise than traditional travel diary options. Many studies for analyzing daily transportation activities and behaviors applied Smartphone and GPS technology to detect the spatial patterns of road users. This method required many researchers to use machine learning and classification algorithms. More recent research applied Geohash clustering to identify the threshold of change for the activities of road users, which represents a significant improvement over any stage-based technique.

For the detection of transportation mode, many studies applied Machine Learning methods either through the use of neural networks, deep learning, or other unsupervised techniques. Integrated data about the travel behavior and physical activities of individuals offers the ability to determine levels of physical activities (PA) associated with transportation.

**<u>Development of Mobile Application and Integrated "PASTA" Platform:</u>**

The study developed an integrated platform named "PASTA" for automated data collection and processing from a mobile application and Fitbit Charge-2. Data collected from the mobile application included the GPS locations of participants' daily trips and activities, while the Fitbit charge-2 collected data regarding physical activities, such as walking/running steps, heart rate, total calories, active minutes.  The research team processed the minute-by-minute data log to determine daily activity (type, location, and duration), travel activity (transportation mode, travel time), and physical activity (e.g. heart rates and duration).

"PASTA" platform included four main components, where the first component was a mobile phone application, which collected data from the phone itself and the smartwatch, and included the following features: the user authentication process, the activity/trip verification, and the location information. The team implemented the mobile application using Ionic 2 cross-platform technology for collecting travel activities data. The server, which was implemented using Spring Framework, received these data via RESTful API and stored the information in the database for processing. The server analyzed the data by identifying from the GPS points the visited locations and the trips between these locations.

The back-end server represented the second component and it received the smartphone data (e.g.  location information, user authentication) and pulled the physical activity (PA) data from the smartwatch. The database management system represented the third elementand handled storage and data recovery. The classifiers, the fourth component, processed raw data within the database and extracted the required information. The research platform included a collection of classifiers; the classification of activities as a trip or non-trip represented the most important classifier. The framework for this classifier used a combined Geohash-GIS technology that clustered the locational activity points.

The integrated approach allowed classification of various activity types and transportation modes. The amount of physical activity by activity type and travel mode was quantified by using the heart rate and the activity duration. Figure (E.1) depicted the flowchart of the "PASTA" platform.

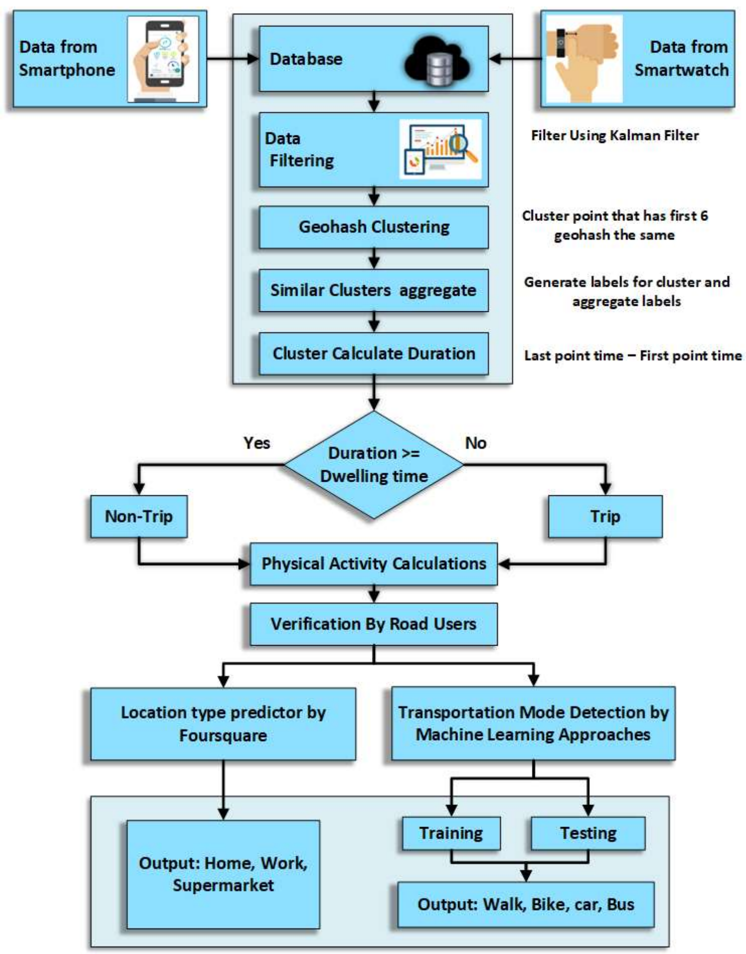



**Figure E.1 The diagram of the data processing through "PASTA" platform**

**Survey Result:**

The study selected a total of 120 participants to collect daily activity and physical activity data in Kalamazoo, Michigan and Arlington, Texas. Each participant received a smartwatch (Fitbit charge 2/3) and installed the PASTA mobile app that tracked the activities of the user in the background and sent these data to the server. The team recorded In-body composition data for each of the participants.

The research team developed an initial survey for pre-screening that included demographic questions, transportation mode choice, and daily commuting travel time. The initial survey showed that those physically active individuals tended to choose active transportation options. Based on the initial survey, the research team classified the participants for a stratified recruitment based on auto use (inactive), auto use (active), and all other modes. All study participants completed the main survey to assess their socio-demographic profile, economic profile, travel options, daily activities. The survey responses indicated that over 70% of the participants perceived that they have good or excellent health. The research team computed the subject's physical activity based on intensity level (from 1 to 10), and duration (minutes) of their activities. Based on the physical activity, both study areas (Kalamazoo vs. Arlington) showed a similar distribution of active and highly active participants (45.8% vs. 44.8%). Based on the body mass index (BMI), the obesity rate remained higher at WMU than UTA (33.9% vs. 25.9%).

The research team combined the transportation modes into three groups and aggregated the travel time by mode, such as active travel time (summation of travel times for walking and biking), private vehicle travel time (summation of travel times for driving, being a passenger of auto, and motorcycle time), and transit travel time (summation of travel times for wait/transfer, bus, rail, and taxi). More than 50% of the participants used private vehicles in terms of commuting for both study areas. For Arlington, the second highest transportation mode was active transportation (38% participants). Public transit appeared more popular for commuting in Kalamazoo in comparison to Arlington because the UTA shuttle bus represented the only fixed-route transit option in Arlington, whereas Kalamazoo transit system instead provided fixed-route transit service to Kalamazoo area cities including Kalamazoo, Portage, Parchment, Texas, and Osthemo. In terms of relationship between transportation mode users and physical activity levels, Arlington participants appeared more physically active in comparison to the Kalamazoo participants because they frequently used bicycle and walking as their transportation mode.

The researchers conducted a cross tabulation analysis to compare the respondent's perceived health status and their actual physical activity. The overall results of the UTA sample showed that perceived health did not always align with objective health measures such as BMI and physical activity level. Although, more Kalamazoo participants belonged to the normal weight with excellent health category than Arlington (10.2% vs. 1.7%), the relationship between perceived health and BMI seemed similar at both study areas.

**<u>Transportation User Activity and Trip Recognition:</u>**

In this study, the research team developed three different approaches to identify user activity and trip based on the GPS trajectories of the participants. The research team applied different thresholds of spatiotemporal change by developing a Geohash clustering approach and the GIS-based approach, forming a combined approach by integrating the Geohash and GIS systems. The team developed and implemented approaches for activity only, trip only, and sequential activity-trip recognition with GPS data from Kalamazoo, Michigan.

| Different testing Scenarios | MAPE | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Geohash-6 | | | GIS-based Approach | | | Combined Approach | | |
| | 5 min | 8 min | 10 min | 5 min | 8 min | 10 min | 5 min | 8 min | 10 min |
| activity/trip | 30.16 | 28.87 | 33.15 | 65.91 | 62.98 | 67.11 | 12.70 | 13.48 | 18.81 |
| activity | 23.81 | 22.82 | 21.22 | 49.72 | 58.59 | 58.29 | 18.65 | 20.95 | 27.83 |
| trip | 42.77 | 40.97 | 51.43 | 44.83 | 34.88 | 46.35 | 15.00 | 20.01 | 25.21 |

**Figure E.2 Prediction accuracy in activity/trip recognition based on MAPE**

For the Geohash clustering approach, Geohash precision level-6 with a dwell time of 5 minutes provided better activity/trip recognition accuracy [Figure (E.2)]. For the GIS-based approach, a dwell time of 10 minutes worked better than 5 minutes in terms of accurately recognizing the activity and trip of the participants. Among the three approaches, the Combined Geohash-GIS approach with dwell time of 5 minutes provided the best accuracy, which could significantly enhance the efficiency and accuracy of a GPS travel survey by correctly (about 88%)

recognizing user activity and trip patterns. This proposed combined approach could serve as a foundation for a future system of full-scale travel information identification with GPS data.

**Transportation Mode Detection:**

Transportation mode identification represents another critical output to allow GPS systems to replace travel diaries. Therefore, the research team developed algorithms for predicting transportation modes using the smartphone and smartwatch data with machine learning techniques. This study used four machine learning methods (Extreme Gradient Boosting, Random Forest, Support Vector Machine, and Artificial Neural Network) for mode prediction after introducing physical activities as a feature not used in previous studies. Figure (E.3) showed the accuracy of the predictive methods considered in this study.

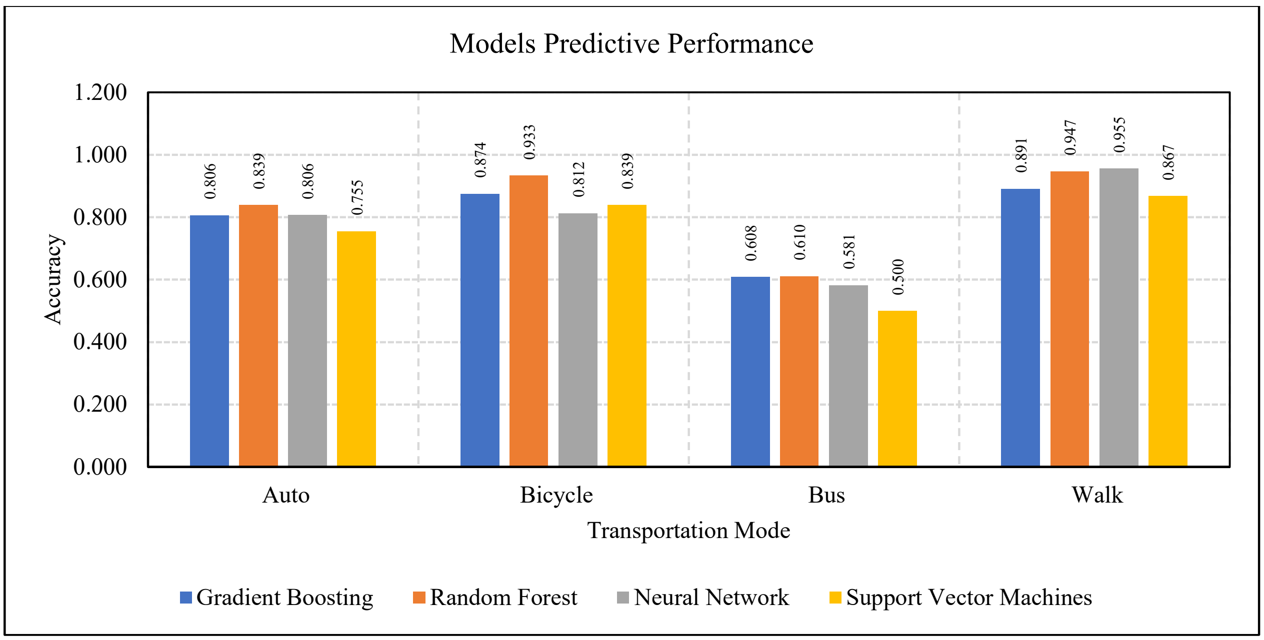



**Figure E.3 Predictive models performances in detecting the transportation mode**

The results showed that the Random Forest method worked better than other methods in detecting non-motorized modes: walking mode (97.2%) and bicycle mode (90.6%). However, the random forest model performed the most poorly for classifying the bus mode (61.4%) because of misclassification cases involved instances where a bus mode was incorrectly classified as an auto mode and vice versa. This study provided the appropriate tools to properly classify most modes, but other features may need to be included to more accurately classify bus.

**Exploring the Association Between Individual's Physical Activity Level with Socio-economic and Body-composition Characteristics:**

In this study, the research team explored the statistical association between the physical activity levels of individuals, their socioeconomic characteristics, and body composition profiles. The physical activity intensity attained from transportation was measured by collecting individual's daily activity, transportation choices, and the amount of physical activity. The in-body composition characteristics included the Body Mass Index (BMI), Body Fat Mass (BFM), and Percent Body Fat (PBF). In this section, the research team conducted a Path analysis, which represents a special case of structural equation modeling (SEM) to explore the causal and non-causal relationships among variables.

From the result of the path analysis [Table (E.1)], partcipant's residing state, body mass index (BMI), number of vehicles per household, and total time spent in different transportation mode showed a direct effect on the weekly equivalent vigorous PA minutes from transportation. On the other hand, the variables with an indirect effect on an individual's physical activity included age and gender. Race and annual income contributed both direct and indirect effects on an individual's physical activity.

## Table E.1 Summary of the significant variables (total, direct, indirect) related to physical activities

| Variables | Total | | | | Direct Effects | | | | Indirect Effects | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coef. | Std. Err. | z | P>z | Coef. | Std. Err. | z | P>z | Coef. | Std. Err. | z | P>z |
| State =Michigan | | | | | | | | | | | | |
| Gender=Male | 0.285 | 0.024 | 12.010 | **0.000** | 0.285 | 0.024 | 12.010 | **0.000** | 0 | | (no path) | |
| Age Group =26-49 | 0.188 | 0.024 | 8.000 | **0.000** | 0.188 | 0.024 | 8.000 | **0.000** | 0 | | (no path) | |
| Race=Black | 0.133 | 0.039 | 3.430 | **0.001** | 0.133 | 0.039 | 3.430 | **0.001** | 0 | | (no path) | |
| Race=Asian | 0.327 | 0.042 | 7.830 | **0.000** | 0.327 | 0.042 | 7.830 | **0.000** | 0 | | (no path) | |
| Race=Hispanic | 0.255 | 0.028 | 9.160 | **0.000** | 0.255 | 0.028 | 9.160 | **0.000** | 0 | | (no path) | |
| Age Group=50-64 | -0.144 | 0.045 | -3.170 | **0.002** | -0.144 | 0.045 | -3.170 | **0.002** | 0 | | (no path) | |
| Constant | 0.059 | 0.030 | 2.000 | **0.045** | | | | | | | | |
| Trip-equivalent=PA min | | | | | | | | | | | | |
| state=Michigan | 4.038 | 1.483 | 2.720 | **0.006** | 4.038 | 1.483 | 2.720 | **0.006** | 0 | | (no path) | |
| BMI-test | 0.573 | 0.145 | 3.960 | **0.000** | 0.573 | 0.145 | 3.960 | **0.000** | 0 | | (no path) | |
| Vehicles No.=2 | 9.929 | 2.501 | 3.970 | **0.000** | 9.929 | 2.501 | 3.970 | **0.000** | 0 | | (no path) | |
| Vehicles No.=1 | 10.383 | 2.203 | 4.710 | **0.000** | 10.383 | 2.203 | 4.710 | **0.000** | 0 | | (no path) | |
| PBF | | | | | 0 | | (no path) | | 0.092 | 0.024 | 3.900 | **0.000** |
| Total time | 0.040 | 0.017 | 2.300 | **0.022** | 0.040 | 0.017 | 2.300 | **0.022** | 0 | | (no path) | |
| Gender=Male | | | | | 0 | | (no path) | | 1.150 | 0.433 | 2.660 | **0.008** |
| Age Group=26-49 | | | | | 0 | | (no path) | | 0.759 | 0.295 | 2.580 | **0.010** |
| Race=Black | | | | | 0 | | (no path) | | 0.536 | 0.251 | 2.130 | **0.033** |
| Race=Asian | -6.751 | 2.680 | -2.520 | **0.012** | -6.751 | 2.680 | -2.520 | **0.012** | 1.319 | 0.513 | 2.570 | **0.010** |
| Race=Hispanic | -3.371 | 1.731 | -1.950 | **0.051** | -3.371 | 1.731 | -1.950 | **0.051** | 1.031 | 0.395 | 2.610 | **0.009** |
| Vehicles No.=3+ | 8.543 | 2.574 | 3.320 | **0.001** | 8.543 | 2.574 | 3.320 | **0.001** | 0.000 | | (no path) | |
| Income=30000-50000 | -10.815 | 2.194 | -4.930 | **0.000** | -10.815 | 2.194 | -4.930 | **0.000** | 2.108 | 0.808 | 2.610 | **0.009** |
| Health Condition=Bad | -11.386 | 3.846 | -2.960 | **0.003** | -11.386 | 3.846 | -2.960 | **0.003** | 0.000 | | (no path) | |
| Professional Status: Administration | 7.937 | 2.844 | 2.790 | **0.005** | 7.937 | 2.844 | 2.790 | **0.005** | 0.000 | | (no path) | |
| Age Group =50-64 | | | | | 0 | | (no path) | | -0.581 | 0.281 | -2.070 | **0.039** |
| Constant | -8.704 | 4.200 | -2.070 | 0.038 | | | | | | | | |
| BMI Test | | | | | | | | | | | | |
| PBF | 0.161 | 0.007 | 22.760 | **0.000** | 0.161 | 0.007 | 22.760 | **0.000** | 0.000 | | (no path) | |
| Constant | 23.232 | 0.242 | 95.810 | 0.000 | | | | | | | | |
| Vehicles No.=2 | | | | | | | | | | | | |
| Income=30000-50000 | 0.357 | 0.024 | 14.870 | **0.000** | 0.357 | 0.024 | 14.870 | **0.000** | 0.000 | | (no path) | |
| Constant | 0.189 | 0.011 | 17.310 | 0.000 | | | | | | | | |
| Vehicles No.=1 | | | | | | | | | | | | |
| Income=30000-50000 | -0.138 | 0.028 | -4.880 | **0.000** | -0.138 | 0.028 | -4.880 | **0.000** | 0.000 | | (no path) | |
| Constant | 0.452 | 0.013 | 35.060 | 0.000 | | | | | | | | |
| var (State =Michigan) | 0.213 | 0.007 | | | | | | | | | | |
| var(trip equivalent =PA min) | 850.939 | 28.154 | | | | | | | | | | |
| var (BMI Test) | 26.821 | 0.887 | | | | | | | | | | |
| var (Vehicles No.=2) | 0.173 | 0.006 | | | | | | | | | | |
| var (Vehicles No.=1) | 0.241 | 0.008 | | | | | | | | | | |

**Integrated Transportation and Health Impacts Model (ITHIM) with a new approach to measuring the relative risk of physical activity related and non-related to travel**

In this study, the research team employed the integrated transportation and health impact model (ITHIM) and enhanced the concept of ITHIM by adding the quantitative data obtained from PASTA. Mechanisms for collecting data on physical activities (PA) related to or not related to transportation in previous studies relied on questionnaires and interviews of specific samples in the community. Instead, the PASTA platform provided an automated mechanism for gathering the daily activities of people, especially with regard to physical activities and travel behaviors. The study introduced the physical activity minute (PAM) (an indicator of changes in activity levels), which represented an alternative to the metabolic equivalence of the task (METs).

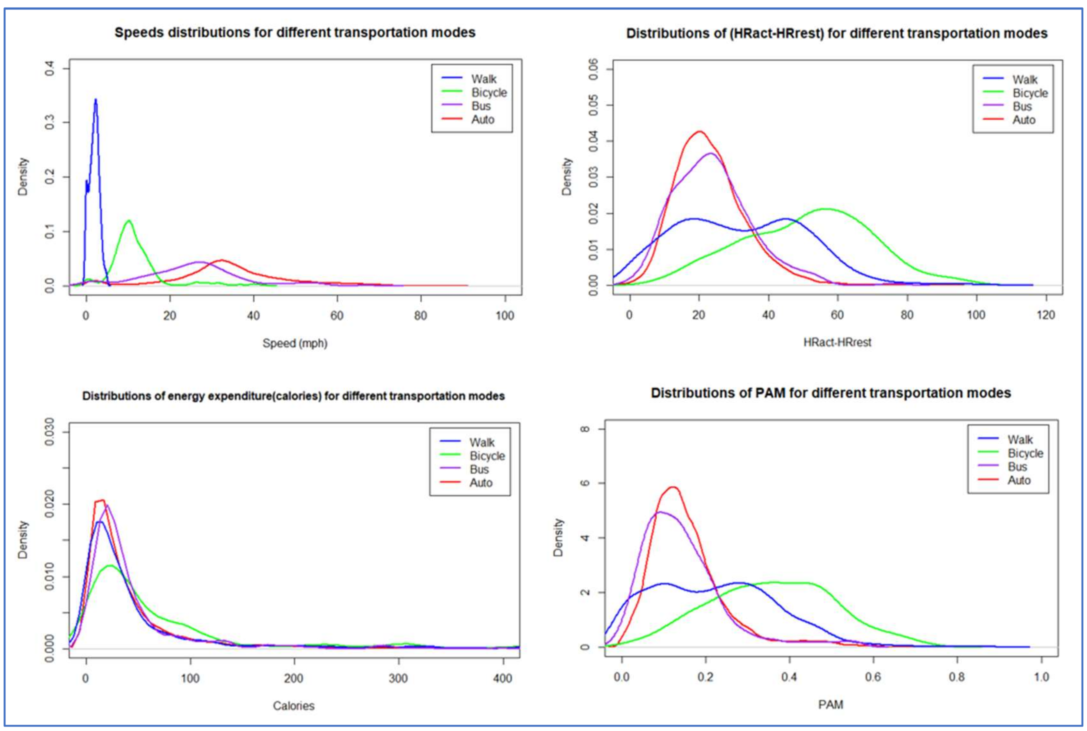



**Figure E.4 Shifting from qualitative data to quantitative data for the Physical Activities impact factor**

Figure (E.4) showed the results obtained from the ITHIM model, where the role of transportation can be observed in different levels of physical activity depending on the transportation mode. The outcome explored the speed, the difference between heart rate activity and heart rate rest for each activity (HRact- HRrest), and calories, in each transportation mode, respectively. Results showed an increasing trend in physical activity with cycling, less physical activity when walking, and then a further decrease when using vehicles. Also, the research team verified a relationship between METs calculated from calorie and the PAM values calculated from the heart rate (both of the data derived from the smartwatches) for both physical activities (PA) related to or not related to transportation. By coupling the data from PASTA, this research substantially reduced the limitations in the previous ITHIM and upgraded the model to conform to the current technological advances.

# Chapter 1 Introduction

## 1.1 Background and Research Problem

Humans require physical activity for healthy lives. Increases in physical and cardiovascular activities tend to decrease diseases (Adesiyun, 2018). Non-motorized transportation options, like walking, running and cycling, deliver natural health benefits and provide one option for increasing physical activity. The potential benefits of active transportation include savings in mobility costs, benefits from related businesses, community savings in costs associated with health and environmental benefits (Cadilhac, 2011). In addition, active transportation provides environmental benefits, enhance road safety by reducing the loss of life, improve traffic efficiency by reducing congestion, contribute to mood improvement of road users and economic benefits derived from many sources (Frank et al., 2006; Leslie et al., 2007; Litman, 2003; ALRMTAT, 2016). Transportation mode and path decisions impact human health in at least three ways: traffic crashes, environmental impacts, and physical activity. While ample efforts in reducing traffic crashes and environmental impacts exist, less attention has been paid to the mode choice impact on physical fitness.

The health benefit of active transportation comes from the participants' physical activities (Ahima and Lazar, 2013). Research related to the health benefits of active transportation represents an important research topic because using active transportation to gain health benefits represents a strategy for integrating physical activity into a daily routine. Studies show that persons with moderate to high levels of physical activity or cardiorespiratory fitness have a lower mortality rate rather than those with sedentary habits or low cardiorespiratory fitness (Wagner et al., 2002). Furthermore, a significant trend in decreasing the risk of death across increasing categories of distance walked, flights of stairs climbed, and degree of intensity of sports played exists (Bouchard et al., 2007). Physical activities for cardiorespiratory endurance reduces the risk of developing or dying from cardiovascular diseases (CVD), hypertension, colon cancer, and non-insulin-dependent diabetes mellitus (NIDDM) and improves mental health while endurance-type physical activity may reduce the risk of developing obesity, osteoporosis, and depression and may improve psychological well-being and quality of life (Gordon-Larsen et al. 2009). According to a report by the Centers for Disease Control and Prevention (CDC), benefits of physical activities include: 1) to help build and maintain healthy bones, muscles, and joints; 2) help control weight, build lean muscle, and reduce fat; and 3) to prevents or delay the development of high blood pressure and helps reduce blood pressure in some adolescents with hypertension (CDC, 2017). Active transportation's magnitude of impact on physical fitness or physical activity requirements remains unknown.

Assessing this impact appears problematic because measuring physical activity directly remains difficult. Three types of physical activity measures have been used in observational studies over the last 40 years. Most studies rely on a self- reported level of physical activity, as recalled by people prompted by a questionnaire or interview (Hatano, 1993). Cardiorespiratory fitness (also referred to as cardiorespiratory endurance), which is measured by aerobic power, represents a more objectively measured characteristic. Some studies use an occupation to classify people according to their likely physical activity at work (Barengo et al., 2004). Although predicting the impact of a particular transportation planning decision on physical fitness appears challenging, the total impacts seem likely to be large. Diseases associated with inadequate physical fitness cause an order of magnitude of more deaths, and cause more deaths than road crashes (HHS, 2018). Even modest reductions in these illnesses could provide significant health benefits; therefore, investigating the impact of transportation mode choices on physical fitness or physical activity levels appears imperative.

Classic clinical interventions promoting a healthy lifestyle primarily rely on counseling, but this requires significant time investments. A structured exercise program based on individual characteristics remains complex and costly to maintain. Wearable devices offer an alternative approach to improve individual health and support the collection and interpretation of data on environments, behaviors, physiology, and well-being. Recent clinical intervention among adults using wearable ECG, a smartphone for real-time data transfer on aerobics sessions, and a Global Positioning System (GPS) have dramatically improved walking distance and depression (Brichetto *et al*., 2019). Improvements to devices, communication devices, and data analysis tools, make the cheap, reliable, wearable sensors available to transport researchers. These devices provide solutions for researchers and multiple streams and allow for new forms of research (Machek, *et* *al.*, 2016). Using the new data collection devices, researchers may be able to characterize the physical activity associated with active transportation choices.

This study intends to identify and categorize the health outcomes associated with different transportation modes. By employing wearable devices with sensing and GPS tracking technology, the research team can collect the amount of physical and cardiovascular activity for different travel activities and transportation mode choices. The study monitors and documents humans' activities to investigate their travel trajectories. Continuous monitoring and data generation provides raw data that can be used by researchers to determine the common physical activity benefits of transportation modes. This research seeks to integrate human health into the transportation planning process by quantifying the health outcomes associated with transportation modal options.

## 1.2 Research Goal and Objectives

The primary goal of the research was to explore the factors impacting the amount of physical activity an individual engages in and the proportion of an individual's daily activity attributable to transportation activities. Specific research objectives include:

1. To develop a strategy for monitoring and recording the daily physical activity of a representative sample of individuals in different urban area contexts.
2. To develop a mobile application for automatically monitoring and recognizing the user's transportation activity and daily travel trajectories.
3. To develop a data fusion strategy to combine wearable data (including heart rate) with smartphone data and Google Map features/data.
4. To create a preliminary scheme for using speed patterns/profiles to classify mode choice.
5. To test the statistical association between the physical activity levels of individuals, their socioeconomic and employment profiles, and their associated health outcomes by using different transportation modes.
6. To identify and categorize health outcomes from daily physical activity.
7. To use the fused data to classify and measure the physical activity for evaluating integrated transportation and health impacts.

## 1.3 Research Scope and Overview

Successful completion of this research should help incorporate human health into the transportation planning process by quantifying the health outcomes associated with transportation mode choices.

This report includes the following components: a review of previous relevant studies, development and implementation of a mobile application for travel trajectory data collection, development of an integrated platform named "PASTA" for combined data collection and storage from mobile application in accordance with physical activity monitoring through Fitbit Charge 2/3, comprehensive data analysis for activity/trip recognition, transportation mode choice analysis, and integrated transportation and health impact model assessment. The researchers accomplished these objectives during the following tasks:

Task 1: Literature Review

Task 2: Mobile Application Development

Task 3: Data Collection and Integrated Platform Development

Task 4: Development of User Activity and Trip Recognition

Task 5: Transportation Mode Detection

Task 6: Exploring Relationship Between Physical Activity and Individual's Characteristics

Task 7: Integrated Transportation and Health Impacts Model (ITHIM) Development

Task 8: Recommendations and Final Report

Figure 1.1 depicts the connectivity of the eight tasks and the overall flow of this research.

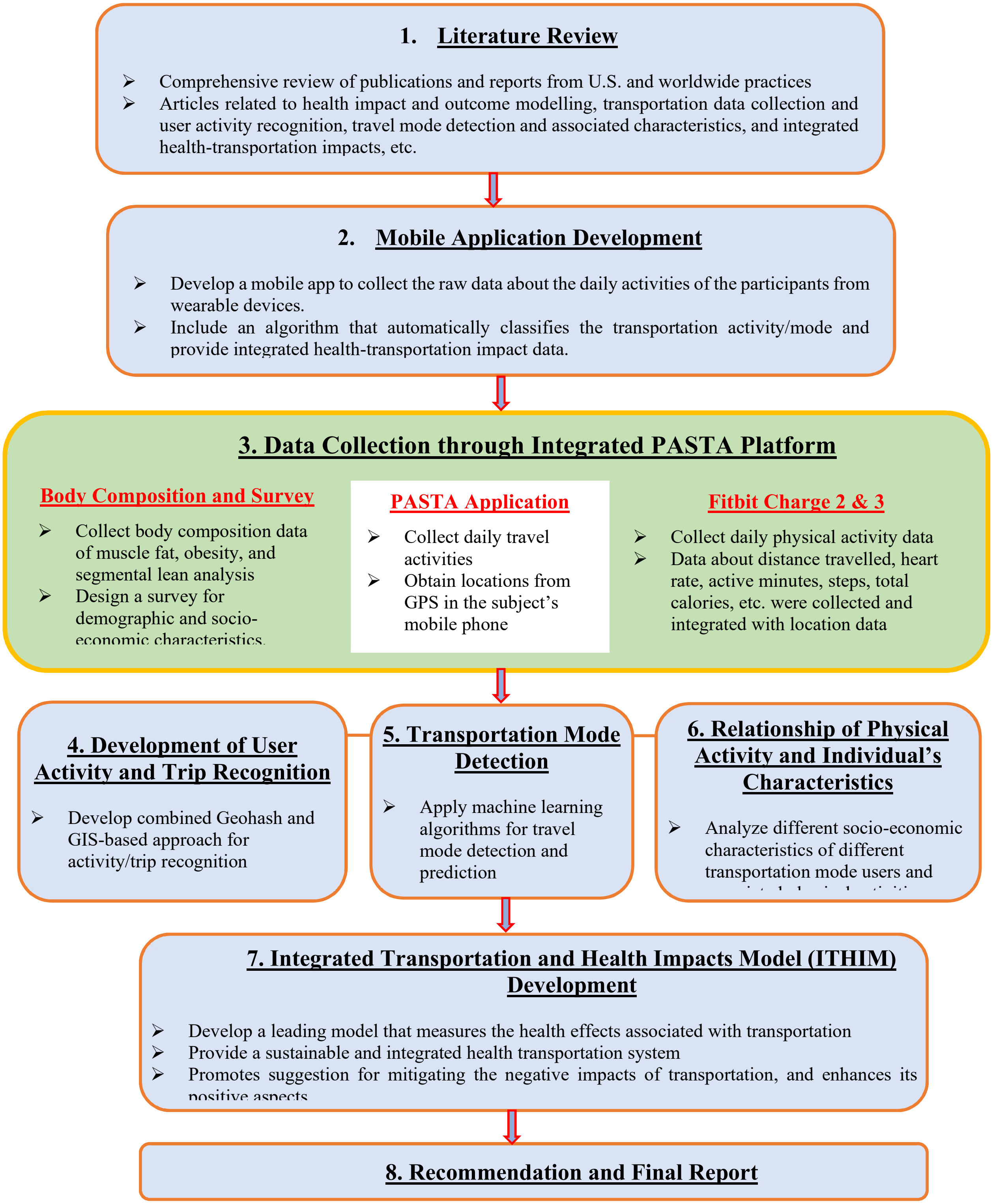



**Figure 1.1 Overall Flow of this Research**

# Chapter 2 Literature Review

## 2.1 Introduction

Physical activity represents a widely considered solution for many health problems and to remain healthier as an older adult. Without increases in physical activity across the broader population, the rates of disease, and the decline in life expectancy will continue to increase and with these trends the cost of health care will also increase. The active transportation system represents an opportunity for an individual to increase physical activity and create a more livable community. This section presents a comprehensive literature review. The review focuses on models that analyze the health impact and outcomes from physical activity, transportation data collection and user activity recognition studies, travel mode detection using GPS data, and integrated health-transportation impacts.

## 2.2 Health Impacts and Outcomes Modeling

According to the Centers for Disease Control (CDC), chronic diseases, including heart disease, cancer, stroke, and diabetes, account for seven out of ten deaths among Americans and represent 75% of the $2 trillion U.S. healthcare spending in 2005 (Tinker, 2017). Although physical exercise as part of one's daily routine, such as walking and biking to mandatory and maintenance activities may easily increase physical activity, most areas lack infrastructure that encourages active transportation and require long commuting travel distances. This leads to private vehicle use for most trips/activities. The sedentary nature of many working environments, a lack of motivation and perceived future body pain represent the main primary barriers for individuals to engage in physical activity (Netz, 2018). A solution to the physical inactivity crisis will decrease nationwide healthcare costs and improve morbidity and mortality across all age groups while enhancing quality of life.

According to Chapman (2019), if someone walks 30 minutes daily for four days a week, he/she will burn an extra 20,000 to 40,000 calories per year, which is equal to a six to twelve-pound weight loss with the same food diet. According to the 10,000-step goal that originated in

Japanese walking clubs[1], 10,000 steps per day results in burning 300-400 kcal energy although the impact depends on walking speed and body composition (Hatano, 1993). According to the CDC (1996), the recommended minimum physical activity (equivalent to 150kcal) can be achieved with 30-minutes of daily moderate intensity physical activity. The study of Tudor-Locke & Bassett (2004) suggests the following thresholds to distinguish different healthy adults based on their physical activity:

- ‘Sedentary lifestyle’: < 5,000 steps/day
- ‘Low active’: 5,000-7,499 steps/day, which is a typical for daily activity excluding volitional sports/exercise
- ‘Somewhat active’: 7,500-9,999 steps/day, including some volitional activities and or elevated occupational activity demands
- ‘Active’: 10,000 steps/day, which is a threshold for classifying active versus non-active people.
- ‘Highly active’: are the people who travel more than 12,500 steps/day.

The Department of Health and Human Services (HHS) (2018), indicates about half of the U.S. adult population has one or more preventable chronic diseases. Regular physical activity can have a positive impact on seven out of ten of the most prevalent chronic diseases. Although half of American adults meet the key guidelines for aerobic physical activity, 80% of them do not meet the guidelines of both aerobic and muscle-strengthening activities. Physical activity contributes $117 billion in annual health care costs and about 10% to premature mortality rates (HHS, 2018). Despite some improvements in the physical activity levels of American adults, only 26% of men, 19% of women, and 20% of adolescents’ report activity levels that meet relevant CDC aerobic and muscle-strengthening guidelines [Figure (2.1)].

[1] In the 1960s, pedometers in Japan were sold under the trade name of “manpo-kei”, meaning “10,000 steps meter”

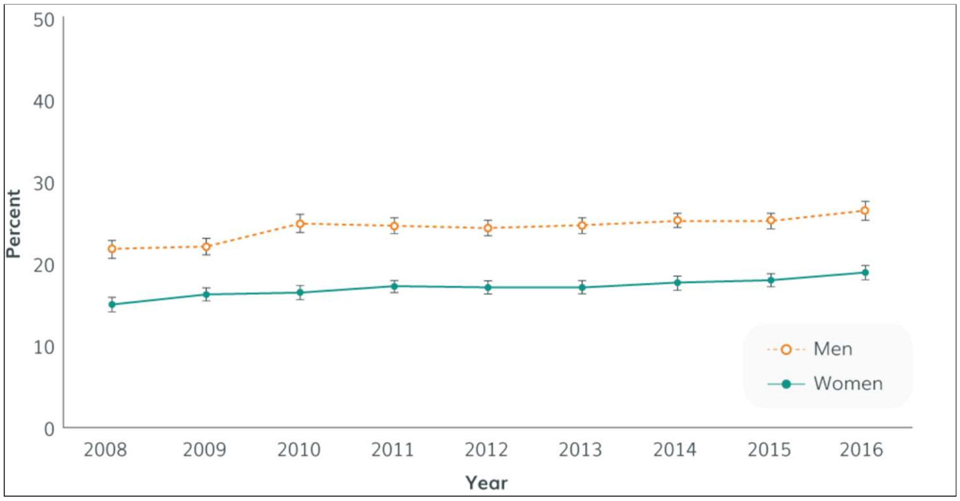



**Figure 2.1 Percentage of U.S. Adults Ages 18 Years or Older Who Met the Aerobic and Muscle-Strengthening Guidelines, 2008–2016 (from Centers for Disease Control and Prevention, National Center for Health Statistics, National Health Interview Survey (NHIS).**

### 2.2.1 Physical Activity and Major Health Outcomes

Significant research regarding the impact of physical activity on health outcomes exists. While some studies examine this linkage for one disease or risk factor, other studies address the association between physical activity and more than one particular disease. The authors the following diseases and risk factors due to their significance in previous studies:

- Obesity
- Diabetes
- Cardiovascular diseases
- Cancer
- Bone health/Osteoporosis

#### *2.2.1.1 Physical Activity and Obesity*

According to the CDC, in 2015-2016, the prevalence of obesity reached 39.8 % in U.S. adults, aged 20 and over, and 20.6% among adolescents, aged 12-19 years (CDC, 2017). Since physical activity provides between 25% and 50% of total daily energy expenditure, it contributes greatly to weight control (Bouchard et al., 2007). Koh-Banerjee et al. (2003) conducts a cohort study to examine the 9-year waist circumference change and changes in some interventions (e.g.

physical activity, dietary intake, alcohol consumption, and smoking) among U.S. men. The results linked an increase of 25 MET-hour/week in vigorous physical activity and more than 0.5 hour/week in weight training to 0.38 cm and 0.91 cm changes in waist circumference, respectively. The high occurrence of obesity and the linkage between physical activity and obesity make the encouragement of greater physical activity a critical public health goal.

Active commuting allows individuals to integrate physical activity into an otherwise sedentary lifestyle. A well-established body of literature examines the positive association between active (walk/bike) modes and health measures, in which include lower rates of obesity (Samimi et al., 2009; Humphreys et al., 2013; Flint et al., 2014; Scheepers et al., 2015; Flint and Cummins, 2016). Men who commute by car appear about 40% more likely to be overweight and obese than those who cycle or use public transit (Wen et al., 2008). Gordon-Larsen et al. (2009) confirm that men using active travel choices have a reduced likelihood of obesity (OR = 0.50; 95% CI 0.33, 0.76). According to Langerudi et al. (2015), a 1% increase in transit use reduces the likelihoods of obesity and heart attack by 1.10% and 1.20%, respectively. She et al. (2019) show that a 1% increase in public transit riders appears to reduce the county population obesity rates by 0.473%. Active transportation and public transit use play a significant role in reducing obesity rates.

#### *2.2.1.2 Physical Activity and Diabetes*

Diabetes causes blood glucose (sugar) levels to rise higher than normal, and Type 2 diabetes represents the most common type. The American Diabetes Association shows that 9.4% (30.3 million) of USA population had diabetes in 2015 (CDC, 2017). Based on many studies (Kriska et al., 2003; Bonora et al., 2004; Jonker et al., 2006; Onat et al., 2007; Fretts et al., 2009; Walker et al., 2010; Reis et al., 2011; Fan et al., 2015), an inverse relationship between physical activity and the risk of Type 2 diabetes exists; however, other research projects (Waki et al., 2005; Burke et al., 2007; Sato et al., 2007; Tonstad et al., 2013) find no association. The linkage between physical activity and Type 2 diabetes appears less clear than the link with obesity; however, it still represents an important relationship.

Many studies investigate the relationship between Type 2 diabetes and physical activity. Hu et al. (2003) examines the associations between occupational, commuting, and leisure-time physical activity and Type 2 diabetes risk and determines the relative risks (RRs) of Type 2 diabetes for light (1.00), moderate (0.70) and vigorous (0.74) work physical activity. Longer active commuting significantly reduces the RR where no active commute, 1-29 minutes of active commute, and more than 30 minutes of active commute generate 1.00, 0.96, and 0.64 relative risk values. The benefits of leisure time physical activity appear quickly with RR values of 1.00, 0.67, and 0.61 for low, moderate, and high levels. Aune et al. (2015) demonstrate that moderate (RR = 0.68) and vigorous (RR = 0.61) physical activity reduce the risk of Type 2 diabetes. Kyu et al. (2016) show major gains (19% reduction) in the risk of five diseases including diabetes with an increase of physical activity from 600 to 3600-MET minutes/week. Tajalli and Hajbabaie (2017) indicate that the change of transportation mode from car to walk reduces obesity by 5.5%, high blood pressure by 16.9%, and diabetes by 4.4%. The relationship between physical activity and Type 2 diabetes appears largely binary where physical activity must exceed some threshold for most of the risk reduction to occur.

#### *2.2.1.3 Physical Activity and Cardiovascular Disease Risk*

Based on the American Heart Association (AHA) (2018), cardiovascular diseases (CVD) account for nearly 836,546 deaths in the U.S., which represents about one out of every three deaths. Many studies investigate the positive association between walking/biking and CVD risk. Many studies determine the positive effects of leisure time and occupational physical activity on reducing CVD risk and mortality (Wagner et al., 2002; Barengo et al., 2004; Hu et al., 2005; Wennberg et al., 2006; Hu et al., 2007; Hamer and Chida, 2008; Gordon-Larsen et al., 2009). The risk factors include:

- systolic blood pressure (Barengo et al., 2004; Hu et al., 2007; Murphy et al., 2007; Gordon-Larsen et al., 2009)
- oscillometric blood pressure (Gordon-Larsen et al., 2009)
- diastolic blood pressure (Murphy et al., 2007; Gordon-Larsen et al., 2009)
- body mass index or BMI (Barengo et al., 2004; Wagner et al., 2002; Wennberg et al., 2006; Hu et al., 2007; Hamer & Chida, 2008)
- body fat (Murphy et al., 2007)
- cholesterol (Barengo et al., 2004; Wennberg et al., 2006; Hu et al., 2007)
- cardiovascular fitness in terms of $VO_2$ Max (Murphy et al., 2007)
- serum measures of lipids, glucose, and insulin (Gordon-Larsen et al., 2009)

Hamer and Chida (2008) identify an 11% reduction in cardiovascular risk associated with active commuting. Gordon-Larsen et al. (2009) uses multivariate linear and logit regression models to determine the associations between self-reported active commuting trip features (time, distance, and mode), body weight, obesity (BMI ≥ 30 $kg/m^2$), fitness (symptom-limited exercise stress testing), moderate-vigorous physical activity (through accelerometer) and CVD risk factors. Their results show that among the four groups of participants (male active and non-active commuters, female active and non-active commuters), active men experience a reduced CVD risk due to higher levels and intensities of active travel than the other groups. Li and Siegrist (2012) show that the high levels of leisure time physical activity (PA) and moderate levels of occupational PA could reduce the total risk of coronary heart disease (CHD) and stroke by 20-30%, among men and 10-20% among women. Bennett et al. (2017) identify that every four MET-h/day increase in total physical activity (equal to one-hour brisk walking per day) reduces the risk major vascular events by 6%, main coronary diseases by 9%, ischemic stroke by 5%, intracerebral hemorrhage by 6%, and CVD death by 12%. Increases in physical activity (including active commuting) play a significant role in reducing the risk of CVD.

#### *2.2.1.4 Physical Activity and Cancer Risk*

Cancer causes about one in every six deaths worldwide, which exceeds the mortality rate of AIDS, tuberculosis, and malaria combined (American Cancer Society, 2018) and makes cancer the second-leading cause of death. Based on the International Agency for research on Cancer (IARC), 17 million new cases of cancer occur annually worldwide (Bray, 2018). In the U.S., an estimated 1,735,350 new cancer cases occur, and cancer causes 609,640 deaths (National Cancer Institute, 2018). Physical activity not only reduces the risk of obesity-related cancer development, but it also improves quality of life of cancer patients. According to Bhaskaran et al. (2014), leisure time physical activity could reduce the risk of thirteen cancers, including liver, lung, kidney, colorectal, and esophageal (adenocarcinoma). Colon and breast cancer risk demonstrate the most consistent inverse relationship with physical activity (Adesiyun and Russell, 2018). Given cancer's significant mortality risk and the physical and economic costs associated with its treatment, any small reduction in cancer risk from physical activity may have significant impact on public health.

Fortunately, the reduction in cancer risk due to increased level of physical activity can be relatively large. Shi et al. (2015) examine the relationship between household physical activity and cancer risk (in general) and find individuals the with greatest levels of household physical activity have a 16% lower cancer risk than individuals with the lowest levels of household physical activity. Cancer risk decreases by 1% for an additional one hour/week and 2% for a 10 MET-hours/week increase in household physical activity. Lynch et al. (2010) shows an average breast cancer reduction risk of 25% for physically active women in comparison with the least active women. Boyle et al. (2012) report that physical activity provides an average proximal/distal colon cancer risk reduction of 26-27%. The relationship between physical activity and prostate cancer appears less clear because some studies (Patel et al., 2005; Giovannucci et al., 2005; Nilsen et al., 2006) show an inverse relationship and, others show no association (Friedenreich et al., 2004; Zeegers et al., 2005; Littman et al., 2006). Other research on the relationship between physical activity cancer investigated rectal cancer (Lee et al., 2007), lung cancer (Kruk et al., 2013), and endometrial cancer (Voskuil et al., 2007). Table (2.1) shows the epidemiologic evidence on the association between physical activity and cancer risk (adapted from Schmid and Leitzmann, 2014). While the link between physical activity and cancer risk requires further investigation, no previous research indicates that physical activity represents a cancer risk.

**Table 2.1 Epidemiologic evidence on the association between physical activity and cancer risk (adapted from Schmid and Leitzmann, 2014)**

| Cancer site | Average risk reduction | Level of epidemiologic evidence |
|---|---|---|
| Colon | 25% | Convincing |
| Breast | 25% | Convincing |
| Endometrial | 20-30% | Probable |
| Lung | 20-50% | Possible |
| Pancreatic | 25% | Possible |
| Gastric | 30% | Possible |
| Prostate | 10% | Insufficient |
| Ovarian | <10% | Insufficient |

#### *2.2.1.5 Physical Activity, Bone Health, and Osteoporosis*

Osteoporosis (or porous bone) causes a reduction in the density and quality of bones, and, the risk of fracture greatly increases when the bones become more porous and fragile (National Osteoporosis Foundation, 2019). The National Institute of Health (NIH) indicates about 54 million Americans either have osteoporosis or appear at risk for low bone mass (NIH, 2018). In addition, one in two women and up to one in four men age 50 and older break a bone because of osteoporosis (NOF 2019). Also, in 2015, 2.3 million osteoporosis-related bone fractures occurred to two million Americans on Medicare (Hansen et al., 2019). Vicente-Rodriguez (2006) and Arab ameri et al. (2012) express that exercise remains the most prominent non-pharmacological way to improve or maintain bone mass. During physical activity, different stresses on the body result in mechanical loading (e.g. jumping, running, resistance training), which enables bones to start the bone remodeling process. Physical activity affects the risk of osteoporosis through this bone turnover process (Miles, 2007). Physical activity appears critical for older adults seeking to reduce the risk of this disease.

Previous studies investigate the specific relationship between physical activity and bone health. According to some studies (Frost, 2003; Marques et al., 2012; Gregov and Salaj, 2014) various intensities and modalities of physical activity have different influences on bone mass; however, the role of intensity, duration, and frequency of physical activities yielding an optimal osteogenic exercise response remains unclear (Bielemann et al., 2013). However, people of all ages and both genders doing sports or physical activity have higher bone mass, bone strength, and greater osteogenic potential in comparison with those who are not physically active (Scott et al., 2008; Arasheben et al., 2011; Quiterio et al., 2011; Gregov and Salaj, 2014). Chastin et al. (2014) examines the association between sedentary behavior, physical activity, and bone health in terms of bone mineral density (BMD). Their results show larger amounts of time spent on moderate to vigorous physical activity (MVPA) increase total bone density. Therefore, programs and policies that keep older adults using active transportation and public transit should improve overall bone health.

#### 2.2.2 Economic Costs of Physical Inactivity

The previous section showed that regular physical activity can reduce the risks of Type 2 diabetes, cardiovascular diseases, different cancer types, bone problems, and mortality. Ding et al. (2016), estimated the world-wide healthcare costs of physical inactivity in 2013 as \$53.8 billion where the public sector paid \$31.2 billion, the private sector paid \$12.9 billion and patients (households) paid \$9.7 billion. Ding et al. (2017) conducted a systematic review of 40 studies about the economic burden of physical inactivity and found that all of the studies measure either direct or indirect healthcare costs related to physical inactivity.

Kirch (2008) defines direct health care costs or the costs due to resource use as those attributable to the use of a healthcare intervention or illness. While the direct non-medical costs include transportation and additional paid caregiver time, the direct medical costs include all expenses related to interventions, follow-ups in ambulatory, in-patient, and nursing care. Indirect health care costs represent the expenses caused by reduction and/or cessation in work productivity due to the morbidity and mortality related to a disease (Boccuzzi, 2003). Studies estimating direct healthcare expenses use either an econometric or a population attributable fraction (PAF)-based approach.

#### *2.2.2.1 Direct Healthcare Costs by Population Attributable Fraction (PAF) Approach*

This method calculates the physical inactivity-induced health care expenditures by applying a PAF to disease-related costs (Ding et al., 2017) where a PAF represents the proportion of a disease eliminating by removing physical inactivity. Katzmarzyk et al. (2000) estimates the effects of physical inactivity on coronary artery disease, stroke, colon cancer, breast cancer, Type 2 diabetes, and osteoporosis. Using the computed risk ratio for each disease and prevalence of physical activity, the study applied a PAF to the total direct healthcare expenditures in 1999, and total number of deaths connected with each disease. The results showed that physical inactivity contributed to about $2.1 billion or 2.5% of the total direct healthcare costs in Canada. In addition, physical inactivity caused 21,000 premature deaths in 1995 and, a reduction of 10% in physical inactivity could save $150 million in annual direct healthcare costs. Stephenson et al. (2000) applied a PAF approach to examine the direct healthcare expenditures of six diseases due to physical inactivity (with the rate of 44%) among the adult Australian population. They estimate the annual direct healthcare expenditures attributable to physical inactivity as $161 million for coronary heart disease, $28 million for non-insulin dependent diabetes, $16 million for colon cancer, $16 million for breast cancer, $101 million for stroke, and $56 million for depressive disorders, which totals more than $370 million per year. Sensitivity analysis suggests that every one percent increase in the proportion of sufficiently active population can result to $3.6 million savings in the health care expenditures of coronary heart disease, non-insulin dependent diabetes, and colon cancer. Garrett et al. (2004) calculates the physical inactivity-related healthcare costs; physical inactivity relates to approximately 12% of depression/anxiety, and 31% of colon cancer, heart disease, osteoporosis, and stroke. Heart disease represents the most expensive outcome of physical inactivity and costs $35.3 million of the total $83.6 million in direct healthcare costs measured during the study. In China, physical inactivity increases the risk of five major non-communicable diseases (NCDs) of type 2 diabetes, cancer, hypertension, stroke, and CHD by 12% to 19% (Zhang and Chaaban, 2013). The results show that physical inactivity accounts for more than 15% of the annual medical and non-medical yearly costs of NCDs in China. In the Czech Republic, physical inactivity appears to contribute only 0.4% of total healthcare costs (Maresova, 2014), and in the United Kingdom, physical inactivity accounts for 6.5% of total healthcare costs (Scarborough et al., 2011). Physical inactivity presents a clear economic burden on public health, but other factors also impact healthcare costs and outcomes.

#### *2.2.2.2 Direct Healthcare Costs by Econometric Approach*

An econometric approach links physical inactivity and healthcare costs at the individual level (Ding et al., 2017). Pronk et al. (1999) examine the association between modifiable health risks including physical inactivity and the subsequent healthcare charges. The results indicate that an additional day of physical activity per week reduces an individual's median healthcare costs ($600) by 4.7%. Regularly physically active individuals spend $1,019 on their mean healthcare costs while physically inactive individuals spend $1,349 (Pratt et al., 2000). Anderson et al. (2005) use multi-variate linear models to estimate health care charges and investigate the effect of physical inactivity and overweight or obesity status among the U.S. white population aged 40 years and older. This study considers age, gender, physical activity, BMI, chronic disease, and smoking status when predicting the average annual health care charges, and the results show that physically inactive and overweight (or obese) individuals incur 23% higher health care costs for the state health plan and 27% higher national health care costs. Carlson et al. (2015) examines the relationship between healthcare expenditures and inadequate levels of physical activity The study classifies respondents as 1) active, reporting at least 150 minutes/week of moderate-intensity equivalent physical activity; 2) insufficiently active, reporting some moderate-intensity equivalent physical activity but not enough to meet active definition; or 3) inactive, reporting no moderate-intensity equivalent physical activity that lasted at least 10 minutes. According to the results, the mean per capita difference between annual healthcare expenditures of inactive versus active adults is $1,437 and insufficiently active versus active individuals is $713. After including BMI to control for body composition, these differences become $1,313 (inactive vs. active) and $576 (insufficiently active vs. active). Min and Min (2016) compare the direct health care costs (inpatient, outpatient, and prescription costs) of physically active individuals who exercise at least once a week with those of physically inactive people using a propensity score-matching method (calculated from multivariable logistic regression) to reduce the bias between the two groups. The results show that the mean of total direct costs for inactive individuals is $1110.5, which remains 11.7% greater than the costs of active individuals. In addition, the specific disease-related medical costs vary from 8.7% to 25.3% higher for inactive individuals compared to active ones. Overall, the studies using an econometric approach identify larger amounts of healthcare costs than the PAF studies (Ding et al., 2017). The econometric models attempt to control for confounding effects and may capture interaction effects.

#### *2.2.2.3 Indirect Healthcare Costs*

The work productivity losses from absenteeism, presentism, and worker replacement can be valued from the perspective of the employer (friction cost approach or FCA by Koopmanschap et al., 1995), or individuals (human capital approach or HCA by Koopamanschap and Rutten, 1996). Moreover, the "value of statistical life (VSL)", which represents a common approach in transportation safety, monetizes an average or statistical life lost (Viscusi and Aldy, 2003). Other studies (Katzmarzyk and Janssen, 2004, Cadhilac et al., 2011, and Krueger et al., 2015) provide more information about the methodology for calculating the indirect costs of physical inactivity.

## 2.3 Transportation User Data Collection and Activity Recognition

Many recent research studies of daily transportation behaviors use applications to detect the spatial travel patterns of road users (Via et al. 2018); the algorithms these studies use to identify the activity or trip differ in their data sources. Different mapping techniques, resulting from spatial positioning systems and spatial data analysis, represent important technologies for user activity/trip recognition (Cho & Choi 2015). However, building algorithms and data analysis tools from these technologies remain difficult and very complex (Zong et al., 2015). The monitoring process of the travelers' behavior involves dealing with geographic data that can be compatible with digital maps, which allows comparison and presentation of data in a visible way (Turner et al., 1998). The amount of data recorded, which is based on the rate of sampling taken (in seconds or minutes), may require approaches compress the data without reducing the accuracy of the desired results (Patire et al., 2015). The magnitude of the data collected and the need to analyze it in the dimensions of space and time complicates the identification techniques.

Spatiotemporal data management (STDM), which observes time thresholds, allows applications to determine the spatial and temporal state of the road user by identifying changes from inside buildings (activity) to exits for travel (trip). Since these thresholds remain difficult to recognize, previous studies adopt a variety of strategies to determine these state changes; however, most of the algorithms seek greater accuracy and reduced consumption of smart-device batteries. This topic remains an active area of on-going research.

### 2.3.1 Data Collection with Mobile Application

Smartphones represent the most popular device for user data acquisition, which records and processes daily trips. Smart-devices such as phones, smart-watches, and other wearable devices have provided enormous data, many of which have been used in the transportation field. Many researchers use machine learning and classification algorithms to monitor the movement of individuals and their travel patterns using smart devices (Das & Winter, 2016). In many cases, the researchers use only GPS data or speed and acceleration data together with GPS logs to detect the activity pattern (Ansari & Golroo, 2015). Many claim the ability to detect the transportation activity pattern with accuracy while ignoring some of the problems associated with these operations (Ganti et al., 2011, Prelipcean et al., 2016) like, excessive energy consumption and battery life. The limited battery life of smart devices creates additional challenges for researchers trying to identify travel and activity patterns.

Gathering data from the daily activities of travelers paves the way for analyzing other travel behaviors such as mode or trip purpose. For years, researchers have tried to develop mechanisms for collecting travel activity data because classic methods, such as paper travel diaries, phone interviews, or web forms, have significant recall and spatiotemporal errors. The use of GPS has facilitated significant leap forward in the study of travel behavior (Rojas et al. 2016). Modern studies and their applications rely on the use of GPS to monitor the movement of all modes of road users (e.g. pedestrians, cyclists, drivers, and public transport passengers) (Rose, 2009). The use of GPS itself has undergone a series of developments that have increased its efficiency, reliability, and versatility (Facchinetti, 2016). The use of GPS appears critical for most future travel behavior studies.

### 2.3.2 Data Clustering by Geohash Method

Identifying the threshold of change for traveler activities of road users represents significant improvement because the threshold of the spatiotemporal change of any person reveals the travel pattern and trip. Clustering algorithms may be used with the Geohash system to compress data, reduce storage volume and facilitate the presentation of results (Le-Khac et al., 2010). The use of the Geohash system, as a geographic coordinate system, represents a modern method of indexing coordinates. The Geohash system relies on a vertical-horizontal coding principle of the earth grid divisions (Suwardi et al., 2015), which approximates taking aerial photographs. The resolution increases based on the number of Geohashes in each case. Without clustering, the Geohash technique has little importance, especially in the topic of spatial-temporal studies. Geohash clustering links similar elements of data in groups for the purpose of evaluating and diagnosing them (Singh et al., 2017, Nin et al., 2014) for achieving better accuracy. Many different sectors, including spatial modelling studies for business (Suwardi et al., 2015), mobile sensing (Lee et al., 2016, Environments, 2017), spatial query (He, 2017), and spatiotemporal mapping (Deiotte & Valley, 2017) previously used the Geohash technique. The Geohash technique appears to be a strong candidate for improving the identification of the threshold of change between activities.

Recently, many transportation researchers have adopted the Geohash method for their studies. Singh et al. (2017) uses the Geohash method to investigate the flow orientation in South Korea. The study uses smart card data for the bus and subway networks to represent the main activity areas for major work and residential areas (Singh et al., 2017). Oh et al. (2017), use the Geohash method to assess the spatial movement patterns of smart card transaction data in a multi-modal transportation network. This method has also been used in New York City to predict the density of taxi pickups throughout New York City as it changes from day to day and hour to hour (Sdaulton, 2018). Large automated communication and data recording companies also use the Geohash technique. Previous research appears to verify the utility of the Geohash technique.

### 2.3.3 Data Clustering by GIS Method

The study of travel behaviors often uses a Geographic Information System (GIS) to study the+ spatial relationships for transport modeling. Spatial analysis with GIS technique has evolved on a large scale for transportation research, since the movement and activity of people vary geographically over space (Kamruzzaman et al., 2011). Some travel behavior studies Use GIS maps to identify origin and destination points that will later determine the path and mode of transportation (Lari & Golroo, 2015). Stenneth et al. (2012) explain the possibilities provided by GIS maps with GPS data in determining transportation activity, by identifying data for speed and acceleration. Kamruzzaman et al. (2011) study the behaviors of the students by developing activity spaces for students to determine the nature of participation in activities (or lack thereof); however, they rely on a traditional method to record data. In addition, Domènech et al. (2017) develop a methodology to assess the effectiveness and spatial coverage of travel patterns in Spanish tourist cities through a GIS system. Loidl et al. (2016) attempt to develop a relationship between activity and travel pattern through geographic information systems (GIS) by applying geospatial data with geo-visualization. Almost all large scale or aggregate traveler behavior studies rely on GIS to help describe travel patterns.

Most of the research uses a certain dwell time for separating activity and trips by setting a minimum duration for activities (Gong et al.,2014). Different researchers use different thresholds based on the available GPS signal, such as more than 120s (Wolf et al., 2001; Tsui & Shalaby, 2006; Stopher et al., 2002, 2005, 2008ab; Schuessler & Axhausen, 2009), more than 180s (Bohte & Matt, 2009), 200s (Gong et al., 2012) or more than 300s (Axhausen et al., 2004). This threshold varies mainly depending on the characteristics of local activities (Gong et al., 2014). Dwell time represents one strategy for integrating GIS and GPS data to separate activities and trips.

## 2.4 Transportation Mode Detection and User Physical Activities

The current land use patterns restrict the viability of active transportation; however, a recognition of its health implications may cause modal shifts. Substantial literature indicates the importance of easy access to destinations and the diversity of transportation modes as features in livable communities. According to US travel data, 11 percent of trips are by foot, 1 percent by bicycle, and 2 percent by public transportation, which often involve walking or biking when moving to and from transportation terminals (ALRMTAT, 2016). The importance of monitoring travel behaviors extends beyond locations and transportation modes to trip purpose, travel experience, and travelling companions.

### 2.4.1 Review of Different Sensors Used for Transportation Mode Detection

Many studies deal with the subject of detecting transportation modes from the past five years (from 2015 to the third quarter of 2019). Table (2.2) summarizes the transportation mode detection (TMD) methods in previous studies by identifying the studies' algorithm, accuracy, modes and sensors. Figure (2.2) graphically sumarizes some of the statistics associated with the previous research.

**Table 2.2 Summary of Transportation Mode Detection in previous Studies (Continuous)**

| Source | First Author | Country | Year | Classification Algorithm | Accuracy (%) | Transportation Mode: Stationary | Walk | Bike | Motorcycle | Car | school bus | Bus | Train | Tram | Subway | Sensors: GPS/GNSS | GIS | GSM | Wi-Fi | Gyroscope | Accelerometer | Magnetometer | Barometer | Bearing | AC voltage data | surveys |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TRB | Nam | Switzerland | 2017 | NL, CNL, MLP, DNN | DBOEM | | | | | • | | | • | | • | | | | | | | | | | | • |
| TRB | Chapleau | Canada | 2019 | RF, ML | 74.5 | | • | • | | • | • | • | | | | | | | | | | | | | | • |
| TRB | Lari | Iran | 2015 | RF | 96.0 | | • | | | • | | • | | | | • | | | | | • | | | • | | |
| TRB | Ji | China | 2017 | RF | DBOEM | | • | • | | • | | • | | | • | • | | | | | | | | | | |
| TRB | Efthymiou | Greece | 2018 | RF, GBM | DBOEM | | | | | • | | • | • | • | | • | | | | • | • | • | | | | |
| TRB | Wang | United State | 2018 | ML, XGB | 94.5 | | • | • | | | • | • | | | | | | | | | | | | | | • |
| TRB | Elhenawy | United State | 2016 | RF, HMM | 97.0 | | • | • | | • | | • | | | | • | | | | • | • | | | • | | |
| TRB | Rezaie | Canada | 2018 | RF, DT, LR, LP | DBOEM | | | • | | | | • | | | | • | | | | | | | | | | |
| TRB | Yang | United State | 2019 | RF, GLM, DNN, WDM | 95.0 | | • | | | • | | • | | | • | • | | | | | | | | | | |
| Scopus | Antar | Bangladesh | 2018 | MNL, NB, SVM, ANN, BOOST, BAG, RF | DBOEM | | • | • | | • | | • | | | | | | | | | | | | | | • |
| Scopus | Hagenauer | Netherlands | 2017 | MNL, NB, SVM, ANN, Boost, BAG, RF | 93.0 | | • | • | | • | | • | | | | | | | | | | | | | | |
| Scopus | Shafique | Japan | 2015 | SVM, NN, DT, BDT, RF, NB | DBOEM | | • | • | • | • | | • | • | | • | • | | | | • | • | | | | | |
| Scopus | Liang | United State | 2017 | CNN | 94.5 | | • | • | | • | | • | • | | | | | | | | • | | | | | |
| Scopus | Tambi | United State | 2018 | CNN | DBOEM | | | | | • | | • | • | | • | | | | | • | • | | | | | |
| Scopus | Zhu | United State | 2015 | GMM | DBOEM | | • | • | | • | | • | | | | • | | | | | | | | | | |
| Scopus | Bedogni | Italy | 2016 | BN, RT, DTs, RF | DBOEM | • | • | • | | • | | • | • | | | | | | | • | • | | | | | |
| Scopus | Song | Japan | 2016 | RNN | DBOEM | • | • | • | | • | | | • | | | • | | | | | | | | | | |
| Scopus | Quintella | Brazil | 2016 | BN, DT, SVM, J.48 | DBOEM | | • | • | • | • | | • | | | • | • | | • | • | | | | | | | |
| Scopus | Wang | China | 2018 | XGBoost, LightGBM, DT | 90.6 | | • | • | | • | | • | | | • | • | | | | | • | | | | | |
| Scopus | Wang | China | 2018 | LSTM, AdaBoost, SVM | 96.9 | | | | | • | | • | • | | • | | | | | • | • | • | • | | | |
| Scopus | Xiao | China | 2019 | SVM, MNL, BN, ANN, DHMM, DBN, MRF, CHMM | 95.3 | | • | • | • | • | | • | | | | • | | • | • | | • | | | | | |
| Scopus | Ballı | Turkey | 2018 | NB, BN, K-NN, LR, J48, DT, RF | 99.6 | | • | • | | • | | • | | | | • | | | | • | • | | | | | |
| Scopus | Xu | Australia | 2019 | SVM, K-NN, NB | 97.0 | | • | | | • | | • | • | | | | | | | | | | | | • | |
| Scopus | Dabiri | United State | 2018 | CNN | 84.8 | | • | • | | • | | • | • | | | • | | | | | | | | | | |
| Scopus | Fang | Taiwan | 2017 | AdaBoost, DT, KNN, SVM, and DNN | 95.0 | • | • | • | • | • | | • | | | • | | | | | | • | • | | | | |

MNL: Multinomial Logit, NB: Naive Bayesian, BN: Bayesian network, SVM: Support Vector Machine, ANN: Artificial Neural Network, RNN: Recurrent Neural Network, BOOST, Boost, AdaBoost: adaptive boosting, BAG: Bagging, LR: Logistic Regression, RF: Random Forest, RT: Random Tree, NNC: Nearest Neighbor Classification, DT: Decision Tree, BDT: Boosted Decision Tree, CNN: Convolutional Neural Network, DNN: Deep Neural Network, NL: Nested Logit model, NN: Neural Network, HMM: Hidden Markove Model, DHMM: Discrete Hidden Markov Model, CHMM: Continuous hidden Markov model, k-NN: k-Nearest Neighbor, J48: J48 decision tree, MP: multilayer perceptron, DTs: decision table, C4.5: C4.5 decision tree, FLR: Fuzzy logic regression, QDA: Quadratic Discriminant Analysis, CRF: Conditional Random Field, ELM: Extreme Learning Machine, CNL: Cross Nested Logit, GBM: Gradient Boosting Machine, XGBoost: extreme gradient boost, LP: Label Propagation, GLM: generalized linear model, WDM: wide-and-deep model, MLP: Multi-layer perceptron models, TSA: trajectory segmentation algorithm, CART: Classification and Regression Tree, LightGBM: Light Gradient Boosting Machine, LSTM: Long short-term memory, GMM: Gaussian Mixture Models, DBN: deep belief network, MRF: Markov random field, DBOEM: Different Based on each mode or based on each method

**Table 2.2 Summary of Transportation Mode Detection in previous Studies (Continuous)**

| Source | First Author | Country | Year | Classification Algorithm | Accuracy (%) | Transportation Mode | | | | | | | | | | Sensors | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Stationary | Walk | Bike | Motorcycle | Car | school bus | Bus | Train | Tram | Subway | GPS/GNSS | GIS | GSM | Wi-Fi | Gyroscope | Accelerometer | Magnetometer | Barometer | Bearing | AC voltage dat | surveys |
| Scopus | Etemad | Canada | 2019 | AdaBoost, DT, NN, RF, SVM, XGBoost | DBOEM | | • | • | • | • | | • | • | | • | | | | | | • | | | • | | |
| Scopus | Rezaie | Canada | 2017 | DT, RF, LP | DBOEM | | | • | | • | | • | • | | | • | • | | | | | | | | | |
| Scopus | Ashqar | United State | 2019 | K-NN, CART,SVM, RF | 97.0 | | • | • | | • | | • | | | | • | • | | | • | • | | | • | | |
| Scopus | Zhu | China | 2018 | DT, RF | DBOEM | | • | • | | • | | • | • | | | • | | | | | • | | | | | |
| Scopus | Fang | Taiwan | 2016 | DT, K-NN, SVM | DBOEM | • | • | • | • | • | | • | • | | | | | | | • | • | • | | | | |
| Scopus | Zhou | China | 2017 | RF | 0.944 | | • | • | • | • | | • | | | • | • | | | | | • | | | • | | |
| Scopus | Wang | China | 2018 | ANN, SVM, RE | 93.0 | | • | • | • | • | | • | | | • | • | | | | | • | | | • | | |
| Scopus | Shafique | Japan | 2016 | RF | 99.9 | | • | • | | • | | • | • | | • | • | | | | | • | | | | | |
| Engineering Village | Das | Australia | 2018 | Fuzzy expert systems | DBOEM | | • | | | | | • | • | • | | • | | | | | | | | | | |
| Engineering Village | Guvensan | Turkey | 2018 | RF, J48, K-NN, NB | 95.0 | • | • | | | • | | • | • | • | • | | • | | | • | • | | | | | |
| Engineering Village | Qin | China | 2018 | AdaBoost, DT, RF, SVM, CNN | 94.0 | | | | | • | | • | • | | • | | | | | • | • | • | • | | | |
| Engineering Village | Jahangiri | United State | 2015 | K-NN, SVM, DT, BAG, RF | DBOEM | | • | • | | • | | • | | | | | | | | • | • | | | • | | |
| Engineering Village | Feng | Netherlands | 2016 | NB, BN, LR, MP, SVM, DT, C4.5 | DBOEM | | | | | | | | | | | • | | | | | | | | | | |
| Engineering Village | Tang | United State | 2015 | DT | DBOEM | | • | • | | • | | • | | | | | | | | | | | | | | • |
| Engineering Village | Endo | Japan | 2016 | DNN | DBOEM | | • | • | | • | | • | • | | • | • | | | | | | | | | | |
| Engineering Village | WANG | UK | 2019 | DT, K-NN, SVM, NB | DBOEM | | • | | | | | • | • | | • | • | | | | • | • | • | • | | | |
| Engineering Village | Zhu | China | 2016 | RF | 82.9 | | • | | | • | | • | | | | • | | | | | | | | | | |
| Engineering Village | Zhou | United States | 2016 | RF | 93.8 | • | • | • | | • | | | | | | • | | | | | • | | | | | |
| Engineering Village | Martin | United States | 2017 | K-NN, RF | 96.8 | | • | • | | • | | • | • | | | • | | | | | • | | | | | |
| Engineering Village | Yang | China | 2016 | ANN, SVM, RF, BN | 90.0 | | • | • | • | • | | • | | | | • | • | | | | | | | | | |
| Engineering Village | Etemad | Canada | 2018 | DT, RF, NN, NB, QDA | 96.5 | | • | • | • | • | | • | • | | • | • | | | | | • | | | • | | |
| Engineering Village | ASGARI | France | 2018 | CRF, SVM, DT, RF | 83.0 | | • | | | • | | • | • | | • | • | | • | | | | | | | | |
| Engineering Village | Zhang | China | 2015 | ELM, SVM | 50.0 | | • | • | | • | | • | | | | • | | | | | | | | | | |
| Engineering Village | Xiao | China | 2015 | SVM, MNL, ANN, BN | DBOEM | | • | • | | • | | | | | | • | | | | | | | | | | |

MNL: Multinomial Logit, NB: Naïve Bayesian, BN: Bayesian network, SVM: Support Vector Machine, ANN: Artificial Neural Network, RNN: Recurrent Neural Network, BOOST, Boost, AdaBoost: adaptive boosting, BAG: Bagging, LR: Logistic Regression, RF: Random Forest, RT: Random Tree, NNC: Nearest Neighbor Classification, DT: Decision Tree, BDT: Boosted Decision Tree, CNN: Convolutional Neural Network, DNN: Deep Neural Network, NL: Nested Logit model, NN: Neural Network, HMM: Hidden Markove Model, DHMM: Discrete Hidden Markov Model, CHMM: Continuous hidden Markov model, k-NN: k-Nearest Neighbor, J48: J48 decision tree, MP: multilayer perceptron, DTs: decision table, C4.5: C4.5 decision tree, FLR: Fuzzy logic regression, QDA: Quadratic Discriminant Analysis, CRF: Conditional Random Field, ELM: Extreme Learning Machine, CNL: Cross Nested Logit, GBM: Gradient Boosting Machine, XGBoost: extreme gradient boost, LP: Label Propagation, GLM: generalized linear model, WDM: wide-and-deep model, MLP: Multi-layer perceptron models, TSA: trajectory segmentation algorithm, CART: Classification and Regression Tree, LightGBM: Light Gradient Boosting Machine, LSTM: Long short-term memory, GMM: Gaussian Mixture Models, DBN: deep belief network, MRF: Markov random field, DBOEM: Different Based on each mode or based on each method

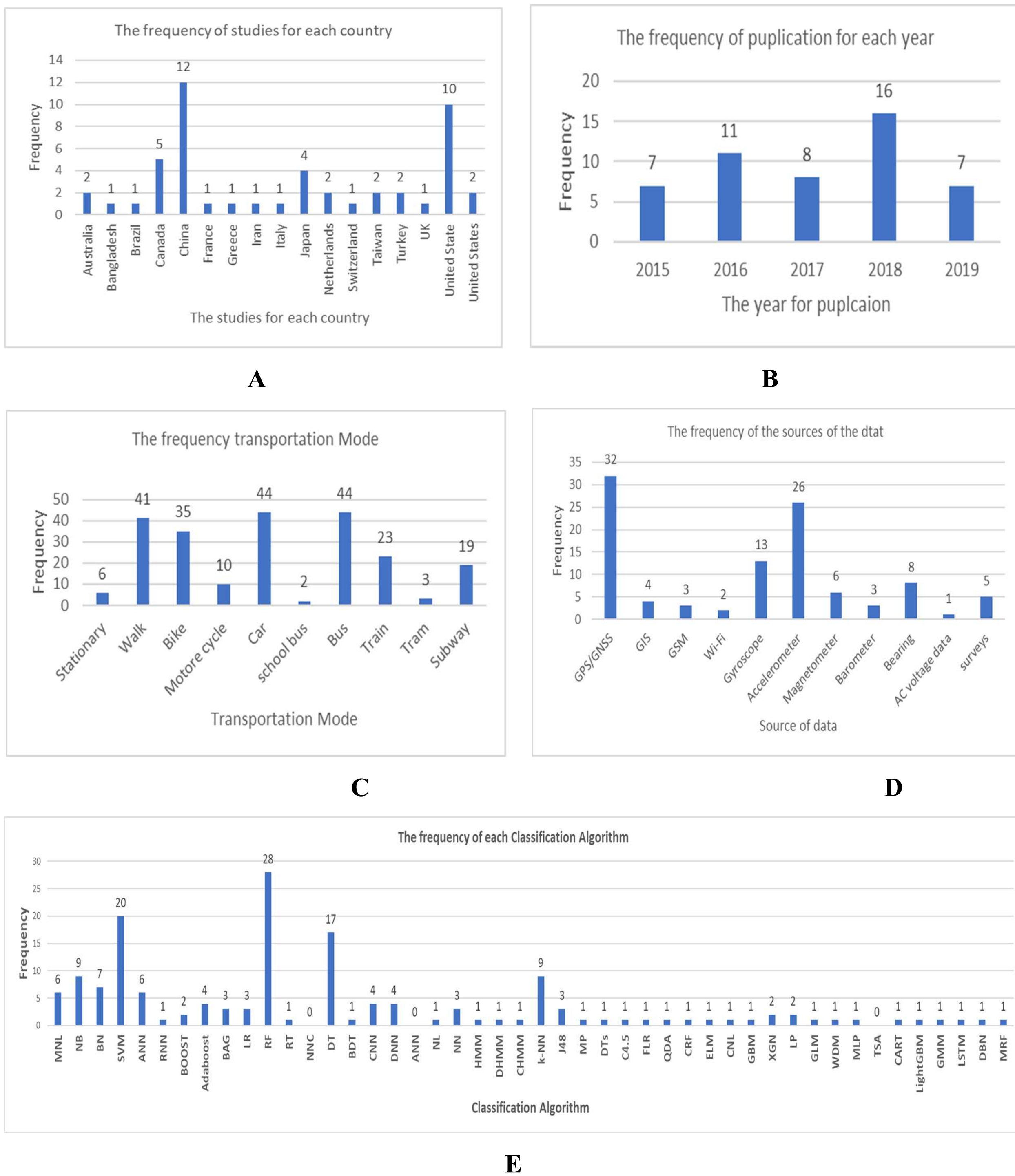



**Figure 2.2 The diagrams (A, B, C, D, E) about the Summary of Transportation Mode Detection in previous Studies**

### 2.4.2 Transportation Mode Detection with Human Activity Profile and Machine Learning

Most of the TMD studies use a phone's sensors and data, while none of these studies address sensors like smartwatches that monitor vital human activities. Smartwatches provide significant and important information that may contribute to predicting transportation modes as well as important information related to road user health. In addition, the literature review shows all of the algorithms used to predict transportation modes and identifies the most used algorithms.

Recent decades have witnessed significant and rapid developments in communications, software, and hardware that have enabled computers to spread and become smaller in size and lower in cost. Many wearable devices and smartphones include advanced sensors, such as Global Positioning System (GPS) devices and physical activity sensors. The development of these devices creates an opportunity for researchers to gather a vast amount of data that requires documentating, processing, and analysis. As previously noted, the issue of battery power to provide the longest possible data collection time represents a critical issue. Reducing the rate of sampling, including reading GPS, represents one of the most effective ways to reduce energy depletion (Liu and Li, 2017). Reducing the sampling rate increases the working time of smart devices but negatively affects the accuracy of the data and the ability to diagnose the thresholds of change in people's activities. Therefore, this study seeks to optimize the tradeoff between accuracy and battery consumption when determining the activities of transportation system users.

For detection of transportation modes, many studies, both conventional and modern, have tried to find the best ways to predict transportation modes through the information provided by a monitoring procedure. Numerous studies use Machine Learning methods in transportation modes detection, either through the use of neural networks, deep learning, or other machine learning techniques, such as Random Forest (RF), Adaptive Boosting (AdaBoost), and Support Vector Machine (SVM). Most of these studies provide a wide range of performance results due to variations in the data in each study. Some limitations have have been associated with several previous studies for TMD regarding the inconsistent accuracy of their predictive methods. A thorough review must compare the accuracy of previous studies as noted in Table (2.2).

### 2.4.3 Relationship of Transportation Mode with User Physical Activities and Characteristics

Diversification of transportation modes in a community represents an important strategy for developing a livable community with more efficient traffic operations and fewer crashes and environmental impacts. Active transportation, such as walking and cycling, provides one of the essential elements of transportation diversification, with physical activities that can contribute to improved health (Edwards et al., 2008). Public transport also promoties physical activity and reduces large traffic volumes (Laverty et al., 2018). Health problems caused by the stable lifestyle of all countries, including Europe and North America, have increased recently (Varo et al., 2003). In 1996, the American Surgeon General acknowledged the potential role of daily physical activity, including walking and cycling, in achieving health benefits (Li et al., 2006). People place different values on their health; some may spend a lot of money to achieve their goals and others may be happy with no investment. A global health recommendation for adults expects at least 150 minutes a week of moderate physical activity or 75 minutes a week of vigorous physical activity (Oja and Titze, 2011). Evaluating and classifying transportation and other activities requires successive stages of data logging, detection of transportation modes, and levels of physical activity (Yang et al., 2018). Previously studies use traditional methods of questioning people about their daily travel activities rather than smart devices (Milne and Watling, 2019). However, the proliferation of smart devices from smartphones and smartwatches makes providing data on the trajectories of movement and acceleration with information on physical activities (heart rate, calories, number of steps, etc.) possible.

## 2.5 Integrated Transportation and Health Impacts

The application of an integrated approach to assessing the health effects of transportation may seem complicated because of the different data, the various methods of analysis and then the evaluation of the results (Smith et al., 2017). An integrated assessment appears useful for evaluating previous studies, which used a non-integrated assessment of health impact. While the links between travel behavior and health outcomes appear clear, integrating health impacts into transportation decision making remains challenging due to the lack of readily available tools, data, and methods familiar to transportation planners (Wu et al., 2017).

Health assessment includes many confounding effects that complicate the analysis. For this reason, health professionals have more problems than others in using "target management" guidelines because of the blurry definition of good health (WHO, 2017). This lack of clarity in the description of goals and methods of health benefits measurement appears when addressing the issue of transportation and health. The cities' built environment and the transportation system represent the most effective factors to promote walking and cycling and improve public health(WHO, 2017). Support for active transportation through infrastructure investment represents an effective stimulus for improving public health (Procyk et al., 2013). Assessing the health benefits of transportation requires controlling for the potentially confounding effects from other factors.

The health outcomes induced by transportation-related physical activities (PA) may reduce healthcare costs (Linkages and Heli, 2009). In recent years, diseases related to sedentary life have increased, particularly in countries that rely on the use of private cars as the dominant modes of transportation (WHO, 2010). Sedentary living represents a significant cause of many deaths, as well as many diseases such as cardiovascular disease, respiratory, obesity, colorectal cancer, indigestion, blood pressure, fat disorders, depression, and anxiety. The health problems associated with people's dependence on private vehicles with an extensive network of roads not only reduces PA and thus causes disease, but also increases traffic crashes, emissions, noise and global warming problems (Haines and Dora, 2012). More than 30 percent of adults worldwide engage in an insufficient level of physical activity; inactivity increases with women more than men with increases in age, and inactivity levels rise in high-income countries (Hallal et al. 2012). The Centers for Disease Control and Prevention (CDC) note that PA benefits not only reduce disease but also in improve the quality of people's lives by strengthening your bones and muscles, improving your mental health and mood, improving one’s ability to do daily activities and prevent falls, and increasing your chances of living longer (CDC, 2018). Transportation unfortunately also represents one of the primary sources of serious injury and mortality (Christie, 2010). By 2030, transportation is expected to become the seventh leading cause of death. Pollution from emissions and noise also poses significant risks to physical and mental health because they impact heart disease, respiratory disease, and anxiety ( Sitlington, 1999). Fully quantifying the health impacts from transportation planning decisions may change the alternative selection process.

### 2.5.1 Transportation and Its Associated Links Related to The Health and Environment

Developed countries, including the United States of America pay significant attention to the impact of transportation on public health and the environment. The integration of all transportation factors with health may not be easy because of the inter-related causality for an individual or community [Figure (2.3)]. However, both Europe and the United States of America have developed methods for measuring and integrating all of the factors of transportation’s impact on health.

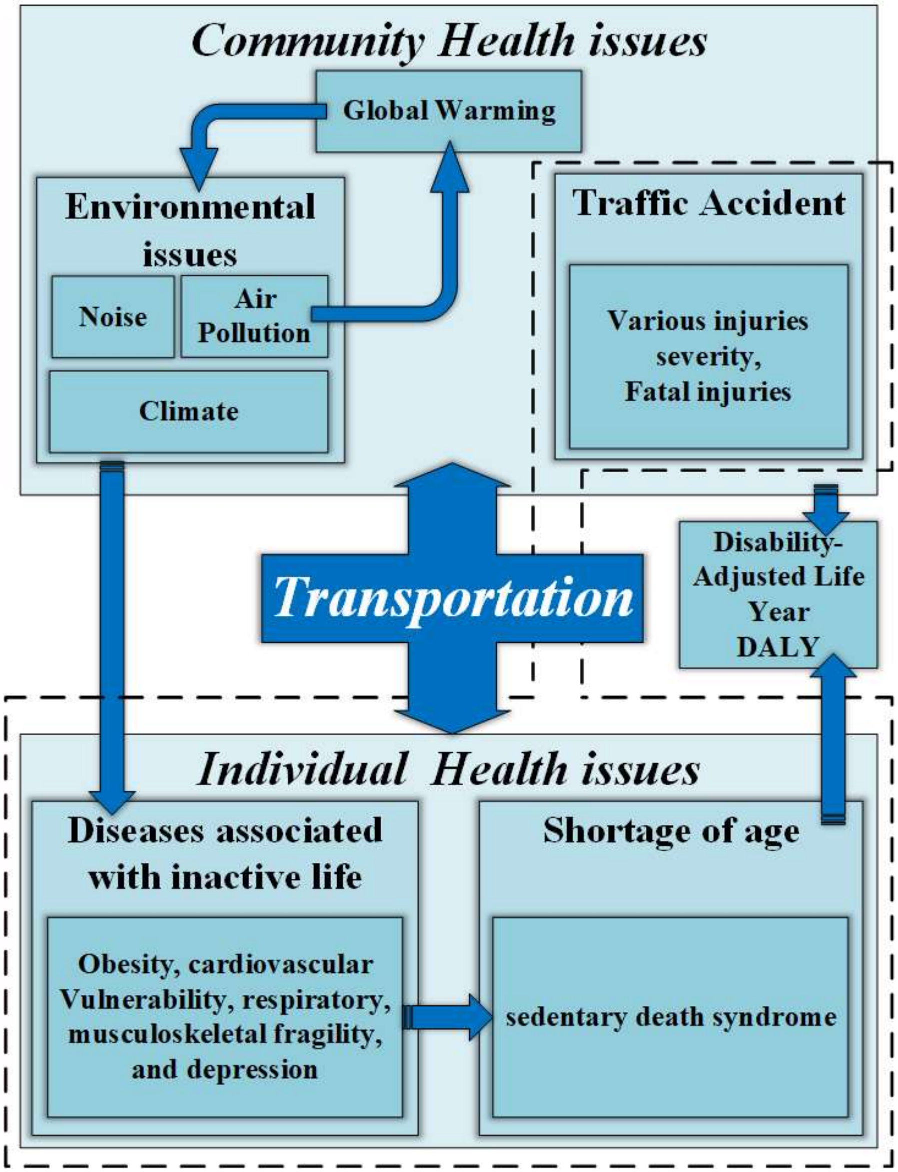



**Figure 2.3 Transportation relationship with community health and individual health**

### 2.5.2 Transportation Issues and Associated Health Risk

The assessment of health risks mainly appears in studies examining the impact of traffic noise, vehicle emissions, and traffic accidents. Many of these studies use available data on mortality or morbidity to determine the Disability-adjusted life years (DALYs). DALYs are used comprehensively for vehicle emissions or traffic noise for the measured group. Meanwhile, DALYs may also be calculated as a percentage of the totals involving traffic crashes. However, the assessment of the transportation health impacts because the studies collectively diagnose both the risks and benefits. Physical activities from walking or biking provide benefits that oppose crash risks and air and noise pollution. Higher benefits than the total risk, in any geographical area, encourages the use of active transportation.

The study of Woodcock et al., 2014 represents a valuable study in the use of the DALYs scale in the stages of assessing the health effects of transportation. The study integrates more than one scenario (substitutions), such as active transportation, emissions, and crash risk, in London, United Kingdom, and Delhi, India by comparing the projection of 2030 with alternative scenarios - vehicles with lower carbon emissions and increased active travel and a combination of the two. The study found that the combination of active travel and low-emission travel would provide benefits by reducing the number of DALYs from ischemic heart disease in the study (Woodcock et al. 2014). A study California determines the health benefits of transportation strategies to reduce greenhouse gas emissions (GHGE) (Maizlish et al., 2013). The results highlight a significant improvement in population health with increased PA associated with active transportation. Previous studies appear to confirm the importance of encouraging active transportation, but they do not provide a clear and consistent strategy for assessing the benefits.

### 2.5.3 Health Benefits Associated with Transportation based on Physical Activity Intensity

Active transportation (walking and biking) represents one of the most effective ways of achieving physical activity. In 1996, the American Surgeon General acknowledged the potential role of daily physical activity, including walking and cycling, in achieving health benefits (Dhondt et al., 2011). Therefore, determining the levels of physical activities for the individuals and communities is a key focus of health institutions (Stewart et al., 2015). Because of different health conditions, the level and type of physical activities that provide improvement for each health condition vary. But in general, the physical activities of any person certainly achieve health benefits in a cumulative way (Schram-bijkerk et al., 2019). The Office of Disease Prevention and Health Promotion (ODPHO) mentioned in its second section, that the physical activities must be done at least 150 minutes per week of moderate intensity or 75 minutes a week of vigorous-intensity aerobic physical activity, or an equivalent combination of moderate- and vigorous-intensity aerobic activity (Chisholm et al., 2012). The 2008 Physical Activity Guidelines for Americans provide information for different age groups and gender when seeking appropriate physical activities for each category. The "2018 Physical Activity Guidelines Advisory Committee Scientific Report" contains a wealth of information about the role of physical activities in health.

The reports note that many of the studies on this topic arrive at unclear interpretations and require continuing research. Therefore, it is highly needed to conduct extensive research on health outcomes and benefits associated with different ways and techniques. Health outcomes related to transportation activities could be further analyzed based on physical activity intensity.

People's concerns may differ in terms of their health goals; however, they likely want to achieve these goals. Therefore, people need to determine the necessary steps to achieve their health goals (Kruk, 2014). The published literature addresses the evaluation of physical activities and the resulting health benefits in two phases. The first phase chooses the method of measuring the physical activities suitable for the researcher according to the tools available to him and matching his research objectives (Sylvia, 2015). The second phase involves determining whether or not the person has achieved health benefits from a physical activity and whether an increase in health benefits due to the increased physical activity occurs. Figure (2.4) groups humans based on their activity level and their use of active transportation. To examine the methods of measuring physical activity, Figure (2.5) shows the most documented methods of measurement within the previous research literature (Ndahimana and Kim, 2017; Kowalski et al., 2012). Figure (2.5) shows the diversity in the daily activities of individuals from intended to unintended. Also, the correlation between Figure (2.4) and Figure (2.5) can be observed. A person may be active because he walks or rides a bike as a mode for his movements.

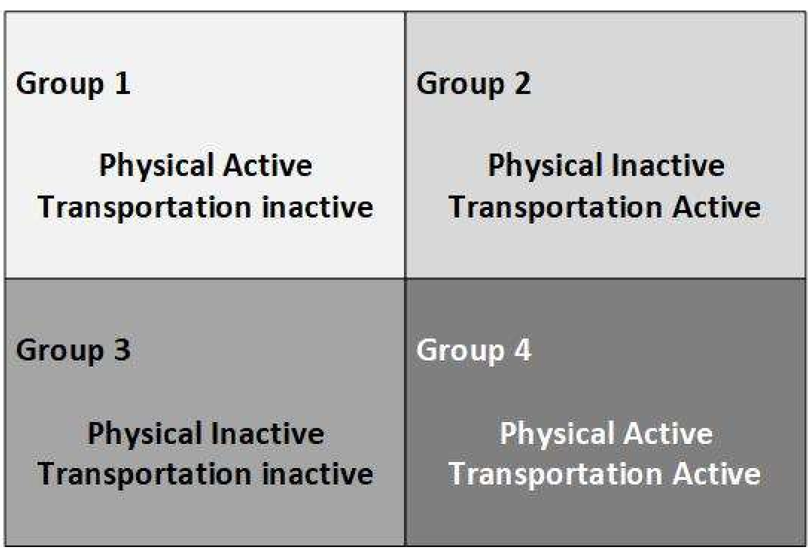



**Figure 2.4 Groups of human activities in conjunction with the usage of PA**

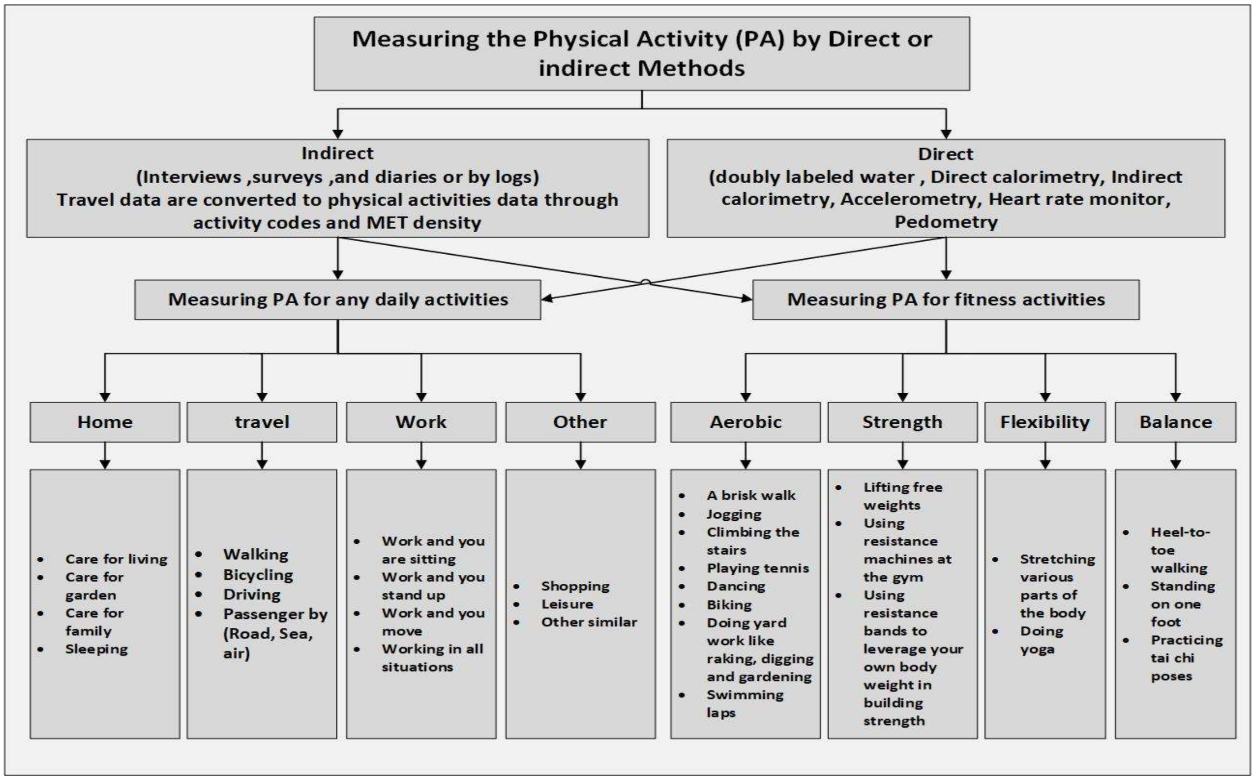



**Figure 2.5 The most documented methods of measuring PA**

The study conducted by Sylvia (2015) provides an extensive explanation of the methods of measuring physical activity, the advantages and disadvantages of each method, and some of the evidence of use cases (Sylvia 2015). This study includes nearly 125 research papers, as reference sources for researchers in this field. Further, Ndahimana et al. (2017) presents a search for methods from 94 research sources of the measurement of physical activity and energy expenditure, which is close to the presentation style of the Sylvia study and presents the advantages and disadvantages of measurement methods.

# Chapter 3 Data Collection and Methodology

## 3.1 Introduction

This research explored the factors impacting the amount of physical activity an individual engages in and the proportion of an individual's daily activity attributable to transportation activities. This research study collected body composition data, physical activity data, and travel data from wearable devices. This study considered four key dependent variables: total physical activity, total physical activity related to transportation, total cardiovascular activity (measured as time spent at age-specific levels of exertion), and total cardiovascular activity related to transportation. The research team identified, and classified health outcomes impacted by physical activities. The effort allowed the research team to incorporate those outcomes into the measurement system in the data collection devices.

## 3.2 Research Approach and Data Collection Technique

The research team developed a mobile app and its associated server-side infrastructure to collect the raw data about the daily activities of the participants using wearable devices. The app included an algorithm that automatically classified the transportation as: walk, run, bike, car, or bus. The algorithm used speed, frequency of stops, accelerometer data, weather conditions, identification of the source and destination locations, and residence time at intermediate locations to identify the transportation mode. Trip length, time of the day, day of the week, and different spatial-temporal conditions data were collected to identify the purpose of the trip. The research team recruited 120 subjects using a stratified sampling strategy from Kalamazoo, Michigan and Arlington, Texas, in order to investigate seasonal variations and locational characteristics in their physical activities.

In terms of data analysis, the research team analyzed a collection of independent variables related to socioeconomic factors (e.g. race, gender, and income). The researchers' hypothesis states that "increasing the role of transportation in achieving physical and cardiovascular activity should have a positive impact on health outcomes." Based on the transportation-related physical activity, the research team hypothesizes that weather, land use, and site design will all impact the total amount of transportation-related physical and cardiovascular activity. Furthermore, the percentage of activity attributable to transportation will be impacted by these factors as well as employment and socioeconomic factors. This project examined the role that transportation-related activities played in an increase of individual's total physical activity and total cardiovascular activity.

The research plan relied on data collected over very different time frames. The study gathered body composition data every six months. At the same time, the study collected travel and physical activity data from the wearable device and the mobile app developed in this project for every 1-minute interval. The travel and physical activity data include activity locations, activity duration, travel distance, heart rate, and total calories during travel and physical activities. Although data were collected for individual subjects, the researchers protected individual privacy by processing all data in an aggregated format. The location information was used for converting their locations into an aggregated format (type of location).

The research team distributed a pre-survey to subjects during intake prior to installing the mobile application and setting up their wearable device. The survey gathered general travel and physical activity information (Pre-survey Questionnaire available in Appendix). In additional to the pre-survey, the subjects completed the device registration form that includes their contact information to assure continuity of data collection during the study period. The research team used the contact information after experiencing system failures while collecting data.

## 3.3 Approach for the Technical Part

This part of the study describes the application to collect the daily activity data of the subjects using smart devices (smartwatch, smartphone). The mobile application utilizes the following Application Programming Interfaces (APIs) to implement its core functionality:

- ***Cordova:*** To build a cross-platform mobile app using HTML5, JavaScript, and CSS3. The use of Cordova allows to easily port the app to Android and iOS.
- ***Fitbit Web APIs:*** To retrieve details about the physical activities of participating subjects including a time series of their activities, heart rate, and sleep logs.
- ***Google Maps or Open Streets APIs:*** To retrieve details about visited locations.
- ***Accelerometer and GPS APIs:*** To retrieve acceleration, estimate the number of steps, speed, and get details about users' traveled paths.
- ***Web Services:*** To interface the mobile app with the back-end database and the data analysis and reporting services. These services allowed raw data access.

The research team uses a server (DELL PowerEdge R210 II) to develop all back-end processes and web services required to grant authenticated and secure access to the collected raw data. This server is currently available at the NEST research lab at WMU.

The mobile application sends a notification in case the participating subject stops sharing data during the study period. In addition, the mobile application sends a daily verification notification for user activity and trip data validation. The research team also developed a "transportation mode" classification algorithm and "purpose of trip" classification algorithm by mining the raw collected data which automatically classifies the transportation mode choices and purpose of activities.

## 3.4 Approach for the Physical Activity Part

This part of the study considers the physical activity data extracted from the smartwatch and mobile application.

### 3.4.1 Body Composition Data Collection

The following body composition data were assessed at the beginning of the survey by using a non-invasive bioelectrical impedance analyzer (InBody 570; InBody Co., Ltd., Seoul, Korea).

- Muscle-fat analysis: weight (lbs.), skeletal muscle mass (lbs.), and body fat mass (lbs.)
- Obesity analysis: body mass index (kg/m2) and percent body fat (%)
- Segmental lean analysis: right & left arms and legs (lbs.) and trunk (lbs.)

### 3.4.2 Physical Activity Data from Fitbit Charge (2 or 3)

The Fitbit charge 2 and 3 version captures various leisure exercises and physical activities including walking, running, aerobic workout, elliptical, outdoor bike, sports, and swimming. The Fitbit measures active minutes, daily steps, number of calories burned and taken in, and heart rate in 60-sec time interval as health measures when an individual engages in a physical activity. The measures include:

- <u>Distance traveled (km):</u> Distance is calculated by multiplying walking (running) steps by walking (running) stride lengths. The stride lengths are estimated using height and gender.
- <u>Heart rate (beats/min):</u> Both resting heart rate and heart rate with physical activities are estimated using a heart rate monitor with photoplethysmography.
- <u>Activity minutes (min):</u> Active minutes are estimated using metabolic equivalents (MET). MET is an indication of how much harder than set a particular activity is. For example, 1-

MET indicates a body at rest, therefore, 3-MET means three times harder than rest, such as stationary cycling or walking at a rate of 4 km/h. MET is estimated in any given minutes by calculating the intensity of physical activity. Active minutes are then earned at or above 3-MET.

- Total calories (Kcal): Total calories are estimated by taking into account basal metabolic rate (BMR) and calories consumed during physical activities in a day.
  a. BMR: BMR is calculated based on gender, age, height, and weight.
     - For men: BMR = 10 × body mass (kg) + 6.25 × height (cm) – 5 age (years) +5
     - For women: BMR = 10 × body mass (kg) + 6.25 × height (cm) – 5 age (years) -161
  b. Calories consumed during physical activities (total calories – BMR): these calories are estimated using the above-mentioned heart rate monitor and a three-axis accelerometer.

### 3.4.3 Analysis Approach for Physical Activity Data

The individual's physical activity data was analyzed based on their physical activity level (including duration and intensity) and one or more health measures, such as BMI (body-mass index), body fat percentage, waist circumference, and waist-to-hip ratio. Physical activity level was categorized based on the Physical Activity Guidelines for Americans, by HHS (2018), as follows:

A) Inactive People: those who do not do any moderate or vigorous-intensity physical activity beyond basic movement from daily activities.
B) Insufficiently Active People: those who do some moderate or vigorous-intensity physical activity, but they still do not meet the key guidelines target range (150 to 300 minutes a week of moderate-intensity physical activity).
C) Active People: those who already meet the key guidelines target.
D) Highly Active People: those who do more than the equivalent of 300 minutes a week moderate-intensity physical activity.

The research team used BMI and daily physical activities as initial grouping criteria, and grouped all the individuals into 16 categories with a combination of four BMI status -- Underweight (BMI less than 18.5), Normal weight (BMI from 18.5 to 24.9), Overweight (BMI from 25 to 29.9), and Obese (BMI equal to 30 or more) and four levels of physical activeness as follows.

$$\begin{pmatrix} \text{Underweight and Inactive} & \cdots & \text{Underweight and Highly Active} \\ \vdots & \ddots & \vdots \\ \text{Obese and Inactive} & \cdots & \text{Obese and Highly Active} \end{pmatrix}_{4\times 4}$$

The research team estimated the amount of daily transportation-related physical activity each group needs to reach the next activity level category, based on their current health (or BMI) condition. The team considered both the duration and intensity of the activities. According to the HHS (2018), individuals achieve additional health benefits by completing more than an equivalent of 300 minutes (5 hours) of moderate-to-vigorous intensity physical activity per week.

## 3.5 Integrated Platform Development for Data Aggregation and Analysis

In this research, the "PASTA" platform integrates mobile application data with an individual's physical activity data. The mobile app collects location information and raw data about the daily activities, which it stores, synchronizes and aggregates to the participants' Fitbit data, such as accelerometer, speed, steps, heart rate, and calories. into the "PASTA" platform. This research study collects data for the following categories and analyzed their relationship.

- Individual characteristics – age, gender, employment, body composition, fitness exercise, amount of physical and cardiovascular activities
- Spatial characteristics – GPS Trajectories and locational information
- Transportation environment and associated built-environment facilities
- Travel activity – trip purpose, transportation mode, travel time, etc.
- Physical activity – amount of physical activity

By using the integrated data from "PASTA" platform, different analysis and relationship between physical activity and transportation options were assessed and evaluated through utilizing the above-mentioned categories.

# Chapter 4 Development of Mobile Application and Integrated "PASTA" Platform

## 4.1 Introduction

Integrated "PASTA" platform includes four components for data collection and data processing. The mobile application provides the user authentication process, activity/trip verification, and location information. The back-end server receives data from mobile phones that have the location information, user authentication, and pulls the physical activity data from Fitbit. The database management system stores and retrieves the data. The final component develops the classifiers.

## 4.2 Mobile Application Overview

The mobile application uses an Ionic 2 cross-platform framework so that the app may be on Android and iOS, but due to the restrictions of iOS background execution, the iOS version does not perform well. The mobile application functionality authenticates the user on Fitbit and grants the application all the required permissions to pull Fitbit data from the server and the app sends the user's access token to the PASTA server. Fitbit provides an OAuth2 authentication mechanism. PASTA application stores the Fitbit access token internally to exchange it with the PASTA server. The mobile app also authenticates the user with the PASTA server, where each user has a registration code assigned and the user has to have a valid registration code. After the user enters this registration code, the mobile app exchanges the Fitbit access token, Firebase access token, mobile device platform, and the time zone. The app and the server do not know anything about the user. The user details are de-identified in PASTA database, for example, no email address, phone number, personal information is used to identify the user. The only way to identify the participant is by the registration code, which is randomly generated, and each user has to have a registration code. The mobile application also verifies the user's activity, where the user receives a notification on a daily basis to verify the activities during that day. The user can also see the activities for any particular day.

### 4.2.1 Backend Side Overview

The backend side handles and manages the data that comes from the mobile phone as well as managing the export of the data to the user for verification. The application server is written using Spring Framework to implement the RESTful APIs, JPA for the persistence layer. The spring application is deployed in a Tomcat Server hosted in the Google Cloud. The server has several RESTful APIs responsible for authenticating the user, receiving locations from users, processing user data, viewing user raw data, etc. The server also hosts static HTML web pages that the team uses by to validate user data and for reporting purposes. After validating the data, the server also stores and retrieves data from the database.

### 4.2.2 Database Management System Overview

The MySQL database system, which is hosted on the Google Cloud platform, stores both raw and processed data using 18 tables. The database schema contains tables for user management, which store user data like user_id, registration_code, and fitbit_access_token. Other tables store data for location information, such as latitude, longitude, time_captured, and speed. Each record in the database is linked with a user_id. Other modules store data for physical activity data, such as heart rate values, time_captured, calories, and steps.

### 4.2.3 Classifiers Overview

The classifiers process the raw data in the database and extract the knowledge from the raw data. This system uses two classifiers. The Activity/Trip classifier takes GPS points and converts them into Activity/Trip segments using a Geohash clustering technique, which encodes coordinates into a Geohash string. Increasing the number of characters increases the proximity between the points. This classifier finds the appropriate clusters for the points and converts them into Activity/Trip segments by validating the duration for each cluster. If the duration is larger than the dwelling time, then this is an activity, else, this is trip. The consecutive trips were merged into one trip from the source and the destination. After finding these segments, the activity type is determined by the Foursquare API that returns the location types. The Physical activity intensity classifier classifies the physical activities into six categories using the heart rate ratio. The classifiers store this information in the database and notify the user to validate the output from the classifiers.

### 4.2.4 Scheduled Jobs Overview

Data processing and knowledge extraction is done offline. After the data is available for a complete day, the scheduled jobs pull the data and manipulate it to become useful. Three scheduled jobs tare deployed in the server. The first one pulls Fitbit Data for each user daily and runs periodically every one-hour. It pulls the data for each user based on the scheduled sync hour, where each user has a scheduled sync hour associated with the user account. The job pulls all the users that match the current system time and processes them. This job populates the Heart Rate, Steps, and Calories tables in the database. The second one runs the Activity/Trip and Physical activity intensity classifiers against the pulled data and the location data. It stores the processed data in the database. The third job notifies users if the user has data needing verification by sending notifications to mobile phones using the Firebase Notification platform.

## 4.3 Integrated "PASTA" System Implementation

This section describes how the PASTA application has been implemented including all application components and technologies. The PASTA application utilizes many technologies; Table (4.1) describes the technologies that have been used by component.

**Table 4.1 Technologies**

| Components | Technology |
|---|---|
| Mobile Application | Ionic 2, Typescript, Java, Objective C |
| Backend Side | Java, Spring Framework, JPA, Spring |
| | Data, Tomcat Server |
| Database Management | MySQL |
| Classifiers | Java, Spring Schedulers |

The user needs to install the PASTA Android application on a phone and sign up for a Fitbit account and link this account with a Fitbit Charge 2/3. Then, the user has to login to the PASTA application using the Fitbit credentials and grant all permissions to the PASTA application. After this process, the PASTA server is allowed to pull the physical activity data from the Fitbit Server directly without involving the mobile application. The user enters a registration code provided by the PASTA administration in the designated text box. After validation, the server exchanges the Fitbit access token, Firebase token, and user's time zone. The server enables the user to receive the locations and pull the Fitbit data on a daily basis. The geolocation tracking plugin sends the user's location every 10 seconds to the server and the server stores these points in the database [Figure (4.1)].

**Figure 4.1 PASTA Ecosystem**

#### 4.3.1 Mobile App Design

This section describes the mobile application implementation, its functionality and its page structure. The mobile application authenticates the user with Fitbit, authenticates the user with PASTA application, and keeps running the background geolocation tracking plugin to send the user's location to the backend. The user must enable some permissions on the phone in order to assure that the app runs in the background mode since this app sends the locations to the server [Figure (4.2)]. The next sections describe each page in detail.

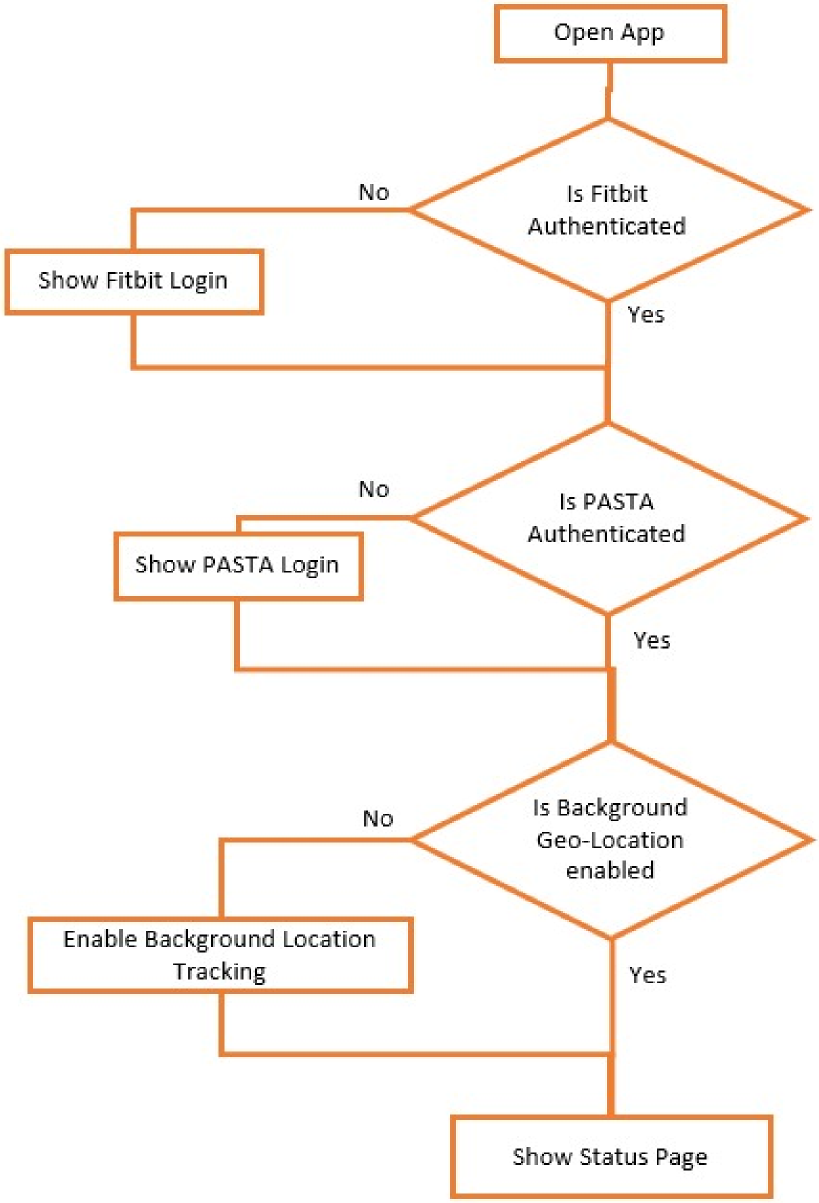



**Figure 4.2 PASTA Mobile App Flowchart**

### 4.3.2 Fitbit Login Page

When the user opens the app, if still not logged in, the first Fitbit Login page (Figure 4.3) appears. The user has to have an account prior to using the PASTA application. The user must enter his or her credentials in the designated text boxes, and app authenticates his/her credentials. The next page requests the user to check the permissions for the PASTA app to access data. The user needs to select and allow all for the PASTA application to pull the fitness data for his/her accounts. The app stores the Fitbit Account Access Token for this user, which permits the server to access the fitness data and store them in the SQLite database. This access token can live for one year. The application changes its status to FITBIT_AUTH, which means that the app has been authenticated from the Fitbit side. The next step is to authenticate from PASTA side. The following screenshots [Figure (4.3)] show the style of the Fitbit login and permission page.

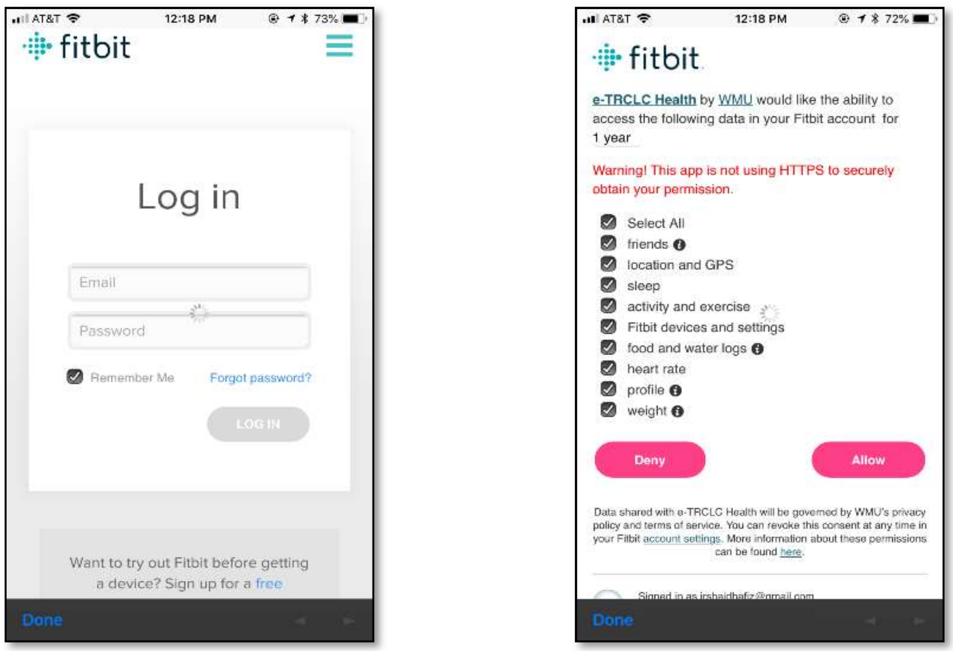



**Figure 4.3 Fitbit Login Page**

### 4.3.3 PASTA Registration Code Page

This page authenticates the user with the PASTA application backend. The user has to have a PASTA registration code assigned by the system administrator, and without this registration code the user will be unable to log in. It should match a registration code in the database. After the user enters the registration code, the application sends the user information to the server and enables this user. It sends a https request to the server carrying information:

a. PASTA Registration Code.
b. Current Time zone.
c. Mobile Platform (Android/iOS).
d. Fitbit Access Token.

e. Firebase User Token.

After sending this information, the application receives a 200-status code, and the application then will change the status of the application to PASTA_AUTH, which means the user is authenticated by PASTA and the app is ready to start tracking the location.

### 4.3.4 Status Page

After the user completed the aforementioned steps, this page will show up. At first, it will change the status of the application into PASTA_AUTH_DONE, and it will enable the background geolocation tracking plugin to send the user's location to the server whenever a significant mode occurred. If the user closed the application and opened it again, the user will not go through the login process again. The user also can log out from the application; after logging out, the application will stop sending the location data to the server and it will go back to the Fitbit Login page.

### 4.3.5 Verification Page

This page allows the user to view his/her activities from any day. The user can verify the locations that he/she visited, and the transportation mode for each trip. It also shows the quantity of the physical activities for each transportation activity. This includes the average heart rate, calories expenditure, and the number of steps, activity duration and the location type [Figure (4.4)].

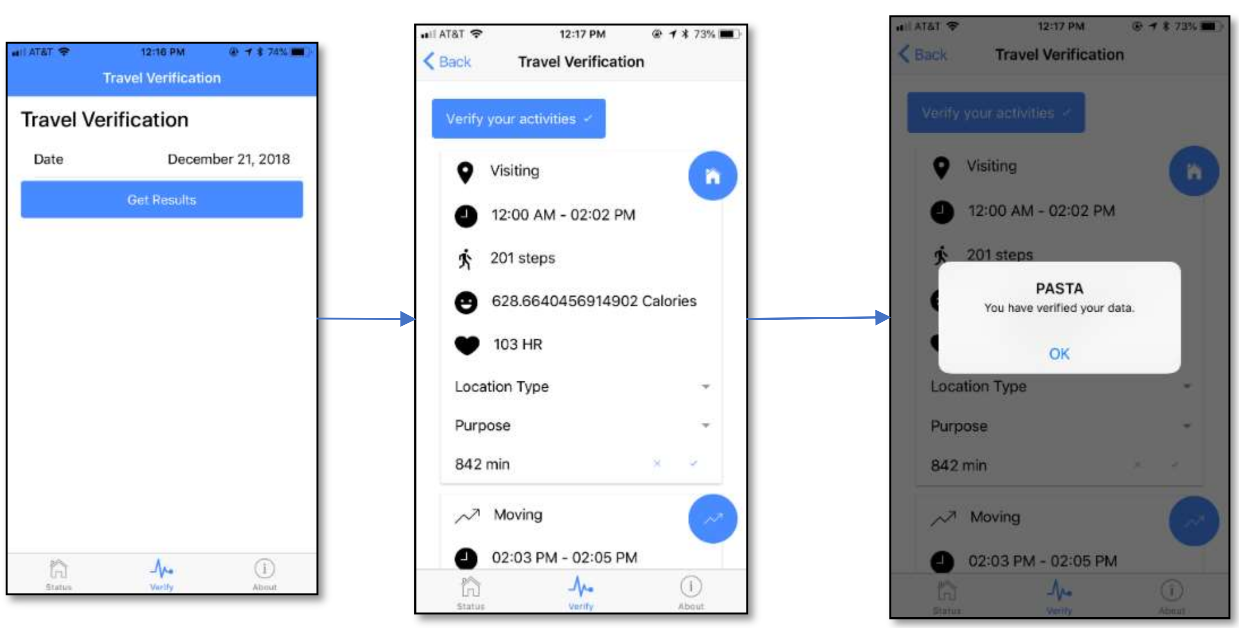



**Figure 4.4 PASTA Verification Page**

### 4.3.6 Background Location Tracking

This plugin tracks the user location and runs in the background. Each 10 seconds, it checks the location and if a significant change in the location occurs, it will report this to the server. This plugin is a battery-saving plugin, and it does not consume much battery time. The next table (4.2) describes each parameter used to setup this plugin.

**Table 4.2 Background Geolocation Plugin Parameters Description**

| Parameter | Description |
|---|---|
| desiredAccuracy | Desired accuracy in meters. Possible values [HIGH_ACCURACY, MEDIUM_ACCURACY, LOW_ACCURACY, PASSIVE_ACCURACY]. Accuracy has direct effect on power drain. Lower accuracy = lower power drain. |
| stationaryRadius | Stationary radius in meters. When stopped, the minimum distance the device must move beyond the stationary location for aggressive background-tracking to engage. |
| debug | When enabled, the plugin will emit sounds for life-cycle events of background-geolocation! See debugging sounds table. |
| distanceFilter | The minimum distance (measured in meters) a device must move horizontally before an update event is generated. |
| stopOnTerminate | Enable this in order to force a stop() when the application terminated (e.g. on iOS, double-tap home button, swipe away the app). |
| startOnBoot | Start background service on device boot. |
| interval | The minimum time interval between location updates in milliseconds. |
| fastestInterval | Fastest rate in milliseconds at which your app can handle location updates. |
| activitiesInterval | Rate in milliseconds at which activity recognition occurs. Larger values will result in fewer activity detections while improving battery life. |
| notificationTitle | Custom notification title in the drawer. |
| notificationText | Custom notification text in the drawer. |
| activityType | [AutomotiveNavigation, OtherNavigation, Fitness, Other] Presumably, this affects iOS GPS algorithm. |
| pauseLocationUpdates | Pauses location updates when app is paused. |
| saveBatteryOnBackground | Switch to less accurate significant changes and region monitory when in background |
| url | Server url where to send HTTP POST with recorded location. |
| syncUrl | Server url where to send fail to post locations |

### 4.3.7 Server-Side Design

This section describes the functionalities of the server for the PASTA application and the RESTful APIs and their functionalities. The server utilizes a Spring Framework technology deployed in a Tomcat server. The server handles the HTTP requests for the RESTful APIs and executes the business logic behind each API. The APIs include:

- /api/classify/{userId}/{dateStr}

The team uses this API to view the data for a particular user for a particular date; it runs the classifier against that user specified in the field {userId} and the date specified in {dateStr}. It returns a JSON that contains all the activities and the trips.

- /api/statusReport/{date}

The team uses this API to view the data available for all active users for a particular date. It takes the parameter {date} as date and returns the number of GPS records, a number of Heart Rate records, Number of Steps Records, etc. for each active user.

- /api/{regCode}/locations

This API posts multiple locations at the same time; this API is necessary when the phone does not have Internet access. The application caches the locations internally until the Internet becomes available when it sends an array of locations to the server.

- /api/{regCode}/location

This API posts one location at a time to the server.

- /api/mobile/{regCode}/dailyActivities/{dateString}

This APIs retrieves the daily activity for each user. the mobile application uses this API for the verification step, and the mobile application populates the {regCode} field and the {dateString} the server will return back the processed activities and trips associated with that user for that particular date.

- /api/{serialNumber}/rowData/{dateString}

This API retrieves the raw data before processing; it does the data fusion process and joins all of the fields together based on the time captured.

- api/{registrationCode}/verify

The phone uses this API to post the verified information about the user's activity and trips.

- /api/user/{serialNumber}/logout

This API disables a user from the system, this API is triggered when the user hits logout button on the mobile app so that user is no longer available.

- /api/user

This API authenticates the user in the PASTA server; this API should have the registration code, Fitbit Access Token, Firebase Access Token to enable the user and allow data retrieval.

### 4.3.8 Database Design

This section describes the database design for the PASTA project, which uses a MySQL database to host the data. The database has 18 tables [Table (4.3)]. Figure (4.5) describes the database schema.

**Table 4.3 Database Schema Details**

| Table | Description |
|---|---|
| ACCELEROMETER | Stores accelerometer Intraday data |
| API_LOG | Stores JSON for each user that contains all physical activity and transportation activity data to be extracted later on |
| CALORIES | Stores Calories Intraday Data |
| CONFIG | Stores Config parameters such that the status of each scheduled task |
| DAILY_ACTIVITIES_VERIFICATION | Stores the processed data from the classifiers |
| DISTANCES | Stores Distance Intraday data |
| ELEVATIONS | Stores Elevation Intraday data |
| FITBIT_APIS | Stores all Fitbit APIs that needs to be called for each user. |
| FITBIT_DATA_STATUS | Stores the status of each day for each user about the data that loader and the classifier status |
| FLOORS | Stores Floor Intraday Data |
| GPS_TRACKING | Stores GPS points per each user |
| HEART_RATES | Stores Heart Rate Intraday Data |
| LOCATIONS | Stores Location Intraday Data and Address associated with each point |
| PREFERRED_LOCATIONS | Stores the preferred locations per each user |
| STEPS | Stores Steps Intraday Data |
| USERS | Stores User Profile data such as User_ID, Registration code, status, Fitbit Access Token, etc. |


**Figure 4.5 Database Schema**

### 4.3.9 Scheduled Job Design

This section describes the functionalities of the scheduled tasks in the PASTA system. These jobs process the raw data offline. The data should be for a complete day, and the job fetches all the enabled users and processes their data. Each user has a field for the sync hour, which specifies the time to process the user data. This technique helps to overcome the limitation in API calls to the third parties, for example, Fitbit does not allow more than 1000 requests each hour, so the users are dissimilated over time to minimize the API consumption and to stay under the maximum limit. The system uses 3 scheduled tasks. The Fitbit Data Loader pulls Fitbit data from the Fitbit Backend and pushes the data to the PASTA database. The second one runs the classifiers against the data pulled from the Fitbit and location data. The third one sends notifications to the users who have data ready from the aforementioned job. The brief job descriptions describe its functionality and frequency:

***Fitbit Data Loader***

- Pull Data from Fitbit for each Participant
- Convert the Data and Store it in the Database
- Frequency: 1 Hour (Users are distributed over time; each one has a particular hour)

***Knowledge Extractor and Classifier Driver***

- Pull Row Data from the Database (Fitbit and Location Data)
- Compute Activity/Trip times.
- Classify Physical Activity Intensity
- Frequency: 1 Hour

***Notification Sender***

- Pull Data Processed Data from Database
- Send Notification to User to Verify Activities by Firebase.
- Frequency: 1 Hour

### 4.3.10 Classifiers Design

This section describes the classifiers implemented to classify trip/activity from row GPS points and to classify the physical activity intensity associated with each activity.

Quantifying the physical activities associated with the transportation activities requires extracting all the trips and activities and identifying the time periods for each activity and trip. The classifier then predicts the transportation mode for each trip and the location type and trip purpose for each activity. The classifiers define six physical activity intensity categories: Very Light, Light, Moderate, Hard, Very Hard and Maximal. These categories show the intensity of the physical activates. These categories can be calculated by the heart rate value and the calories. The classifier divides the heart rate time series data into these categories and calculates the duration of each category.

Before running the Geohash clustering and the activity/trip classifier, the data needs to be retrieved from the database for a particular user and for a particular day. The data also needs to be smoothed and interpolated to provide any missing points due to mobile phone limitations. The interpolation identifies the segments that have a lot of missing data and provides the missing data by measuring the distance between every two consecutive points. If the distance is more than 1 mile, the process requests all these points from a GIS system to provide all the points on the street applies a Kalman Filter to smooth the GPS error and enhance the accuracy.

### 4.3.11 Activity Trip Classifier

After the Geohash clustering method generates all the clusters, the Activity trip classifier generates the list of activities and trips, based on the dwelling time. The dwelling time is the time that the participant spends in a certain place. To register as an activity rather than a trip a participant has to remain at a location for at least the value of the minimum dwelling time; for example, a visit to the supermarket requires a 5-minute dwell time to classify this cluster as an activity. If the duration of the cluster is greater than the dwelling time, then the trip/Activity classifier labels this cluster as an activity. If the duration is less than the dwelling time, then the classifier labels this as a trip. The classifier merges all continuing trips and considers them as one trip.

The research team set the dwelling time to 5 minutes, based on observations of pilot data gathered from team members. This dwelling time is flexible and may be tuned for different testing scenarios. The dwelling time threshold only applies to activities; the classifier still identifies a trip of less than 5 minutes as a trip. The following diagram [Figure (4.6)] shows the process flowchart.

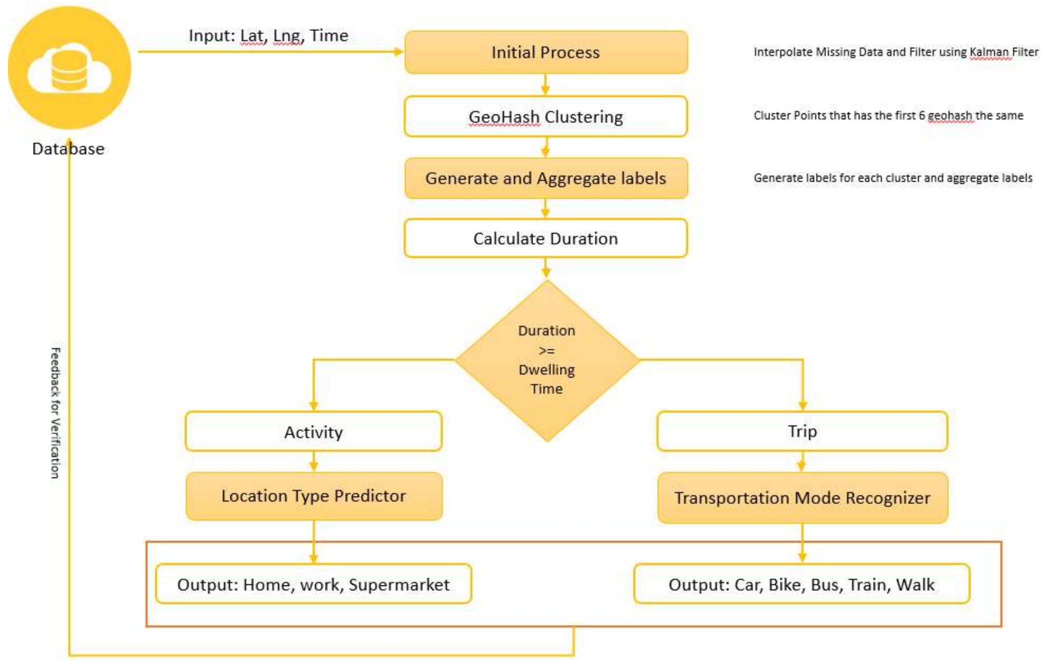



**Figure 4.6 Activity/Trip Classifier**

### 4.3.12 Physical Activity Classifier

This classifier extracts the duration of each physical activity intensity category: Very Light, Light, Moderate, Hard, Very Hard and Maximal. It uses the heart rate ratio values for the all of the heart rate values for a particular activity and then it calculates the duration of each category. The heart rate ratio can be calculated by the following equation:

$$\%\mathbf{HRR} = \frac{HR_{act} - HR_{rest}}{HR_{max} - HR_{rest}} \quad \textbf{(4.1)}$$

Where:

$HR_{Fitbit}$ : Heart Rate

$RHR_{Fitbit}$ : Resting Heart Rate from Fitbit

$HR_{max}$ = (220 – Age)

This equation should be applied for each heart rate value associated with that activity. After that, each heart rate ratio can be classified into the physical activity intensity categories by the following function:

$$f(hrr) = \begin{cases} vl, & 0 \leq hrr < 25 \\ l, & 25 \leq hrr < 45 \\ mo, & 45 \leq hrr < 60 \\ h, & 60 \leq hrr < 85 \\ vh, & 85 \leq hrr \leq 99 \\ ma, & hrr = 100 \end{cases} \quad (4.2)$$

Where:

$vl$: Very Low

$l$: Low

$mo$: Moderate

$h$: Hard

$vh$: Very Hard

$ma$: Maximal

After applying this function to each heart rate ratio value, the classifier needs to calculate the duration of each intensity category. This uses six counters where each counter counts the number of points for each intensity category. The counter value represents the number of minutes since Fitbit provides minute by minute heart rate data.

## 4.4 Limitations of the Mobile Application

### 4.4.1 Background Execution After Terminating the App

The mobile phone is not designed to keep running processes periodically precisely. After closing the app, the operating system (OS) assigns a low priority even with a sampling rate of every 10 seconds. Since data acquisition depends on the phone's status, getting the location data every 10 seconds remains challenging. Android uses Alarm Manager to trigger a broadcast receiver when the app is closed, but this does not function precisely. Based on the collected data to the Android case appears okay with missing data. However, in the iOS case, Apple has a limit of a 15-minute sampling rate, which would lead into miscalculating the physical activity intensity.

### 4.4.2 Location Accuracy

The phone sometimes captures the GPS location with low accuracy (especially inside the building), which leads to inaccurate location points. If the GPS location is not available, the phone captures the location from the network (i.e. Wi-Fi or Cellular network), which has very low accuracy compared with the GPS. In addition, the user has to keep the GPS enabled at all times to get the most accurate results, which might lead into battery drain.

### 4.4.3 Transportation Mode Using Heart Rate

Relying on the heart rate to predict the transportation mode represents a difficult task. Heart rate characteristics vary from individual to individual. The testing for one participant reveals a very low accuracy because creating a machine learning model to predict the transportation mode using heart rate data appears difficult.

### 4.4.4 iOS Version

Apple has limitations in running the applications in the background. Once the app is terminated, one can execute a task that runs minimum, and every 15 minutes to get a sample of the current location. Therefore, feeding this data to the Activity/Trip classifier would lead into an issue calculating the duration. Google Timeline provides such information; however, the Google timeline results are 2 hours, which leads to another problem calculating the physical activity since the time remains critical to the physical activity intensity classifier.

# Chapter 5: Survey Result

## 5.1 Introduction

The research team recruited participants from two different areas, Arlington, Texas and Kalamazoo, Michigan. The team expects that these two locations could provide distinct seasonal and geographical variations in physical and transportation activities. Figure (5.1) shows the location of Arlington and Kalamazoo. Arlington is located between Fort Worth and Dallas in the north Texas region. The Dallas-Fort Worth-Arlington metropolitan statistical area (MSA), is the fourth most populated MSA (7,539,711 estimated as of July 1, 2018) in the U.S (Census Bureau 2018). Kalamazoo is located in southwestern Michigan in the Kalamazoo-Portage MSA, which has a population of less than 500,000. Table (5.1) provides the population demographic details of these cities. While the gender and age distributions appear similar, the annual household median income remains much higher in Arlington ($55,000) than Kalamazoo ($37,000).

**Figure 5.1 Location of Arlington and Kalamazoo**

**Table 5.1 Socio-demographic characteristics of Arlington and Kalamazoo**

| Socio-Demographic | Arlington | Kalamazoo |
|---|---|---|
| Population estimate (July 1, 2018) | 398,112 | 76,545 |
| MSA population | 7,539,711 | 340,318 |
| City Land area (square mile) (2010) | 96.50 | 24.68 |
| Female population (%) | 51.0% | 49.7% |
| Population under 18 years (%) | 26.3% | 19.5% |
| Population 65 years and over (%) | 10.0% | 10.0% |
| White alone, not Hispanic or Latino origin (%) | 40.1% | 64.0% |
| Hispanic or Latino origin (%) | 28.9% | 7.1% |
| Black or African American alone (%) | 21.9% | 21.3% |
| American Indian and Alaska Native alone (%) | 0.4% | 0.3% |
| Asian alone (%) | 6.9% | 2.2% |
| Native Hawaiian and Other Pacific Islander alone (%) | 0.1% | 0.0% |
| Two or more races (%) | 3.1% | 6.3% |
| High school graduate or higher (persons of 25 years and over) (%) | 84.7% | 90.5% |
| Bachelor's degree or higher (persons of 25 years and over) (%) | 29.4% | 34.2% |
| Median household income (in 2017 Dollars), 2013-2017 | $55,562 | $37,438 |

## 5.2 Public Transportation in Arlington and Kalamazoo

Arlington has a unique position in Texas and in the US, because it is the largest (population of 398,112) city without public transportation service. Since 2009, the City of Arlington has operated a door-to-door paratransit service, called Handitran, for older adults and persons with disabilities residing within Arlington. Individuals eligible for Handitran must apply for the service with a $10 application fee and receive a certificate to use the service. Customers can use the service for a $55 monthly pass or $2 per one-way trip. Handitran operates within the city boundaries from Monday to Saturday, and trips can be scheduled up to 14 days in advance. The City of Arlington also supports a micro-transit rideshare service, Via, to connect community members to key destinations around the city center through a Public Private Partnership (PPP). Unlike Handitran, any individual can request the Via service in real time using either Via app or by phone for a $3 flat fee per ride

Kalamazoo provides fixed route transit "KMetro" to the adjacent cities of Portage, Parchment, Texas, and Oshtemo (KMetro, 2019). This system includes four sub-systems as follows.

- Bus system: fixed-route and schedule system operation on weekdays (6:00 AM to 12:15 AM), Saturdays (6:00 AM to 10:15 PM), and Sundays (8:15 AM to 5:15 PM).
- Metro Connect: a shared origin-destination service for all the passengers of Kalamazoo County. In this system, multiple passengers may ride together in the same vehicle.
- Metro Share: is similar to Metro Connect, but only for approved agencies serving seniors and individuals with disabilities at no cost.
- RideShare: a ride matching service helping commuters to find carpool, vanpool, transit, or bike options to get around Kalamazoo County and southwest Michigan.

## 5.3 Initial Survey Development

The research team developed an initial survey to pre-screen participants. This short survey includes questions about age, gender, main daily transportation mode, and approximate daily commuting travel time. Respondents also answered about all their physical activities, number of times and average duration (minutes) in the seven days prior to the time of survey. Since the PASTA app can only be operated with an Android system, the survey asked the type of smart phone (iPhone, Android, or other) they use. The respondents also indicated their willingness to share their activity data if they were provided a Fitbit smart watch. The report contains the complete survey questionnaire in Appendix. The Institutional Review Board (IRB) approval for this survey was accepted in February 2018 for both WMU and UTA. In March 2018, the link of the online initial survey (designed in Qualtrics$^{XM}$) was sent to students, faculty, and staff at both universities. The UTA team sent two reminder emails in the second weeks of April and May 2018. Initially, UTA and WMU collected 1,160 and 880 survey responses and selected 388 and 250 Android smart phone users who were willing to share their activity data through Fitbit as a pre-screened participant pool.

The team compared the number of survey participants based on two criteria, main transportation mode and being physically active (120 minutes or more per week of physical activity). The survey participants of UTA showed the following distributions:

1. Car driver and physically inactive (191 individuals)

2. Car driver and physically active (113 individuals)
3. Non-driver and physically active (78 individuals)

At UTA, the research team sought to oversample the non-driver and physically active group to capture more active transportation data. As a result, they set their target sample sizes for the three groups as 25 for the car driver and physically inactive group, 40 for the car driver and physically active group, and 60 for the non-driver and physically active group. Within each of these subgroups, the research team used simple random sampling to select participants within each group.

## 5.4 Main Survey

The main data collection includes three parts: pre-survey, body composition measurement, and physical activity measurement through Fitbit. Among the survey respondents, the research team recruited 120 individuals (60 from WMU and 60 from UTA) for the Fitbit data collection. Starting from the second week of February 2019, WMU participants visited the Biomechanical Laboratory (in the Student Recreation Center) at WMU for the main survey, and UTA participants visited the Kinesiology Department (in the Maverick Activity Center) at UTA. At the beginning of the measurement sessions, participants completed an informed consent form after reading the terms and conditions, and the pre-survey. The research team conducted body composition measurements for each participant. To facilitate the process, the team sent an email to each participant with the following documents:

1. Link to the online pre-survey
2. Consent form
3. User manual (including a brief description about the participants' responsibilities and how to install and sign up for the Fitbit and PASTA apps)
4. Body composition measurement guidelines and requirements (e.g. dietary restrictions and exercise recommendations during the 24 hours before the tests)

### 5.4.1 Mobile App Installation and Fitbit Activation

The team installed and synchronized the Fitbit Charge 2/3 and PASTA apps on the participants' smartphones. The team instructed the participant about the methods to check/validate the device and app to maintain data validity from both the PASTA app and a Fitbit tracker. For example, the participants had to check the Bluetooth and Location setting daily and keep them running. The research team regularly checked the data from all the participants. During the data collection period, eight participants at UTA and five participants at WMU returned their device due to their poor participation in the study. The research team distributed the returned device to new recruited subjects from the participant pool.

### 5.4.2 Intake Measurements

The In-body measurements include the following:

- Height
- Weight (through traditional and digital scale)
- Body fat percentage (through a digital scale and a hand-held body fat monitor)
- BMI (through hand-held body fat monitor)
- Girth measurements (abdomen as the smallest girth around the abdomen and hip as the largest girth around the buttocks)

The research team provided a few instructions for the in-body measurement to the participants:

- Well hydrate the day before the test
- Do not drink caffeine and eat food 3-4 hours prior to test
- Do not exercise 6-12 hours before the test

## 5.5 Survey Data Analysis

The main survey asked participants about their socio-demographics characteristics, physical activities, and transportation activities (see Appendix for the questionnaires). Table (5.2) compares the survey findings between UTA and WMU. The team also added the average profiles of US population to the table to better understand the distribution of the participants' sociodemographic profiles. UTA has 53% male participants while WMU has 73%. At both universities, the majority of participants are young adults (18-25 age group) or students (67.2% at UTA and 78% at WMU). Whites and Asians represent the majorities at both universities (UTA with 37.9% Whites and 37.9% Asians, WMU with 52.5% Whites and 23.7% Asians). The oversampling of the Asian/Pacific Islander participants primarily reduces the number of white participants; however, both the Black and Hispanic populations appear slightly under-sampled.

The survey responses showed that over 70% of participants perceived that they have good or excellent health and only 5% thought they are not healthy. Regarding physical activity, respondents identified the type, intensity level (from 1 to 10), and duration (minutes) of their activities during the week prior to the survey. HHS Physical Activity Guidelines (2018) indicate that moderate activities have a relative intensity of 5 or 6 on a scale of 0 to 10, while relatively vigorous activity begins at a 7 or 8 on the same scale. Of note, one minute of vigorous physical activity is considered as two minutes of moderate physical activity (HHS, 2018).

Using the suggested criteria, the team estimated each individual's moderate activity level and excluded the light intensity activities, which have the relative intensity of 4 or less. Therefore, each activity with the self-reported intensity of equal to 5 or 6 is considered as moderate, while higher than 6 is taken as a vigorous intensity level. The researchers added the total minutes of moderate-intensity activities for each person and assigned their physical activity level. While the proportion of highly active subjects appears much higher at WMU than UTA (37.3% vs. 20.7%), both Universities show a similar distribution of active and highly active participants (45.8% vs. 44.8%). The National Health Interview Survey (NHIS) (2017) showed that only 51.7% and 21.7% of U.S. adults aged 18 and over met the 2008 federal physical activity guidelines for aerobic activity, and aerobic muscle strengthening activity, respectively. America's Health Rankings Annual Report also showed that 26.2% of U.S. adults were inactive in 2008. The participant pool appears to align relatively well with the national patterns for physical activity, but the inactive group may be slightly oversampled.

The team also investigates the BMI ($\frac{\text{weight}}{\text{height}^2}$) from the in-body examination. This project uses four BMI categories of underweight (less than 18.5), normal (between 18.5 and 24.9), overweight (between 25 and 29.9), and obese (30 and higher). Although the rate of being overweight is very similar at both schools (37.9% for UTA and 35.6% for WMU), the obesity rate remains higher at WMU than UTA (33.9% vs. 25.9%).

**Table 5.2 Summary of UTA (n=58), WMU (n=59), total respondents (n=117), and US characteristics**

| Attribute | | Frequency (%) UTA | Frequency (%) WMU | Frequency (%) Total | Relative Frequency (%) U.S. |
|---|---|---|---|---|---|
| **Gender** | male | 31 (53.4%) | 43 (72.9%) | 74 (63.2%) | 49.2% |
| | female | 27 (46.6%) | 16 (27.1%) | 43 (36.8%) | 50.8% |
| **Age** | Under 18 | 0 (0.0%) | 2 (3.4%) | 2 (1.7%) | 22.8% |
| | 18 - 25 | 27 (46.6%) | 24 (40.7%) | 51 (43.6%) | 9.6% |
| | 26 - 49 | 25 (43.1%) | 30 (50.8%) | 55 (47.0%) | 32.8% |
| | 50 - 64 | 6 (10.3%) | 3 (5.1%) | 9 (7.7%) | 19.6% |
| | 65 and above | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 15.2% |
| **Race** | American Indian or Alaskan Native | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0.7% |
| | Asian / Pacific Islander | 22 (37.9%) | 14 (23.7%) | 36 (30.8%) | 5.3% |
| | Black or African American | 7 (12.1%) | 5 (8.5%) | 12 (10.3%) | 12.3% |
| | Hispanic American | 6 (10.3%) | 9 (15.3%) | 15 (12.8%) | 17.3% |
| | White / Caucasian | 22 (37.9%) | 31 (52.5%) | 53 (45.3%) | 62.0% |
| | Other | 1 (1.7%) | 0 (0.0%) | 1 (0.9%) | 2.5% |
| **Education** | some high school education, but no diploma | 0 (0.0%) | 4 (6.8%) | 4 (3.4%) | 12.6% |
| | high school graduate with a diploma or equivalent (for example: GED) | 1 (1.7%) | 3 (5.1%) | 4 (3.4%) | 27.7% |
| | some college credits, but no bachelor's degree | 13 (22.4%) | 11 (18.6%) | 24 (20.5%) | 31.0% |
| | bachelor's degree or higher | 44 (75.9%) | 41 (69.5%) | 85 (72.6%) | 28.7% |
| **Status (Position)** | Student | 39 (67.2%) | 46 (78.0%) | 85 (72.6%) | N/A |
| | Administration position | 6 (10.3%) | 3 (5.1%) | 9 (7.7%) | N/A |
| | University faculty | 6 (10.3%) | 3 (5.1%) | 9 (7.7%) | N/A |
| | Office worker | 7 (12.1%) | 4 (6.8%) | 11 (9.4%) | N/A |
| | Outdoor worker | 0 (0.0%) | 2 (3.4%) | 2 (1.7%) | N/A |
| | Not currently employed / home with family | 0 (0.0%) | 1 (1.7%) | 1 (0.9%) | N/A |
| **Income** | Less than $30,000 | 37 (63.8%) | 42 (71.2%) | 79 (67.5%) | 27.2% |
| | $30,000 - $50,000 | 14 (24.1%) | 10 (16.9%) | 24 (20.5%) | 18.2% |
| | $50,000 - $100,000 | 7 (12.1%) | 7 (11.9%) | 14 (12.0%) | 30.0% |
| | More than $100,000 | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 24.6% |
| **Having Driver's License** | Yes | 47 (81.0%) | 49 (83.1%) | 96 (82.1%) | 68.6% |
| | No | 11 (19.0%) | 10 (16.9%) | 21 (17.9%) | 31.4% |
| **No. of household motorized vehicles** | 0 | 10 (17.2%) | 8 (13.6%) | 18 (15.4%) | 8.7% |
| | 1 | 21 (36.2%) | 32 (54.2%) | 53 (45.3%) | 33.2% |
| | 2 | 18 (31.0%) | 12 (20.3%) | 30 (25.6%) | 37.1% |
| | 3+ | 9 (15.5%) | 7 (11.9%) | 16 (13.7%) | 21.0% |
| **Weight Status** | underweight | 1 (1.7%) | 2 (3.4%) | 3 (2.6%) | 1.5% |
| | normal | 20 (34.5%) | 16 (27.1%) | 36 (30.8%) | 27.7% |
| | overweight | 22 (37.9%) | 21 (35.6%) | 43 (36.8%) | 31.8% |
| | obese | 15 (25.9%) | 20 (33.9%) | 35 (29.9%) | 39.8% |
| **Moderate/Vigorous Physical Activity Level** | inactive | 20 (34.5%) | 22 (37.5%) | 42 (35.9%) | ----- |
| | insufficiently active | 12 (20.7%) | 10 (16.9%) | 22 (18.8%) | ----- |
| | active | 14 (24.1%) | 5 (8.5%) | 19 (16.2%) | ----- |
| | highly active | 12 (20.7%) | 22 (37.3%) | 34 (29.1%) | ----- |
| **Perceived Health** | very bad | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 2.2% poor |
| | bad | 4 (6.9%) | 2 (3.4%) | 6 (5.1%) | 7.7% fair |
| | fair | 15 (25.9%) | 13 (22.0%) | 28 (23.9%) | 23.7% good |
| | good | 35 (60.3%) | 34 (57.6%) | 69 (59.0%) | 31.1% very good |
| | excellent | 4 (6.9%) | 10 (16.9%) | 14 (12.0%) | 35.3% excellent |

The participants also identified the travel time they spent using each transportation mode when they traveled from home to work. The travel options include walk, bike, auto (as driver), auto (as passenger), wait/transfer, bus, rail, taxi, motorcycle, and other. They also marked their primary mode. The research team combined the transportation modes into three groups and aggregated the travel time by mode as follows:

- Active travel time (summation of travel times for walking and biking)
- Private vehicle travel time (summation of travel times for driving, being a passenger of auto, and motorcycle time)
- Transit travel time (summation of travel times for wait/transfer, bus, rail, and taxi)

Figure (5.2) compares the proportion of active, private, and transit use. A majority of the participants from both UTA (53%) and WMU (51%) primarily use (> 50%) private vehicles for commuting. A few UTA (3%) participants primarily use transit while almost 38% of UTA participants primarily use active transportation modes. At WMU, 22% of the participants primarily use transit while 17% primarily use active transportation modes. At UTA (35%) and WMU (54%), active transportation represents the primary secondary mode (10-50% use). Furthermore, transit represents the infrequent mode (<10% use) for most UTA (86%) and WMU (66%) participants. Public transit appears more popular for commuting in Kalamazoo than Arlington because the UTA shuttle bus represents the only fixed-route transit option in Arlington. In Kalamazoo, the transit system (called "KMetro") instead provides fixed-route transit service to Kalamazoo area cities including Kalamazoo, Portage, Parchment, Texas, and Oshtemo.

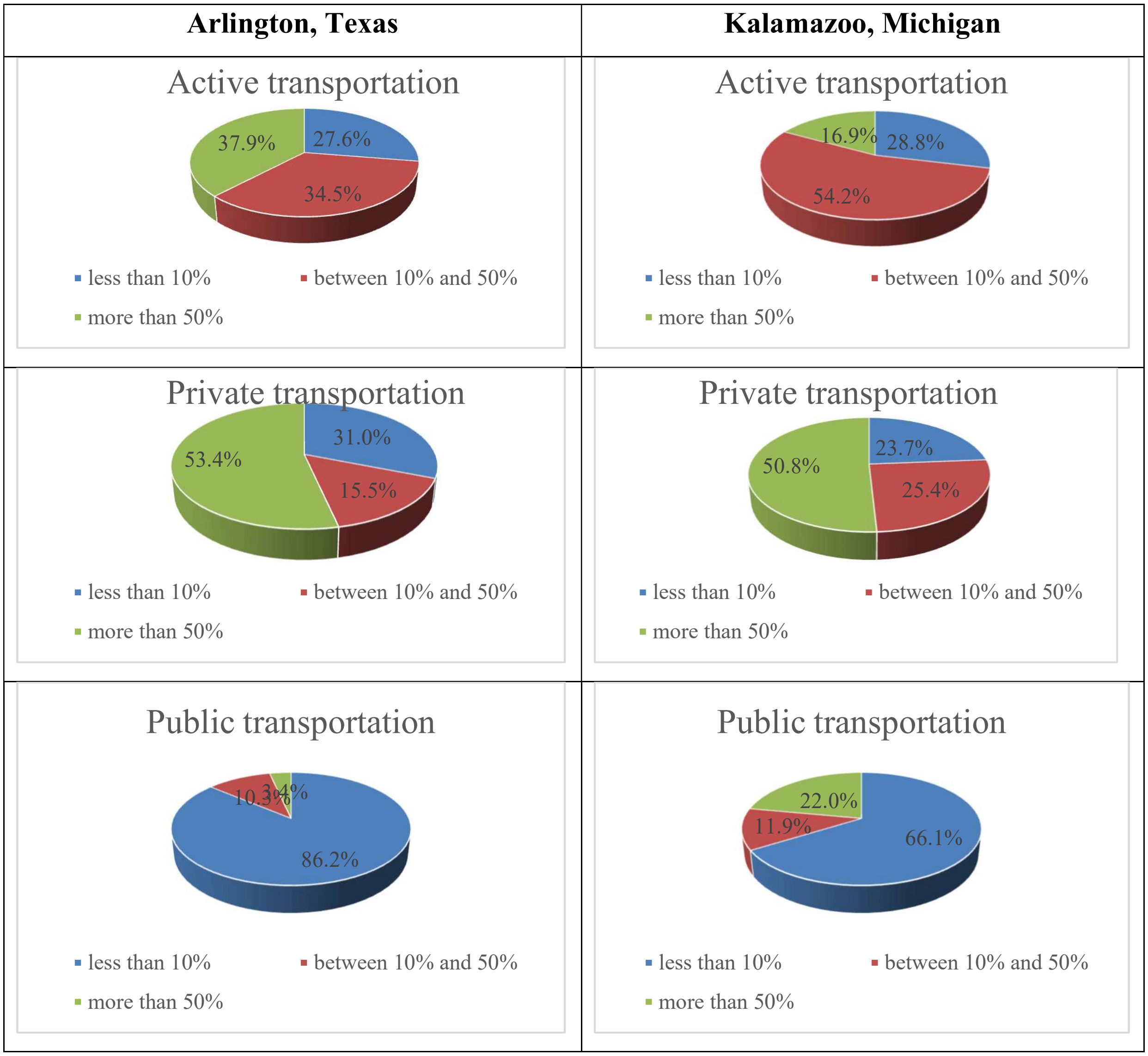



**Figure 5.2 Relative frequency of active, private, and public transportation use for the proportions of commuting time at Arlington (UTA) and Kalamazoo (WMU)**

## 5.6 Discussion and Cross Tabulation Analysis

Tables (5.3), (5.4), and (5.5) compare the perceived health, objective health (BMI), physical activity level, and main transport mode collected from the survey and in-body examination in a cross-tabulation format. Overall, the majority of the participants (59%) have a good assessment of health, followed by the fair (23.9%) and excellent (12%) groups. In addition, higher BMI appears to relate to lower perceived health. Highly active individuals show an excellent perception of health. Furthermore, active commuters show better perceived health than car commuters.

At UTA, the highest proportion of the participants (29.3%) has normal weight and good health, followed by overweight and good health (22.4%). Among the survey participants, 34% are overweight and 26% are obese; however, 68% of overweight participants believe their health is good or excellent, which is much higher than those who answered bad (0%) or fair (32%). In addition, only 27% of obese participants perceive their health is bad.

**Table 5.3 Comparisons among self-assessed health, weight, physical activity, and main transport mode for total population**

| | Perceived Health | | | | Total(%) |
|---|---|---|---|---|---|
| | Bad | Fair | Good | Excellent | |
| **BMI** | | | | | |
| Underweight | 0(0.0%) | 0(0.0%) | 3(2.6%) | 0(0.0%) | 3(2.6%) |
| Normal | 0(0.0%) | 3(2.6%) | 26(22.2%) | 7(6.0%) | 7(6.0%) |
| Overweight | 0(0.0%) | 12(10.3%) | 26(22.2%) | 5(4.3%) | 43(36.8%) |
| Obese | 6(5.1%) | 13(11.1%) | 14(12.0%) | 2(1.7%) | 35(29.9%) |
| **Total(%)** | 6(5.1%) | 28(23.9%) | 69(59.0%) | 14(12.0%) | **117(100.0%)** |
| **Physical Activity** | | | | | |
| Inactive | 5(4.3%) | 13(11.1%) | 23(19.7%) | 1(0.9%) | 42(35.9%) |
| Insufficiently Active | 0(0.0%) | 7(6.0%) | 13(11.1%) | 2(1.7%) | 22(18.8%) |
| Active | 0(0.0%) | 4(3.4%) | 12(10.3%) | 3(2.6%) | 19(16.2%) |
| Highly Active | 1(0.9%) | 4(3.4%) | 21(17.9%) | 8(6.8%) | 34(29.1%) |
| **Total(%)** | 6(5.1%) | 28(23.9%) | 69(59.0%) | 14(12.0%) | **117(100.0%)** |
| **Transport Mode** | | | | | |
| Walk/Bike | 1(0.9%) | 8(6.8%) | 22(18.8%) | 5(4.3%) | 36(30.8%) |
| Transit | 0(0.0%) | 4(3.4%) | 7(6.0%) | 4(3.4%) | 15(12.8%) |
| Car | 5(4.3%) | 16(13.7%) | 40(34.2%) | 5(4.3%) | 66(56.4%) |
| **Total(%)** | 6(5.1%) | 28(23.9%) | 69(59.0%) | 14(12.0%) | **117(100.0%)** |

The researchers compare the perceived health and their physical activity. Overall, 20% and 12% of population has inactive and insufficiently active physical activity, respectively. Among them, 60% indicate that their health is good or excellent, which is much higher than those who answered that their health is bad or fair. The results show that 79% of the active and 75% of the highly active individuals indicate a good or excellent health condition. The percentage of people who engage with the lowest level of physical activity and indicate good perception of health (19%) is almost four times higher than those who perceive bad health with the same activity level (5.2%). While 62.5% of car commuters have good/excellent perceived health, 75% of active commuters evaluate their health to be good and excellent. The overall results of the UTA sample show that perceived health does not always align with objective health measures such as BMI and physical activity level. However, the sample also shows that transportation-related physical activity relates to better perceived health.

**Table 5.4 Comparisons among self-assessed health, weight, physical activity, and main transport mode for UTA**

| | **Perceived Health** | | | | **Total(%)** |
|---|---|---|---|---|---|
| | Bad | Fair | Good | Excellent | |
| **BMI** | | | | | |
| Underweight | 0(0.0%) | 0(0.0%) | 1(1.7%) | 0(0.0%) | 1(1.7%) |
| Normal | 0(0.0%) | 2(3.4%) | 17(29.3%) | 1(1.7%) | 20(34.5%) |
| Overweight | 0(0.0%) | 7(12.1%) | 13(22.4%) | 2(3.4%) | 22(37.9%) |
| Obese | 4(6.9%) | 6(10.3%) | 4(6.9%) | 1(1.7%) | 15(25.9%) |
| **Total(%)** | 4(6.9%) | 15(25.9%) | 35(60.3%) | 4(6.9%) | **58(100.0%)** |
| **Physical Activity** | | | | | |
| Inactive | 3(5.2%) | 6(10.3%) | 11(19.0%) | 0(0.0%) | 20(34.5%) |
| Insufficiently Active | 0(0.0%) | 4(6.9%) | 6(10.3%) | 2(3.4%) | 12(20.7%) |
| Active | 0(0.0%) | 3(5.2%) | 9(15.5%) | 2(3.4%) | 14(24.1%) |
| Highly Active | 1(1.7%) | 2(3.4%) | 9(15.5%) | 0(0.0%) | 12(20.7%) |
| **Total(%)** | 4(6.9%) | 15(25.9%) | 35(60.3%) | 4(6.9%) | **58(100.0%)** |
| **Transport Mode** | | | | | |
| Walk/Bike | 1(1.7%) | 5(8.6%) | 15(25.9%) | 3(5.2%) | 24(41.4%) |
| Transit | 0(0.0%) | 1(1.7%) | 1(1.7%) | 0(0.0%) | 2(3.4%) |
| Car | 3(5.2%) | 9(15.5%) | 19(32.8%) | 1(1.7%) | 32(55.2%) |
| **Total(%)** | 4(6.9%) | 15(25.9%) | 35(60.3%) | 4(6.9%) | **58(100.0%)** |

**Table 5.5 Comparisons among self-assessed health, weight, physical activity, and main transport mode for WMU**

| | Perceived Health | | | | Total(%) |
|---|---|---|---|---|---|
| | Bad | Fair | Good | Excellent | |
| **BMI** | | | | | |
| Underweight | 0(0.0%) | 0(0.0%) | 2(3.4%) | 0(0.0%) | 2(3.4%) |
| Normal | 0(0.0%) | 1(1.7%) | 9(15.3%) | 6(10.2%) | 16(27.1%) |
| Overweight | 0(0.0%) | 5(8.5%) | 13(22.0%) | 3(5.1%) | 21(35.6%) |
| Obese | 2(3.4%) | 7(11.9%) | 10(16.9%) | 1(1.7%) | 20(33.9%) |
| **Total(%)** | 2(3.4%) | 13(22.0%) | 34(57.6%) | 10(16.9%) | **59(100.0%)** |
| **Physical Activity** | | | | | |
| Inactive | 2(3.4%) | 7(11.9%) | 12(20.3%) | 1(1.7%) | 22(37.3%) |
| Insufficiently Active | 0(0.0%) | 3(5.1%) | 7(11.9%) | 0(0.0%) | 10(16.9%) |
| Active | 0(0.0%) | 1(1.7%) | 3(5.1%) | 1(1.7%) | 5(8.5%) |
| Highly Active | 0(0.0%) | 2(3.4%) | 12(20.3%) | 8(13.6%) | 22(37.3%) |
| **Total(%)** | 2(3.4%) | 13(22.0%) | 34(57.6%) | 10(16.9%) | **59(100.0%)** |
| **Transport Mode** | | | | | |
| Walk/Bike | 0(0.0%) | 3(5.1%) | 7(11.9%) | 2(3.4%) | 12(20.3%) |
| Transit | 0(0.0%) | 3(5.1%) | 6(10.2%) | 4(6.8%) | 13(22.0%) |
| Car | 2(3.4%) | 7(11.9%) | 21(35.6%) | 4(6.8%) | 34(57.6%) |
| **Total(%)** | 2(3.4%) | 13(22.0%) | 34(57.6%) | 10(16.9%) | **59(100.0%)** |

WMU participants feature fewer people with bad perceived health but more people with excellent perceived health, compared to UTA. The number of people having an excellent health perception and normal weight is two and six times higher than the overweight and obese people in the same health category. The relationship between perceived health and BMI seems similar at both schools. For example, 94% and 90% of the WMU and UTA respondents in the normal weight category have a good or excellent assessment of health. Nevertheless, more WMU participants belong to the normal weight with excellent health category than UTA (10.2% vs. 1.7%). While the UTA sample has almost three times more active participants than WMU (24.1% vs. 8.5%), more WMU participants fall in the highly active category. Tables (5.4) and (5.5) indicate that 41.4% of UTA participants are active commuters while 22% of the WMU participants use public transit. The percentage of active commuters having a good/excellent health perception is two times higher in Arlington than Kalamazoo (31.1% vs. 15.3%).

# Chapter 6: Transportation User Activity and Trip Recognition

## 6.1 Introduction

Traveler behavior research requires identifying individuals' activity and trip information. Automatic and digital methods with Global Positioning System (GPS) data logs offer the promise of higher quality data than traditional traveler behavior data collection. Therefore, this section develops and evaluates strategies that recognize the activity and trip with different thresholds of spatiotemporal change by applying a Geohash clustering approach, a GIS-based approach, and a combined approach by integrating the Geohash and GIS systems. The study develops and implements these approaches for activity only, trip only, and sequential activity-trip recognition with GPS data as a case study in Kalamazoo, Michigan.

## 6.2 Data and Methodology

This section discusses the major data collection methods and the approaches this study applies to recognize activity/trip information.

### 6.2.1 Data Collection

This research collected a GPS dataset of more than thirty users' daily activities over a 12-month period from January to December of 2018. Each respondent had an average of about 100,000 records for the study period. With missing values being eliminated from the records, the final sample included nearly 10 million GPS records. Each record in the dataset represented a GPS signal that was captured consecutively in every 1-second interval by the Android GPS device and contained information on index, date and time (ET), latitude, longitude, altitude (m), speed (m/h), distance (m), and satellite information. The researchers used a moderate GPS accuracy (100) for user trip/activity recognition.

### 6.2.2 Research Methodology

This research developed three different approaches to recognize user activity/trip from GPS data logs, including a Geohash Clustering Approach, a GIS-based Approach, and a Combined

Geohash-GIS Approach. The research team developed different models based on dwelling times of 5, 8, and 10 minutes for each of the approaches.

#### *6.2.2.1 Geohash Clustering Approach:*

Geohash is a public domain geocoding system that encodes a geographic location into a short string of letters and digits (Wikipedia Contributors, 2018). It maintains a hierarchical spatial data structure that subdivides space into buckets of grid shape by using latitude and longitude points (Sing el al. 2017). This research clusters the GPS points based on the Geohash approach. A Geohash algorithm hashes all points during a day to cluster the adjacent GPS points. Increasing the number of Geohash string characters (precision) increases the neighboring points that incorporate into the same cluster. Based on the geographic extent of the user data, the researchers test 5, 6, and 7-character precision at a spatial resolution level to cluster the adjacent points. Therefore, the 5-character precision develops a 4.9 km × 4.9 km area as a cluster, 6-character precision creates a 1.22 km × 0.61 km area as a cluster, and 7-character precision uses a 152.9 m × 152.3 m area as a cluster. Figure (6.1) shows an example of the Geohash clustering technique where the data include the latitude, longitude, time, and activity type.

**Figure 6.1 Geohash clustering example**

The technique converts and hashes the latitude and longitude points into a Geohash format using a specific Geohash length. The clustering technique aggregates all similar points with the same Geohash and adds them to a cluster labeled by the Geohash string. the process calculates the duration by subtracting the start and the end times of the corresponding trip/activity.

#### *6.2.2.2 GIS-based Approach*

This approach uses a GIS-based boundary shapefile to detect and recognize a user's daily activity/trip. The research team uses the boundary shapefile from open street map for this analysis, which is readily available on the Web. The ArcPy code in the Python language environment can reverse geocoding to get spatial boundary information from GPS points. Each of the GPS points in the user's trajectory return the spatial location information, which may be aggregated to get the duration of each activity or trip. The study tests the dwelling time as 5, 8, and 10 minutes to keep consistent analysis with other approaches. During this approach, the researchers develop the code based on "identity analysis tool" using ArcGIS 10.5.

This GIS-based method uses a Kalman Filter to smooth the GPS data and applies a 10-foot (3 meter) buffer for polygon and line features to incorporate the outlier points. If the spatial points for a user appear inside the boundary (e.g., home, school, or market) for more than the specified dwelling time, the GIS-based approach defines the data point cluster as an activity rather than a trip. The method calculates the activity duration by calculating the difference between the first and last points of any specific boundary feature. After this step, the activity/trip classifier aggregates every duration and generates the final outcome of the activity or trip classification.

#### *6.2.2.3 Combined Geohash-GIS Approach*

This study develops a Combined Geohash-GIS approach by integrating the GIS-based and Geohash Clustering approaches. Figure (6.2) shows a schematic diagram for the combined approach. Geohash alone could not explain the internal activity/trip analysis within each cluster if the Geohash precision level is too big; correspondingly, it generates multiple sub-sections for a single trip/activity if it considers a minimal Geohash precision level. The GIS-based approach lacks the flexibility to incorporate all of the points into the boundary shapefile because of GPS data outliers. The combined approach seeks to overcome these problems. This method uses a two-stage algorithm. For the first stage, a Geohash clustering algorithm uses the approach methodology described in section 6.2.2.1. The research team uses a Geohash precision level-6 to hash the entire study area with a Geohash precision level-7 inside each block of level-6 precision to identify and check the inner activities among them. The identification of the internal activities triggers the second stage of the algorithm by applying the GIS-based approach inside each internal hashed activity. This stage applies a reverse geocoding algorithm with shapefile (e.g., object boundary, line features, etc.) information to accurately identify the user activity or trip information. This combined approach detects the inner activities/trip with the correct precision level and validates the activity/trip by applying the GIS-based boundary shapefiles. The researchers also reduce the outlier effect since the Geohash considers those outliers by cluster aggregation at the final stage. Thus, the combined approach minimizes the errors to identify the user activity/trip accurately.

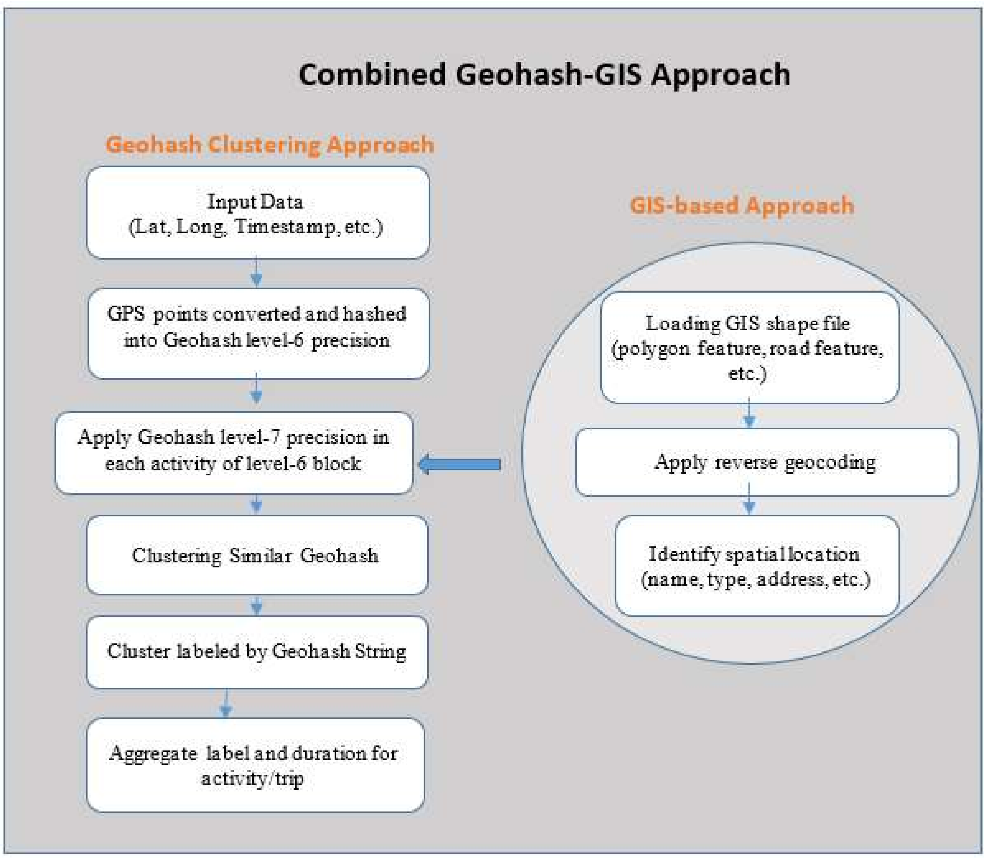



**Figure 6.2 Schematic diagram for Combined Geohash and GIS-based approaches**

### *6.2.2.4 Sample Dataset*

The testing dataset randomly collects 90 samples (90 different days/dates of user activities) from all users for the study period of 12 months. The testing dataset contained about 50,000 GPS data records. The study considers three different testing scenarios based on specific tolerance levels, including simple, moderate, and critical testing to identify trip only, activity only, and sequential activity-trip recognition analysis. From the sample data, the research team randomly selects 20% for simple testing, 30% for moderate testing, and 50% for critical testing. The three different testing scenarios identify the accuracy of activity/trip recognition using the following standard and tolerance levels:

- Simple testing with 99.7% confidence bounds (tolerance range = ±3E from true mean value)
- Moderate testing with 95% confidence bounds (tolerance range = ±1.96E from true mean value)
- Critical testing with 68.3% confidence bounds (tolerance range = ±E from true mean value)
- Where, $E=\sigma/(\sqrt{N})$ …. $\sigma$ is for standard deviation and N is sample size.

The research team prepares the true dataset based on the actual duration of activities and trips performed by users. The user's feedback data together with GPS tracking logs were displayed and carefully observed in a GIS map to identify the true data. The activity or trip duration usually varies because of the user's behavioral pattern for different activity types, which seems somewhat problematic for calculating the true mean value for the observed dataset. For example, in terms of a mean value of trip duration, some of the trips were very long (e.g., 1 hr. or 2 hrs.) and some were very short (e.g., 10-15 minutes). So, it is inappropriate to calculating the mean value by combining those long and short trips, which will make our results incorrect. Therefore, we sorted the sample data with similar types of duration times and considering some group criteria (e.g., less than 15 minutes, 15-30 minutes, 31-45 minutes, 46-60 minutes, 61-80 minutes, more than 80 minutes, etc.) and calculated the mean value for that sample group.

The accuracy for user's activity/trip recognition was calculated based on equation (6.1) for different scenario i, as critical, moderate, and simple testing. Here, the accuracy means that the defined model accurately recognizes the user trip or activity in such a way that the test/observe value is within 68%, 95%, and 99% confidence bounds of true mean value for critical testing, moderate testing, and simple testing, respectively.

$$\boldsymbol{Accuracy\ in\ activity/trip\ recognition\ (A_i)\ = \frac{accurately\ classified\ data\ for\ scenerio\ i}{total\ data\ for\ scenerio\ i} * 100} \qquad \mathbf{(6.1)}$$

In addition to the testing data and accuracy measurement for activity/trip recognition, the study evaluates the approaches based on model training and prediction accuracy for the whole dataset. The training accuracy checks the misclassification error rate (MER) and calculates the accuracy based on equation (6.2). The misclassification error rate checks whether the model accurately classifies the trip as a trip and an activity as an activity, or vice versa.

$$\boldsymbol{Model\,Training\,Accuracy, A_{training} = (1 - MER) * 100} \tag{6.2}$$

The study compares the different approaches based on model prediction accuracy and evaluates the prediction accuracy based on equation (6.3) by calculating the Mean Absolute Percentage Error (MAPE) by observing the absolute differences between the actual and predicted user's activity duration.

$$\boldsymbol{Model\,Prediction\,Accuracy, A\_prediction = 1 - MAPE} \tag{6.3}$$

Where:

$$\boldsymbol{MAPE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\frac{observed\ activity}{trip\ duration} - \frac{predicted\ activity}{trip\ duration}}{\frac{observed\ activity}{trip\ duration}}\right) * 100} \tag{6.4}$$

Where:

N: is the total number of observations.

## 6.3 Analysis and Numerical Results

The researchers analyzed the individual approach outputs and compare the three proposed approaches across the 15 models based on different dwell times (5, 8, and 10 minutes) and geohash levels. The model evaluation analyzes the accuracy of the user activity/trip recognition.

### 6.3.1 Geohash Clustering Approach

This study develops nine models using three Geohash character levels (character sizes 5, 6, and 7) and three different dwelling times (5, 8, and 10 minutes). The researchers compute the confusion matrix based on the output for each of the nine models to identify the misclassification between trip and activity. The confusion matrix shows the accuracy as to whether the predicted activity is actually an activity, or the predicted trip is actually a trip for the specific model. For example, Geohash-5 with Dwell-5 minute model returns the following result: $\begin{pmatrix} & \boldsymbol{activity} & \boldsymbol{trip} \\ \boldsymbol{activity} & 123 & 11 \\ \boldsymbol{trip} & 52 & 102 \end{pmatrix}$, which translates into the model accurately classifying the activity as activity and trip as trip with an accuracy of 78.1%.

Table (6.1) indicates that the Geohash-6 and Geohash-7 clustering shows better accuracy in comparison to the Geohash-5 clustering. The Geohash-5 clustering does not deliver a good outcome since it covers a larger block of geographical area to cluster the GPS data where some of the trip/activities were overlooked and misclassified. The Geohash-6 clustering model with a dwell time of 5 minutes provides the best accuracy (89.23%).

**Table 6.1 Accuracy in activity/trip recognition for Geohash clustering approach**

| | | **Activity/trip Recognition Accuracy (Percentage)** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Geohash-5 | | | Geohash-6 | | | Geohash-7 | | |
| | | Dwell 5 min | Dwell 8 min | Dwell 10 min | Dwell 5 min | Dwell 8 min | Dwell 10 min | Dwell 5 min | Dwell 8 min | Dwell 10 min |
| **Model Accuracy ($A_{training}$) based on misclassification error rate of training data** | | 78.11 | 77.71 | 78.74 | 89.23 | 85.71 | 83.63 | 80.91 | 80.92 | 77.08 |
| **Accuracy ($A_i$) of different testing scenarios based on sample dataset** | | | | | | | | | | |
| **Critical testing** | Sequential activity-trip | 34.62 | 38.46 | 37.18 | 46.15 | 38.46 | 33.33 | 20.51 | 17.95 | 16.67 |
| | Activity | 53.85 | 46.15 | 43.59 | 51.28 | 43.59 | 41.03 | 17.95 | 15.38 | 12.82 |
| | Trip | 15.38 | 30.77 | 30.77 | 41.03 | 33.33 | 25.64 | 23.08 | 20.51 | 20.51 |
| **Moderate testing** | Sequential activity-trip | 50.68 | 52.05 | 50.68 | 57.53 | 56.16 | 49.32 | 52.05 | 50.68 | 43.84 |
| | Activity | 69.70 | 69.70 | 63.64 | 60.61 | 63.64 | 51.52 | 51.52 | 48.48 | 45.45 |
| | Trip | 33.33 | 35.90 | 38.46 | 53.85 | 48.72 | 46.15 | 53.85 | 53.85 | 43.59 |
| **Simple testing** | Sequential activity-trip | 75.18 | 75.64 | 73.08 | 80.77 | 75.64 | 75.64 | 67.95 | 67.95 | 67.95 |
| | Activity | 73.77 | 73.77 | 73.77 | 77.05 | 75.41 | 77.05 | 65.57 | 65.57 | 65.57 |
| | Trip | 76.32 | 75.00 | 72.37 | 80.26 | 75.00 | 75.00 | 68.42 | 68.42 | 68.42 |

The evaluation tests the accuracy for three different classifications (i.e. activity only, trip only, and sequential activity/trip) based on different Geohash character clustering levels and dwell times. Table (6.1) shows the outcome of the models based on different testing scenarios. Overall, the Geohash-6 with dwell time 5 min shows the best result with about 50% accuracy for critical testing, 60% accuracy for moderate testing, and 80% accuracy for simple testing. Since the critical testing (68% confidence interval from mean value) considers a narrow interval from mean value, it shows less accuracy in activity/trip recognition in comparison to other testing scenarios. Table 6.1 indicates that the activity recognition accuracy appears better than trip recognition accuracy for all testing scenarios except simple testing. The Geohash-6 clustering with dwell time 5 min showed higher accuracy (more than 80%) to recognize the user activity successfully based on simple testing scenario.

### 6.3.2 GIS-based Approach

The study develops three different models to recognize activity/trip by applying the GIS-based approach with 5, 8, and 10-minute dwell times. Table (6.2) shows the model Accuracy ($A_{training}$) based on the misclassification error rate of the training data. The GIS-based model with a dwell time of 5 minutes shows good accuracy (70.41%) in comparison to the other GIS-based models. The classification accuracy for the GIS-based models remains lower than the Geohash models.

**Table 6.2 Accuracy in activity/trip recognition for GIS-based approach**

| | | Activity/trip Recognition Accuracy (Percentage) | | |
|---|---|---|---|---|
| | | Dwell-5 min | Dwell-8 min | Dwell-10 min |
| **Model Accuracy ($A_{training}$) based on misclassification error rate of training data** | | 70.41 | 69.72 | 70.15 |
| **Accuracy ($A_i$) of different testing scenarios based on sample dataset** | | | | |
| **Critical testing** | Sequential activity-trip | 26.92 | 33.33 | 34.62 |
| | Activity | 35.90 | 38.46 | 41.03 |
| | Trip | 17.95 | 28.21 | 28.21 |
| **Moderate testing** | Sequential activity-trip | 26.03 | 31.51 | 35.62 |
| | Activity | 42.42 | 48.48 | 51.52 |
| | Trip | 10.26 | 15.38 | 20.51 |
| **Simple testing** | Sequential activity-trip | 63.50 | 51.28 | 51.28 |
| | Activity | 60.66 | 52.46 | 54.10 |
| | Trip | 65.79 | 51.32 | 51.32 |

The model evaluation assesses the outcome accuracy for the three GIS-based models for the activity only, trip only, and sequential activity/trip recognition testing scenarios. Table (6.2) shows that the dwell time 10 minutes' model outperforms the other models for the critical and moderate testing; however, the dwell time 5-minute model shows better accuracy for the simple testing case. In general, the dwell time 10-minute model works better than the other GIS-based approaches with an overall accuracy of about 50% for simple testing scenarios, which was less than the accuracy of the Geohash clustering approach.

### 6.3.3 Combined GIS and Geohash Approach

The study proposes three combined approach by using different dwell times. Table (6.3) shows the accuracy based on a confusion matrix, where all models have good accuracy (above 90%), which outperforms the previous models.

Table (6.3) shows the accuracy of the models based on a combined approach, and they all show very good accuracy to recognize activity only, trip only, and sequential activity/trip for all testing scenarios in comparison to other approaches. Among these models, the dwell time 5-minute model shows the best accuracy (above 90%) for critical, moderate, and simple testing.

**Table 6.3 Accuracy in activity/trip recognition for Combined Geohash-GIS approach**

| | | Activity/trip recognition Accuracy (Percentage) | | |
|---|---|---|---|---|
| | | Dwell-5 min | Dwell-8 min | Dwell-10 min |
| **Model Accuracy ($A_{training}$) based on misclassification error rate of training data** | | 94.10 | 92.12 | 92.01 |
| **Accuracy ($A_i$) of different testing scenarios based on sample dataset** | | | | |
| **Critical testing** | Sequential activity-trip | 93.59 | 93.59 | 89.74 |
| | Activity | 87.18 | 87.18 | 79.49 |
| | Trip | 100.0 | 100.0 | 100.0 |
| **Moderate testing** | Sequential activity-trip | 91.78 | 87.67 | 89.04 |
| | Activity | 81.82 | 72.73 | 75.76 |
| | Trip | 100.0 | 100.0 | 100.0 |
| **Simple testing** | Sequential activity-trip | 91.97 | 91.44 | 91.44 |
| | Activity | 83.61 | 72.13 | 77.05 |
| | Trip | 98.68 | 98.68 | 98.68 |

## 6.4 Discussion and Comparison of Different Approaches

This section discusses the predicted values (duration in minutes) of user activity/trip based using different approaches. The analysis compares models based on the Geohash-6 clustering approach, GIS-based and combined Geohash-GIS approach under different dwell time scenarios. The comparison uses the Geohash-6 approach because it shows better accuracy than the other Geohash precision levels. The evaluation also considers dwell time because it remains a key feature to identify the activity/trip.

The research team compares the predicted sequential activity-trip duration values with the true duration values by developing a scatter plot diagram in Figure (6.3). The dwell time 5-minute models show a good relationship between the true and predicted values with a r-square value of above 0.8 for all models, but the combined approach shows the best accuracy with a higher r-square value (above 0.9). Therefore, the combined approach can explain more than 90% of the data variability of the predicted activity-trip duration.

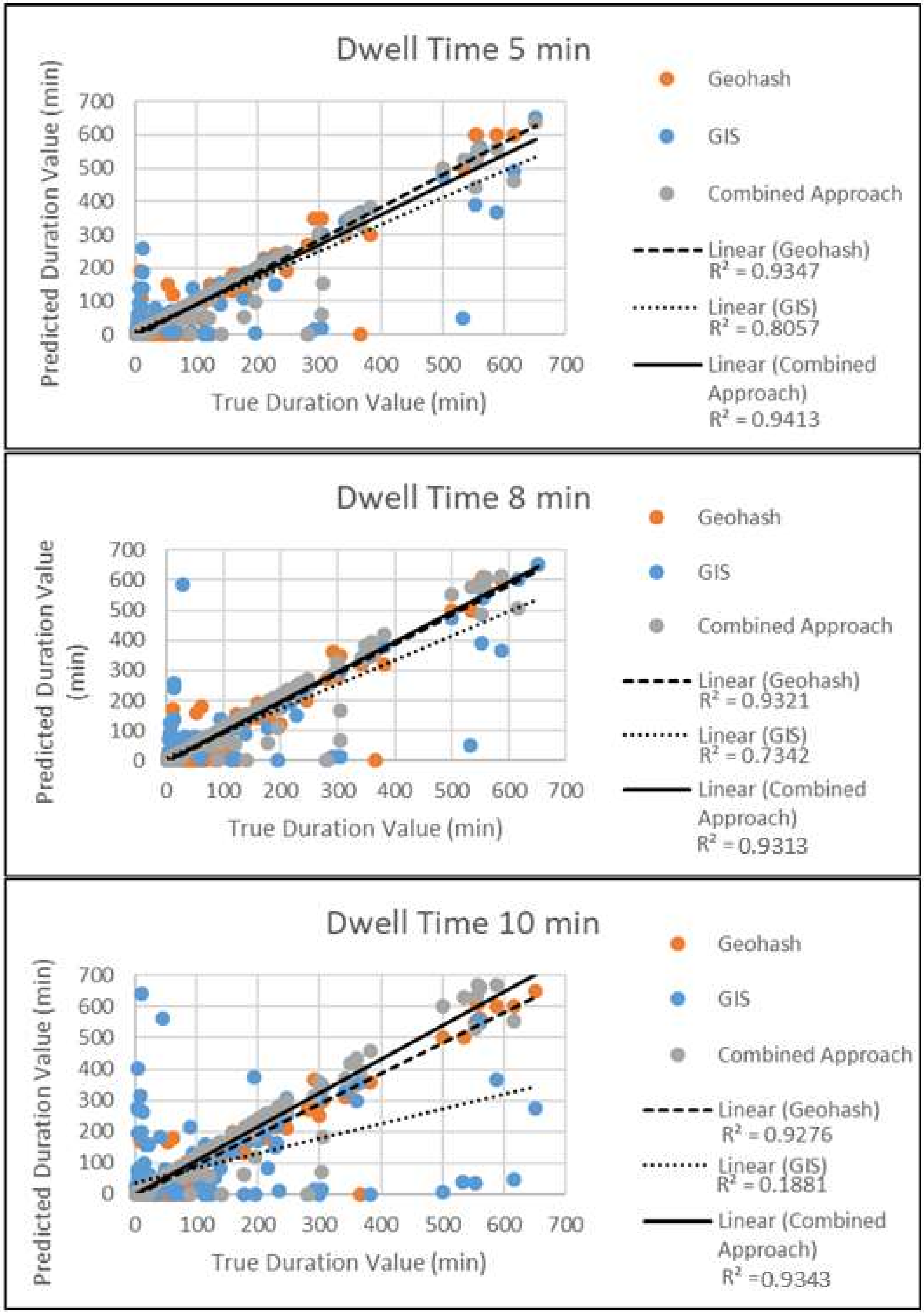



**Figure 6.3 Predicted accuracy comparisons for different approaches**

The Mean Absolute Percentage Error (MAPE) shows the actual deviation of predicted activity/trip duration from the true value. Figure 6.4 shows the MAPE values for the activity only, trip only, and sequential activity-trip accuracy-based cases. All of the models based on the GIS-based approach show poor accuracy for predicting the activity/trip duration. The Geohash-6 clustering shows better accuracy in the sequential activity-trip duration in comparison to GIS-based methods. However, the combined approach with a dwell time of 5 minutes shows the best accuracy, where the MAPE values remain less than 19 percent.

| Different testing Scenarios | MAPE | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Geohash-6 | | | GIS-based Approach | | | Combined Approach | | |
| | 5 min | 8 min | 10 min | 5 min | 8 min | 10 min | 5 min | 8 min | 10 min |
| activity/trip | 30.16 | 28.87 | 33.15 | 65.91 | 62.98 | 67.11 | 12.70 | 13.48 | 18.81 |
| activity | 23.81 | 22.82 | 21.22 | 49.72889 | 58.59237 | 58.29276 | 18.65 | 20.95 | 27.83 |
| trip | 42.77 | 40.97 | 51.43 | 44.83809 | 34.885 | 46.35985 | 15.00 | 20.01 | 25.21 |

**Figure 6.4 Prediction accuracy in activity/trip recognition based on MAPE**

This research also compares the accuracy in user activity/trip recognition by developing individual Receiver Operating Characteristics (ROC) for different models on different approaches. Figure (6.5) shows the diagnostic test based on ROC curves, where the relationship between sensitivity and false-positive rate (FPR) are explained for the predicted activity/trip values by different models. The study also calculates the area under curve (AUC) based on different dwell times for different approaches. The combined approach with a dwell time 5 minutes shows the highest AUC value with 0.878.

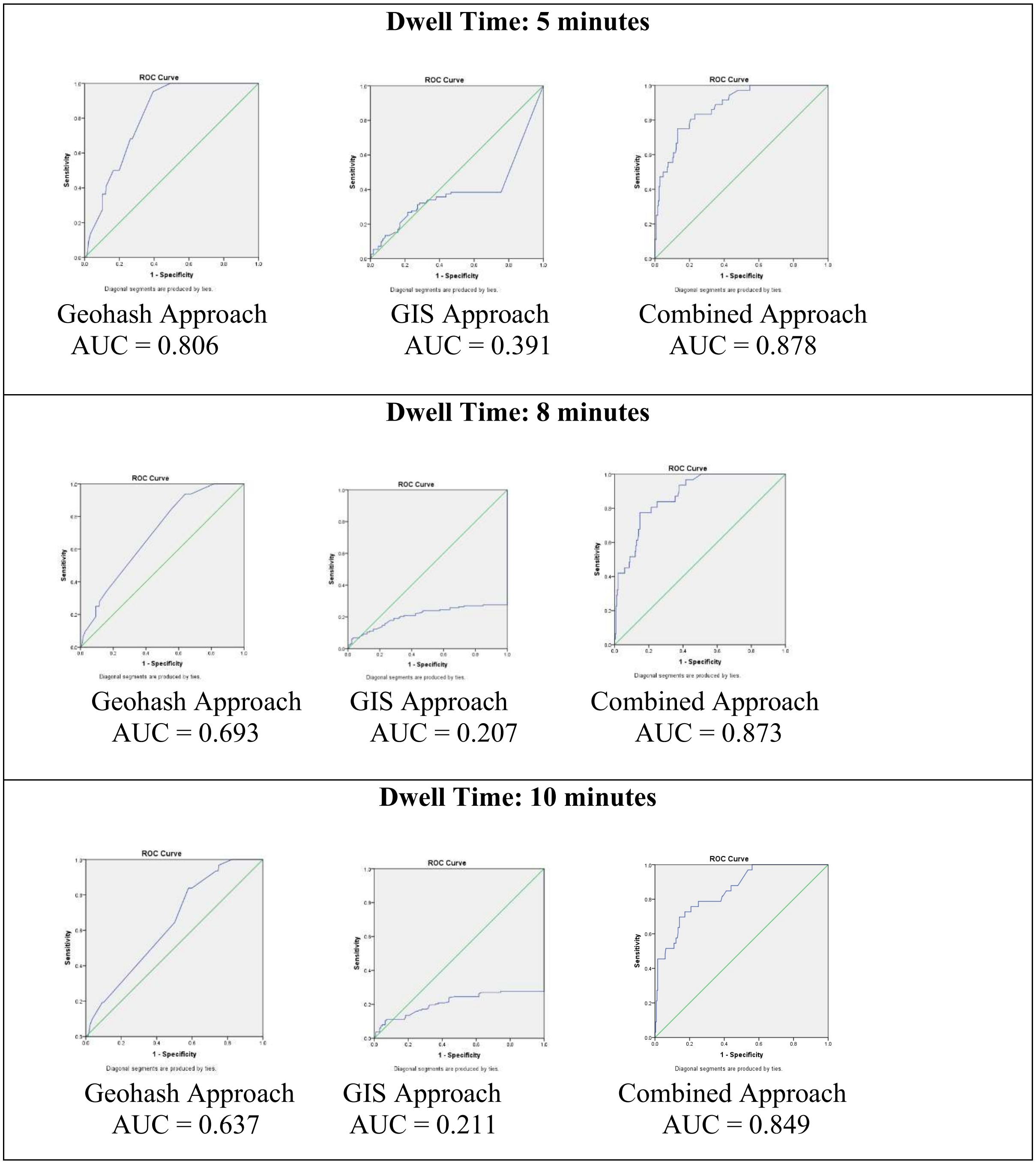



**Figure 6.5 Predicted accuracy comparisons based on ROC curve**

## 6.5 Summary and Concluding Remarks

This research studies three different approaches for user activity and trip recognition based on GPS log data. For Geohash clustering analysis, Geohash precision level-6 with dwell time 5 min shows the best accuracy for user activity/trip detection. For the GIS-based approach, a dwell time of 10 minutes has the best performance. However, the Combined Geohash-GIS approach with a dwell time of 5 minutes achieves the best overall accuracy. The proposed combined approach may significantly enhance the efficiency and accuracy of GPS travel survey data by correctly recognizing user activity and trip patterns. This approach can serve as a foundation for a future system of full-scale travel information identification with GPS data. The combined approach could contribute to improvements in the modeling and analysis of travel behavior. The proposed approach seems easy to replicate and could contribute to transportation and city planning research by replacing the traditional survey methods with automatic recognition of user travel patterns.

# Chapter 7: Transportation Mode Detection

## 7.1 Introduction

This section describes strategies to detect different transportation modes from smartphone and smartwatch data. The study develops a series of different machine learning algorithms to predict and detect transportation modes. This section highlights the benefits of monitoring human vital activities for mode detection.

## 7.2 Data and Methodology

### 7.2.1 Data Description

The study uses data from a total of 120 participants to predict the transportation mode. The PASTA application transfers the data, which includes information about the date, time, latitude, longitude, altitude, heart rate (HR), calories, steps, elevation, and metabolic equivalent tasks (METs), from the phone and the watch to a server.

### 7.2.2 Measurement of Physical Activity

This study relies on the heart rate as a method to measure physical activity. This study uses a smartwatch (Fitbit Charge 2 or Charge 3 for measuring and recording physical activity data. The study requires knowing the values of the: 1) resting heart rate; 2) age-predicted maximum heart rate (i.e. [220-age]); 3) and heart rates with physical activities recorded each minute. The Fitbit smartwatch provides each participant's resting heart rate once per day, and it records the heart rate of physical activity every minute.

The study adapts the Karvonen equation or "Heart rate (HR) Reserve Method" to automate the measurement for the changes that may occur in the physical activities of the person and is known as the. The original equation of the Karvonen method, shown in equation (7.1), presents the target range of HR based on the predetermined value of a person's percent intensity for physical activity.

$$\boldsymbol{Target\ HR_{range}} = ([\boldsymbol{HR_{max}} - \boldsymbol{HR_{rest}}] \times \boldsymbol{Percent\ intensity}) + \boldsymbol{HR_{rest}} \qquad \textbf{(7.1)}$$

The researchers adjust the Karvonen equation formula by substituting the real value of the %HRR recorded from the Fitbit smartwatch, which changes in each minute, for the target HR range. Equation (7.2) illustrates the modified formula for measuring the physical activity intensity per minute:

$$\%\boldsymbol{HRR} = \frac{(\boldsymbol{HR_{act.}} - \boldsymbol{HR_{rest}})}{(\boldsymbol{HR_{max}} - \boldsymbol{HR_{rest}})} \times \mathbf{100} \tag{7.2}$$

Where:

$HR_{act.}$: Actual Heart Rate (from Fitbit)

$HR_{rest}$: Resting Heart Rate (from Fitbit)

$HR_{max}$= (220 – Age)

The physical activity intensity value when calculated per minute (PAM) uses Equation (7.3):

$$\boldsymbol{PAM_i} = \int_{i_{start}}^{i_{end}} \%\boldsymbol{HRR}\,.\,\boldsymbol{dt} \tag{7.3}$$

### 7.2.3 Training and Verification Algorithms

This study uses four algorithms (Extreme Gradient Boosting, Random Forest, Support Vector Machine, and Artificial Neural Network) to extract the classification models and compares their performance. The following provides a brief explanation of these algorithms:

***Extreme Gradient Boosting (XGBoost):***

The XGboost is a machine learning algorithm, which can be used for supervised learning tasks such as Regression, Classification, and Ranking. XGBoost produces a predictive model in a set of weak prediction models, usually decision trees. XGBoost constructs the model in a stage-wise fashion as other boosting methods do, and it generalizes them by enabling optimization of an arbitrary differentiable loss function (Chen et al., 2016).

Basically, the training is done using an "additive strategy": Given a molecule $i$ with a vector of descriptors $x_i$ , a tree ensemble model uses $K$ additive functions to predict the output (Sheridan et al., 2016 ).

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}} = \boldsymbol{\phi}(\boldsymbol{x_i}) = \sum_{\boldsymbol{k}=\boldsymbol{1}}^{\boldsymbol{k}} \boldsymbol{f_k}(\boldsymbol{x_i}), \qquad \boldsymbol{f_k} \in \boldsymbol{\mathcal{F}} \tag{7.4}$$

Where:

$\boldsymbol{\mathcal{F}}$: is the set of all possible regression trees.

$\boldsymbol{f_k}$: is a function at each of the k steps maps the descriptor values in $\boldsymbol{x_i}$ to a certain output

***Random Forest (RF) algorithm:***

The RF is a method designed for classification, which creates a forest of trees where each tree represents a set of training data. Every tree in the forest is given the opportunity to grow as far as possible and without pruning. The fruits of the prediction attempts are then harvested for the purpose of evaluating the accuracy provided by the algorithm. Therefore, the number of trees and the number of variables is the two most important parameters in RF (Zhang et al., 2018).

Random forest using bagging ensemble algorithim to produce unbiased models with low variance. The random forest procedure can be summarized as follows (Akinkunmi et al. 2019);

1. For $b = 1$ to $B$:
   (a) A bootstrap sample $\boldsymbol{Z}^*$ of size N from the training data is drawn.
   (b) Grow a random-forest tree $T_b$ to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size $n_{min}$ is reached.
      i. Select $m$ variables at random from the $p$ variables.
      ii. Pick the best split-point among the m.
      iii. Split the node into two daughter nodes.
2. Output the ensemble of trees $\{T_b\}_1^B$

A prediction at a new point $x$ can be accomplished by;

Let $\widehat{C_b}(x)$ be the class prediction of the $b$th random-forest tree.

Then $\quad \hat{C}_{rf}^B(x) = majority\ vote\ \{\hat{C}_b(x)\}_1^B$

***Support Vector Machine (SVM):***

SVM is a method of supervised machine learning, which explores data sets to determine results. SVM is widely used for binary classification and prediction. In addition, SVM can use the "kernel trick", which sets instances in high-dimensional space to provide non-linear prediction or classification (Alanazi et al., 2017). Given the training data ($\overrightarrow{x_\iota}$, $\overrightarrow{y_\iota}$) ,……..( $\overrightarrow{x_n}$, $\overrightarrow{y_n}$), we wish to to find maximum margin hyperplane that devides the group of points x$i$ for which $\overrightarrow{y_\iota} = 1$ from the other group with $\overrightarrow{y_\iota} = -1$. The hyperplane can be represented as (Leskovec et al. 2014 ) :

$$\boldsymbol{\vec{w}.\vec{x} - b = 0} \tag{7.5}$$

With $\vec{w}$ running parpendicular to the hyperplane. The two parallel hyperplanes at edge of the two data sets are defined as;

$$\boldsymbol{\vec{w}.\vec{x} - b = 1} \tag{7.6}$$

$$\boldsymbol{\vec{w}.\vec{x} - b = -1} \tag{7.7}$$

Whereby the objective function is to minimize $\|\vec{w}\|$ from the equation (7.6) and equation (7.7).

***Artificial Neural Network (ANN):***

ANN is a widely used machine learning method that classifies and predicts. ANN is a computational model derived by the connectivity of neurons to animate nervous systems. Any function of assignment from training inputs to training outputs can be used if nonlinear functions are used in the network. ANN requires enough neurons in the network and enough training examples (Alanazi et al. 2017). Inputs of flow signals are $x_1, \ldots, x_n$ and it is unidirectional (Parida et al., 2012), while outputs of flow signals for neurons are referred to as ($O$). The output of neuron signals is as follows:

$$\boldsymbol{O = f(net) = f(\sum_{j=1}^{n} w_j x_j)} \tag{7.8}$$

Where:

$w_j$ = The weight vectors

$f(net)$ = The activation functions

Also, the variable *net* is defined as a scalar product of the weight and input vectors by

$$\boldsymbol{net = w^T x = w_{\ 1} x_1 + \cdots + w_n \ x_n} \tag{7.9}$$

Where:

$T$ = Is the transpose of a matrix.

The output value $O$ is computed as

$$\boldsymbol{O \ = \ f(net) \ = \begin{cases} 1 & if \ w^T x \geq \ \theta \\ 0 & otherwise \end{cases}} \tag{7.10}$$

The $(\theta)$ is called the threshold level; also, this type of node is also called a linear threshold unit. The internal activity of the model of neurons is given by

$$\boldsymbol{v_k = \sum_{j=1}^{p} w_{kj} x_j} \tag{7.11}$$

Then the output of the neuron would be the outcome of some activation function on the value of $v_k$.

## 7.3 Results and Discussions

### 7.3.1 Descriptive Statistics

The results section provides the descriptive statistics of variables used in developing a machine learning tool for detecting the transportation mode. The selected features for each trip duration include the HRact, HRrest, speed (mph), and energy expenditure expressed in calories. The PAM and MET values are derived from heart rate and calorie information, respectively. Personal information, such as age, gender, weight, and height, may also be used by the machine learning algorithms. The study also calculates the body mass index (BMI) of each participant from the weight and height information.

Figure 7.1 shows the distribution of speed, energy expenditure, PAM, and deviation of actual heart rate from the resting heart rate for each transportation mode type.

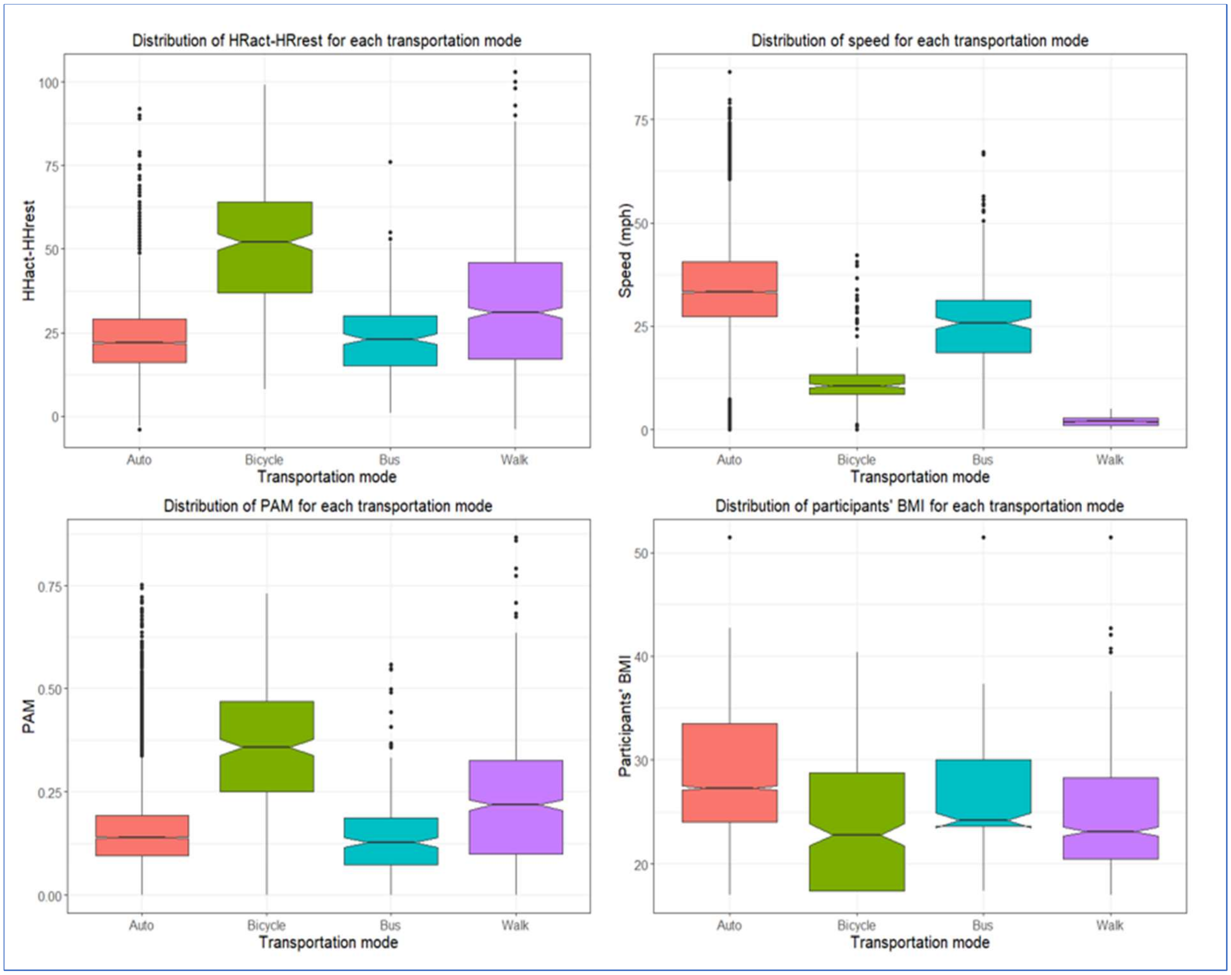



**Figure 7.1 Distribution of classifiers/features by transportation mode type**

Table (7.1) shows the ANOVA test results, which explores the significance level of any observed differences of the selected features across transportation modes. The study uses the post-hoc Tukey test for pairwise comparisons of means as shown in Table (7.2).

On average, auto speeds were 34±13.9 mph, bus speeds were 26±12.3 mph, bicycle speeds were 12±6.6 mph and walking speeds were 2±1.1 mph. The ANOVA test (F (3,8768) =1744.8, p=0.000) indicated a significant difference in the observed speeds across transportation modes. The post hoc Tukey test showed a significant difference in speed for each pair of transportation modes.

The analysis calculates the difference between the HRact and HRrest and (PAM) for each transportation mode. The bicycle mode has the highest *HRact-HRrest* (51±18.4) and PAM value

(0.36±0.15) followed by walking, which has a *HRact-HRrest* of 32±18.8 and PAM value of 0.22±0.15. The ANOVA test denotes a significant difference in the observed HRact-*HRrest* and PAM values across the transportation modes. However, the Tukey test indicates an insignificant difference in *HRact-HRrest* and PAM value between the bus and auto modes.

The BMI represented the physical characteristics of participants for each mode. The participants using a bicycle had the lowest BMI (24.7±7.1) and those walking had the second lowest (25.9±7.4). Participants who used the automobile for most of their trips had the highest BMI (29±6.4). The ANOVA test and post-hoc Tukey test showed a significant difference in participants' BMI across all transportation modes. A pairwise comparison of BMI means for the transportation modes appeared significant for all modes except walk and bus.

**Table 7.1 ANOVA Test for the Selected Features by Transportation Mode**

| Feature | Source | SS | df | MS | F | Prob > F |
|---|---|---|---|---|---|---|
| Speed (mph) | Between groups | 876981.326 | 3 | 292327.109 | 1744.48 | 0.000 |
| | Within groups | 1469273.82 | 8768 | 167.572288 | | |
| HRact-HRrest | Between groups | 272799.901 | 3 | 90933.3004 | 659.81 | 0.000 |
| | Within groups | 1208383.2 | 8768 | 137.817427 | | |
| PAM | Between groups | 14.4698168 | 3 | 4.82327227 | 445.23 | 0.000 |
| | Within groups | 94.985774 | 8768 | 0.01083323 | | |
| BMI | Between groups | 12478.1398 | 3 | 4159.37992 | 95.7 | 0.000 |
| | Within groups | 381086.261 | 8768 | 43.4633053 | | |

**Table 7.2 A Tukey Post Hoc Test**

| Feature | Transportation Mode | Contrast | Std. Err. | t | P>t |
|---|---|---|---|---|---|
| Speed (mph) | Bicycle vs Auto | -22.080 | 0.760 | -29.060 | 0.000 |
| | Bus vs Auto | -8.209 | 0.896 | -9.170 | 0.000 |
| | Walk vs Auto | -31.958 | 0.472 | -67.780 | 0.000 |
| | Bus vs Bicycle | 13.871 | 1.155 | 12.010 | 0.000 |
| | Walk vs Bicycle | -9.878 | 0.869 | -11.370 | 0.000 |
| | Walk vs Bus | -23.749 | 0.990 | -24.000 | 0.000 |
| HRact-HRrest | Bicycle vs Auto | 27.873 | 0.689 | 40.450 | 0.000 |
| | Bus vs Auto | **0.667** | **0.812** | **0.820** | **0.844** |
| | Walk vs Auto | 8.976 | 0.428 | 20.990 | 0.000 |
| | Bus vs Bicycle | -27.206 | 1.048 | -25.970 | 0.000 |
| | Walk vs Bicycle | -18.898 | 0.788 | -23.990 | 0.000 |
| | Walk vs Bus | 8.308 | 0.897 | 9.260 | 0.000 |
| PAM | Bicycle vs Auto | 0.202 | 0.006 | 33.110 | 0.000 |
| | Bus vs Auto | **-0.012** | **0.007** | **-1.660** | **0.346** |
| | Walk vs Auto | 0.065 | 0.004 | 17.130 | 0.000 |
| | Bus vs Bicycle | -0.214 | 0.009 | -23.070 | 0.000 |
| | Walk vs Bicycle | -0.137 | 0.007 | -19.670 | 0.000 |
| | Walk vs Bus | 0.077 | 0.008 | 9.660 | 0.000 |
| BMI | Bicycle vs Auto | -4.279 | 0.387 | -11.060 | 0.000 |
| | Bus vs Auto | -2.547 | 0.456 | -5.580 | 0.000 |
| | Walk vs Auto | -3.040 | 0.240 | -12.660 | 0.000 |
| | Bus vs Bicycle | 1.732 | 0.588 | 2.940 | 0.017 |
| | Walk vs Bicycle | 1.239 | 0.442 | 2.800 | 0.026 |
| | Walk vs Bus | **-0.493** | **0.504** | **-0.980** | **0.762** |

### 7.3.2 Machine learning-based transportation mode detection tool

The study uses features generated from smartwatch and smartphones devices to develop a machine learning-based transportation mode detection tool. The procedure involves two steps: (1) Model calibration, and (2) Model validation. The research team uses accuracy, which represents the percentage of correctly classified observations, to test the performance of each model. It provides by the model. The present study compares four machine learning models (Extreme Gradient Boosting (xgboost), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN)). The dataset uses a ratio of 80% to 20% for model calibration and model validation, respectively.

### *7.3.2.1 Model Calibration*

The research team determines the optimum tuning parameters for each machine learning model using a cross-validation approach. Figure (7.2) provides the optimal tuning parameters for Extreme Gradient Boosting, RF, SVMS, and ANN. The tuning parameters that yielded the highest accuracy are used to calibrate the final model. For Gradient boosting, the final values used for the model are nrounds =50, max_depth =2, eta =0.3, gamma =0, colsample_bytree = 0.8, min_child_weight = 1 and subsample = 0.5. The final ANN values are hidden units = 7 and decay = 0.1. A cost value C, of 0.5 is optimal for SVM. The number of features (mtry=2) yields the optimal results for the RF model.

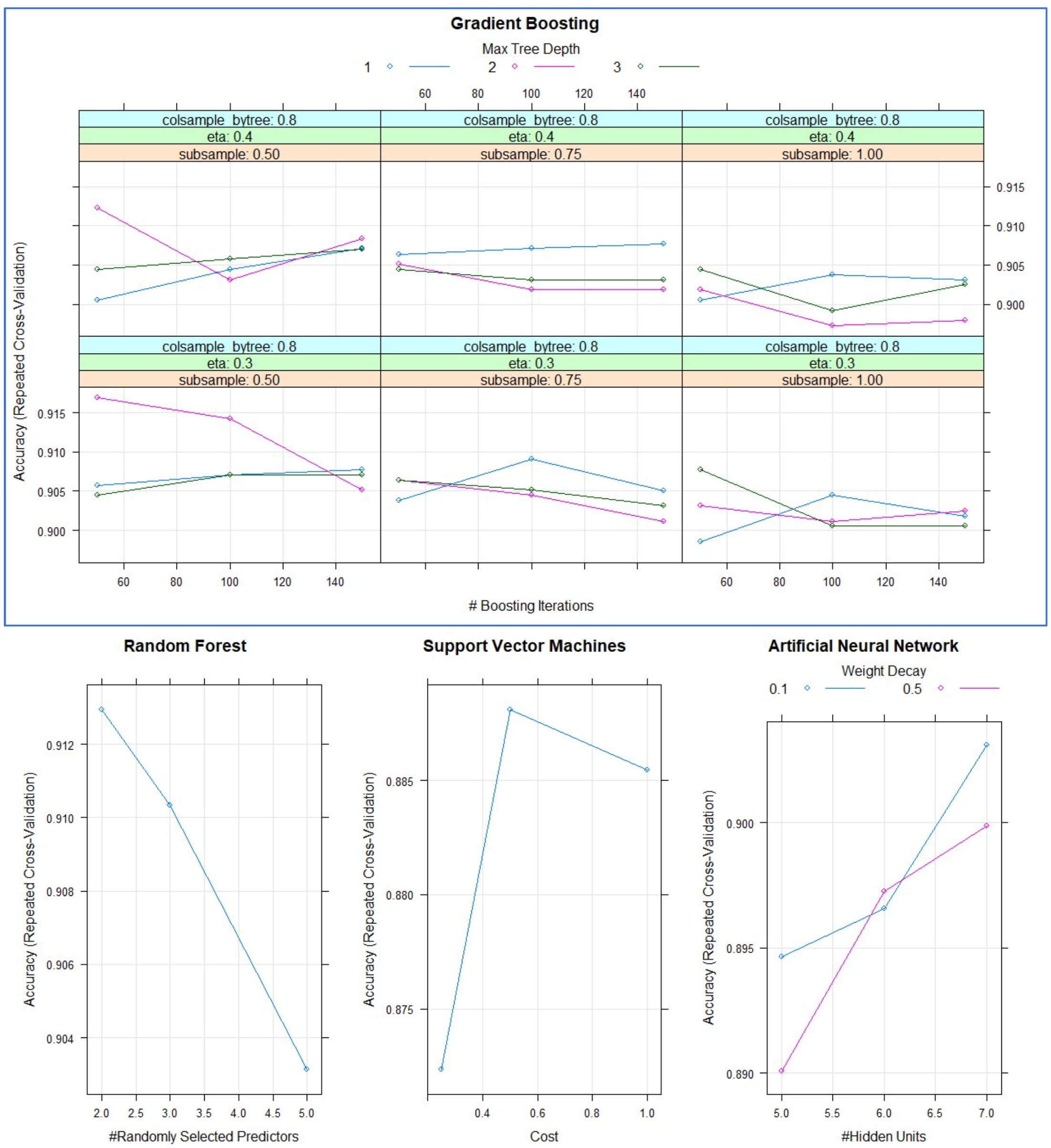



**Figure 7.2 Model tuning parameters for XGBoost, RF, SVMS, and ANN**

#### *7.3.2.2 Model Validation*

Random forest outperforms the other competing models in detecting each transportation model using features generated from smartwatches and smartphones [Figure (7.3)]. The RF model yields the highest accuracy in detecting non-motorized modes, namely, walking mode (97.2%) and bicycle mode (90.6%). The lowest accuracy occurs when classifying the bus mode (61.4%). The previous statistical testing indicates an insignificant difference in the heart rate information expressed as *HRact-HRrest* and PAM between the bus and auto mode. As a result, the bus mode has the lowest accuracy and most of the misclassification cases involve instances with misclassifications between the auto and bus modes.

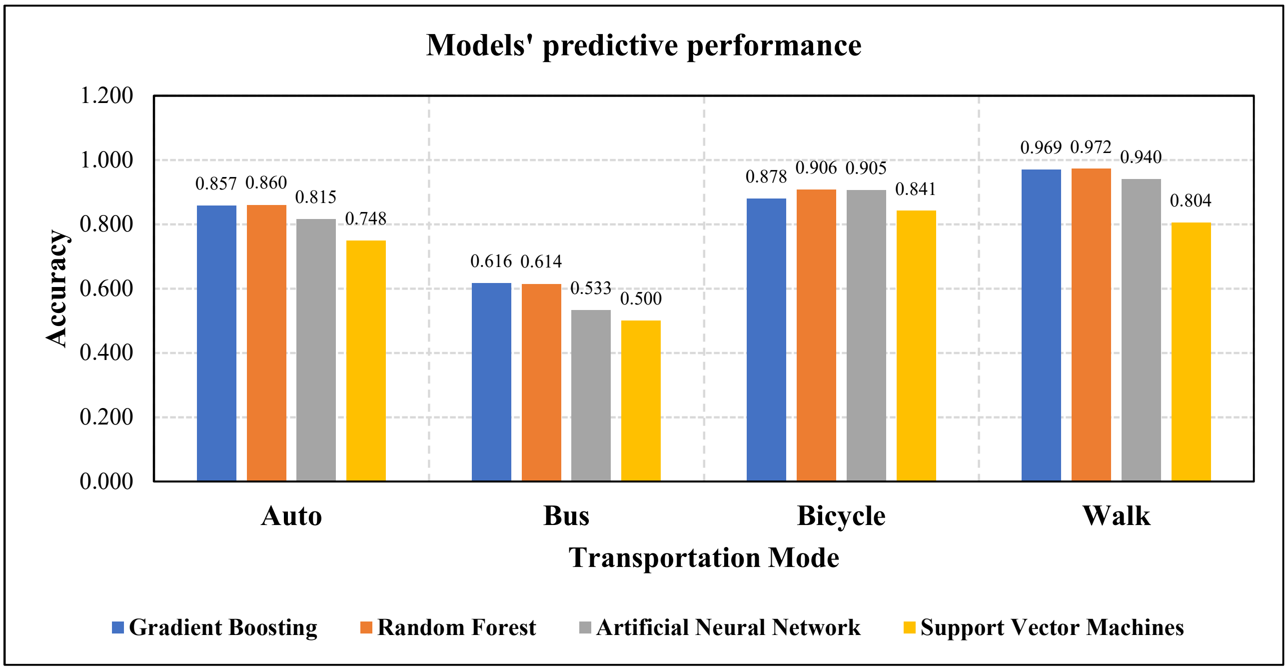



**Figure 7.3 Predictive models performances in detecting the transportation mode**

## 7.4 Conclusion

The study uses data collected from smartphones and smartwatches to predict transportation modes. The study does not compare the performance of predictive models with previous studies because this study uses PA as a new feature in predicting transportation modes. The study seeks to address some of the most important limitations of previous predictive processes, such as the problem of battery consumption and the continuity of predictive processes for as long as possible. The study introduces the PASTA platform, which works as a mobile application linked to servers that conduct a process of compiling, classifying and analyzing data from both the phone and the watch. The PASTA platform shows high potential to monitor people's activities and classify these activities into trips and non-trips. The PASTA platform also provides daily reports of the spatial and temporal data associated with these trips and the accompanying PA. The researchers investigate several Machine learning methods to predict the transportation modes. The Random Forest method appears to be the most accurate in terms of detecting different modes of transport.

# Chapter 8: Exploring the Association Between Individual's Physical Activity Level with Socio-economic and Body-composition characteristics

## 8.1 Introduction

The study tests the statistical association between an individual's physical activity level, socioeconomic characteristics, and body composition profiles. This study incorporates the actual measurement of physical activity intensity and body composition through the PASTA platform for data collection. The research team measures the more accurate parameters of body composition characteristics, such as Body Mass Index (BMI), Body Fat Mass (BFM), and Percent Body Fat (PBF), using InBody Test equipment.

## 8.2 Research Methodology

This analysis uses data from the PASTA platform, data from the questionnaire (pre-survey/main survey), and data from the body composition test of the study participants.

Figure (8.1) shows the framework for the study of the transportation modes associated with physical activities. This diagram shows that the final outputs of the study required a series of steps.

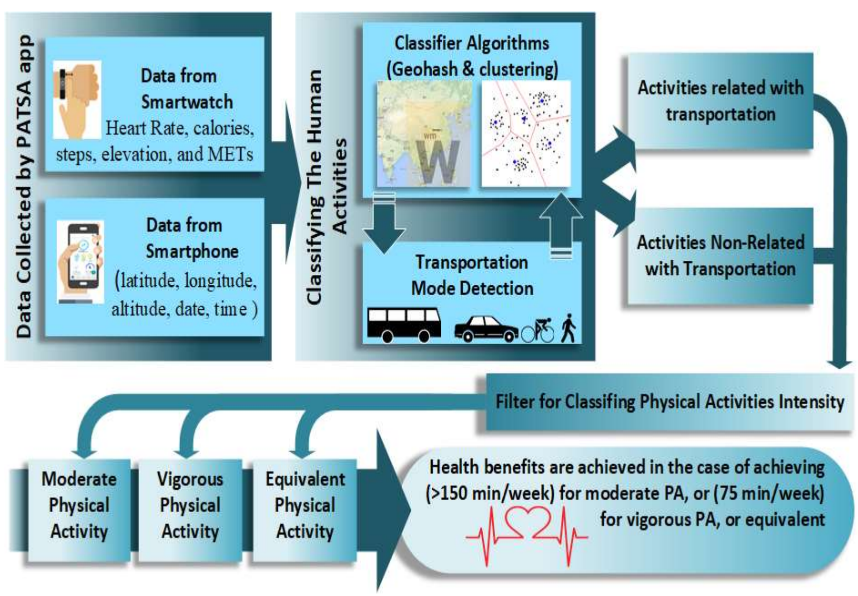



**Figure 8.1 A framework for the study of the transportation modes association with physical activities**

The research team creates six physical activity levels using heart rate or Metabolic Equivalent of Task (MET) values [Table (8.1)], adopted from Sirven and Varrato (1999).

**Table 8.1 Classification of physical activity intensity, based on physical activity lasting up to 60 minutes**

| Intensity | Endurance-type activity | | | | | | | | Strength-type exercise |
|---|---|---|---|---|---|---|---|---|---|
| | Relative intensity | | | Absolute intensity (METs) | | | | | Relative intensity |
| | $VO_2$ max (%) heart rate reserve (%) | Maximal heart rate (%) | PRE | Young (20-39) | Middle-aged (40-39) | Old (65-79) | Very old (80+) | PRE | Maximal voluntary contraction (%) |
| **Very Light** | <25 | <30 | < 9 | <3.0 | <2.5 | <2.0 | ≤1.25 | <10 | <30 |
| **Light** | 25-44 | 30-49 | 9-10 | 3.0-4.7 | 2.5-4.4 | 2.0-3.5 | 1.26-2.2 | 10-11 | 30-49 |
| **Moderate** | 45-59 | 50-69 | 11-12 | 4.8-7.1 | 4.5-4.4 | 3.6-4.7 | 2.3-2.95 | 12-13 | 50-69 |
| **Hard** | 60-84 | 70-89 | 13-16 | 7.2-10.1 | 6.0-8.4 | 4.8-6.7 | 3.0-4.25 | 14-16 | 70-84 |
| **Very Hard** | ≥85 | ≥ 90 | ≥16 | ≥10.2 | ≥8.5 | ≥6.8 | ≥4.25 | 17-19 | >85 |
| **Maximal** | 100 | 100 | 20 | 12 | 10 | 8 | 5 | 20 | 100 |

**Maximal values are mean values achieved during maximal exercise by healthy adults. Absolute intensity (METs) values are approximate mean values for men. Mean values for women are approximately 1–2 METs lower than those for men.**

### 8.2.1 Data Collection

The data collection period lasted for nearly six months from March 2019 to August 2019. Most of the research participants were students, staff, and faculty members at Western Michigan University (60 participants) and the University of Texas at Arlington (60 participants). The study collected the minutes of PA for a week and determined the PA level of intensity to determine the impacts of daily activities and transportation on health. The research team excluded PA that fell within the light and very light levels.

The researchers reorganize several of the data fields; the vigorous physical activities include values from the hard, very hard, and maximal categories. The research team classify moderate or more intense physical activities into trips or non-trips. Table (8.2) shows a sample of the results provided by the PASTA platform through a computerized interface, which provides data for any week over the study duration. The equivalent minutes of vigorous PA is computed by dividing the total minutes of moderate PA by two and adding the minutes to vigorous PA minutes as shown in equation (8.1), (8.2).

$$\textbf{Equivalent to moderate PA minutes} = \textbf{moderate} + \textbf{vigorous} \times \textbf{2} \qquad \textbf{(8.1)}$$

Or

$$\textbf{Equivalent to vigorous PA minutes} = \textbf{vigorous} + \frac{\textbf{\textit{moderate}}}{\textbf{2}} \qquad \textbf{(8.2)}$$

The red and green colors in Table (8.2) indicate the level of PA intensity. The green color indicates that a person weekly PA minutes met the required standard value to achieve health benefits for a given category while the red color indicate otherwise.

**Table 8.2 Sample of the weekly report for PASTA application about the PA intensity achieved for each participant (from 3/25/2019 to 3/31/2019)**

| User ID | Trip Moderate | Trip Vigorous | Trip Equivalent | Non-Trip Moderate | Non-Trip Vigorous | Non-Trip Equivalent | Total Equivalent (Trip + Non-Trip) |
|---|---|---|---|---|---|---|---|
| 76 | 0 | 1 | 1 | 37 | 122 | 140.5 | 141.5 |
| 77 | 2 | 2 | 3 | 94 | 86 | 133 | 136 |
| 78 | 8 | 6 | 10 | 111 | 95 | 150.5 | 160.5 |
| 79 | 0 | 6 | 6 | 12 | 66 | 72 | 78 |
| 80 | 0 | 1 | 1 | 43 | 129 | 150.5 | 151.5 |
| 81 | 0 | 2 | 2 | 7 | 41 | 44.5 | 46.5 |
| 82 | 0 | 2 | 2 | 6 | 64 | 67 | 69 |
| 83 | 0 | 0 | 0 | 1 | 16 | 16.5 | 16.5 |
| 84 | 82 | 35 | 76 | 401 | 170 | 370.5 | 446.5 |
| 85 | 126 | 92 | 155 | 35 | 64 | 81.5 | 236.5 |
| 86 | 0 | 2 | 2 | 33 | 104 | 120.5 | 122.5 |
| 87 | 3 | 14 | 15.5 | 9 | 45 | 49.5 | 65 |
| 88 | 0 | 4 | 4 | 20 | 53 | 63 | 67 |
| 89 | 3 | 34 | 35.5 | 30 | 166 | 181 | 216.5 |
| 90 | 0 | 2 | 2 | 1 | 17 | 17.5 | 19.5 |
| 91 | 0 | 17 | 17 | 1 | 66 | 66.5 | 83.5 |
| 92 | 0 | 3 | 3 | 4 | 60 | 62 | 65 |
| 93 | 0 | 7 | 7 | 2 | 49 | 50 | 57 |
| 94 | 0 | 9 | 9 | 18 | 74 | 83 | 92 |
| 95 | 103 | 47 | 98.5 | 401 | 215 | 415.5 | 514 |
| 96 | 3 | 9 | 10.5 | 180 | 415 | 505 | 515.5 |

Overall, a person can look into the total equivalent PA minutes (sum of trip and non-trip equivalent) to see if he/she has achieved the health benefits. A person may have less than required trip or non-trip PA minutes but have total equivalent PA minutes which meets the required PA minutes that provide health benefits [see UserID 79 and 91 in Table (8.2)]. A person can look into the category where he/she can increase his/her PA weekly minutes if the total equivalent PA minutes is less than the required minimum PA minutes to achieve the health benefits.

The other two sources of data of the study, the survey, as well as body composition tests, are intended to form a vision of the demographic and social characteristics of the participants. The results obtained from the questionnaire and body composition test specified the number and types of variables, determining the shape of the relationships between the variables with the physical activities of people.

The questionnaire (pre-survey) which it is paper-based, involved twelve questions, including general questions about gender, age, race/ethnicities, etc., with other questions related to physical activities. The pre-survey was completed by 120 participants distributed among 60 participants in Kalamazoo-Michigan, and 60 other participants in Arlington-Texas. The summary of the participants in the pre-survey is shown in table (8.3). Each participant signed the consent form document to participate in the study before participating in the study. Therefore, each participant that received a Fitbit smartwatch (charge 2 or 3) was asked to download Fitbit and PASTA application on their phone.

**Table 8.3 Characteristics' Summary of participants in pre-survey**

| Variables | Details | Michigan | | Texas | | Grand Total |
|---|---|---|---|---|---|---|
| | | Female | Male | Female | Male | |
| **Race and ethnicity** | Asian | | 11 | 3 | 15 | 29 |
| | Black | 1 | 4 | 3 | | 8 |
| | Hispanic | 4 | 2 | 2 | 4 | 12 |
| | White | 9 | 20 | 15 | 5 | 49 |
| **Education level** | Some high school | 1 | 2 | | | 3 |
| | High school graduate | 1 | 1 | | 1 | 3 |
| | Some college credits | 3 | 8 | 7 | 6 | 24 |
| | | | | | | 68 |
| **Professional status** | Administration position | 2 | 1 | 4 | 1 | 8 |
| | | | | | | 3 |
| | Office worker | 2 | 1 | 4 | 1 | 8 |
| | | | | | | 2 |
| | Student | 8 | 32 | 12 | 18 | 70 |
| | | | | | | 8 |
| **Age group** | <18 | 1 | | | | 1 |
| | 18 - 25 | 4 | 16 | 11 | | 45 |
| | 26 - 49 | 8 | 18 | 8 | 10 | 44 |
| | | | | | | 9 |
| **Annual income** | <30000 | 8 | 28 | 10 | 20 | 66 |
| | | | | | | 19 |
| | 50,000 - 100,000 | 3 | 3 | 4 | 3 | 13 |
| | 100000-150000 | | 1 | | | 1 |
| **Health condition** | Excellent | 2 | 6 | 1 | 3 | 12 |
| | | | | | | 58 |
| | Fair | 4 | 8 | 8 | 4 | 24 |
| | | | | | | 5 |
| **driving license** | No | 2 | 5 | 3 | 6 | 16 |
| | | | | | | 83 |
| **Number of vehicles** | 0 | | 6 | 2 | 7 | 15 |
| | | | | | | 43 |
| | 2 | 5 | 4 | 11 | 6 | 26 |
| | 3 + | 2 | 5 | 5 | 2 | 14 |

Also, the study uses a device named "InBody 570," which provides complete printed results for each body composition test. Figure (8.2) shows the sample results of the InBody test. The body composition information includes the Body Mass Index (BMI), Body Fat Mass (BFM), Percent Body Fat (PBF), and Skeleton Muscle Mass (SMM) as shown in Table (8.4). The researchers also collect the waist and hip circumference manually, using a tape measure, to find the waist to hip ratio, which represents a good indicator of the level of obesity in people.

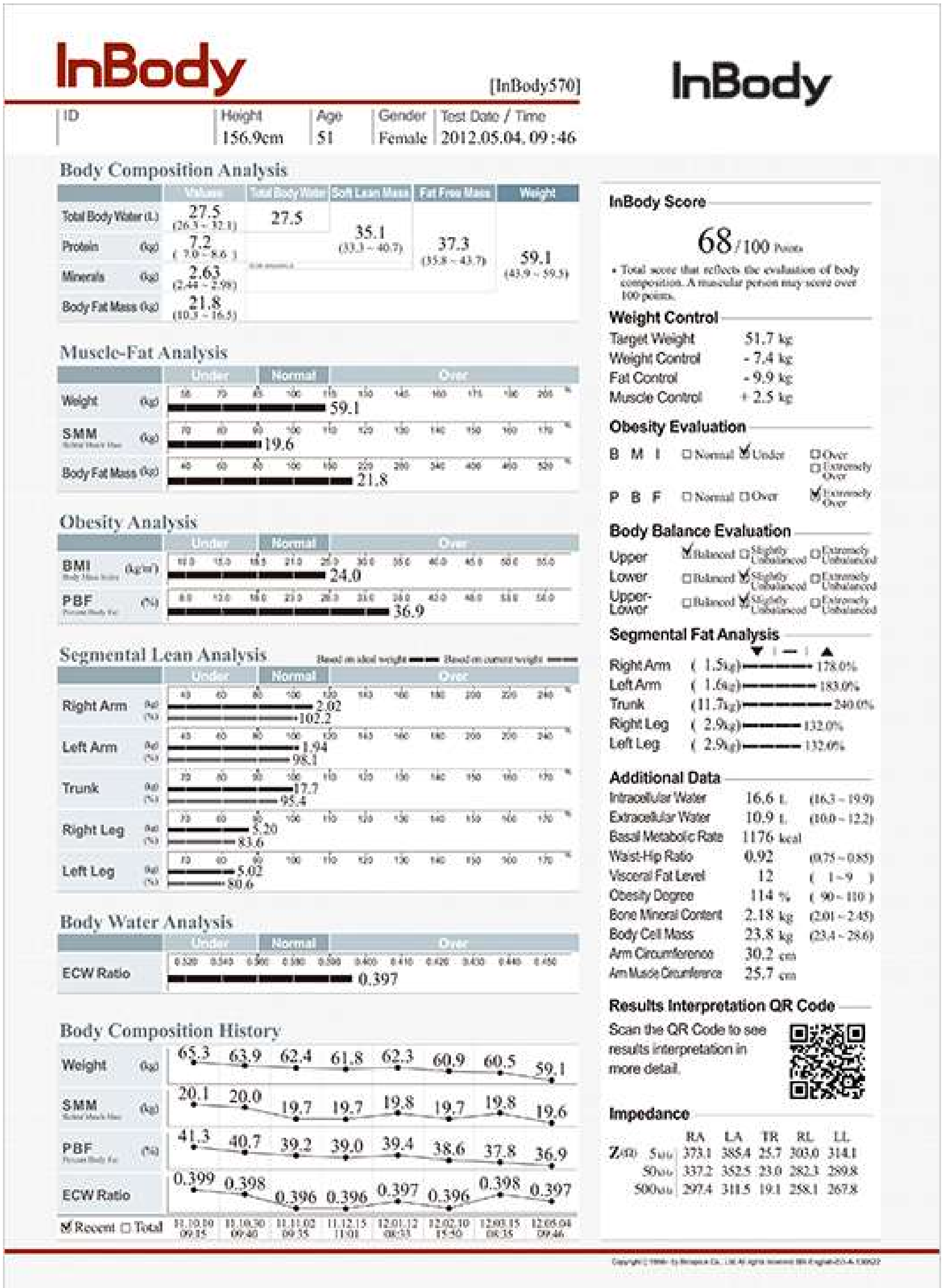



**Figure 8.2 The InBody test report**

**Table 8.4 Characteristics' Summary of participants in the InBody test**

| State | Variable | Female Average | Male Average | Female Max | Male Max | Female Min | Male Min | Female SD | Male SD |
|---|---|---|---|---|---|---|---|---|---|
| Michigan | Age | 38.08 | 29.55 | 52 | 57 | 17 | 20 | 10.91 | 9.74 |
| | Weight | 169.52 | 195.83 | 285 | 300 | 108 | 115 | 40.98 | 42.47 |
| | Height | 64.23 | 68.51 | 69 | 77 | 56 | 61 | 2.77 | 2.97 |
| | BMI | 29.36 | 29.19 | 42.90 | 40.70 | 18.00 | 17.70 | 7.02 | 6.37 |
| | PBF | 38.19 | 31.60 | 53.80 | 244.00 | 23.70 | 6.80 | 10.08 | 29.20 |
| | LBM | 102.79 | 144.98 | 140.00 | 222.00 | 67.70 | 103.40 | 13.07 | 22.16 |
| | Waist | 90.20 | 97.51 | 126.00 | 130.00 | 49.00 | 68.00 | 16.14 | 16.03 |
| | Hip | 109.48 | 108.78 | 140.00 | 134.00 | 90.00 | 85.00 | 13.51 | 11.57 |
| | waist-hip ratio | 0.82 | 0.89 | 1.04 | 1.05 | 0.52 | 0.78 | 0.07 | 0.07 |
| Texas | Age | 34.17 | 27.63 | 60 | 52 | 18 | 19 | 12.79 | 8.72 |
| | Weight | 168.28 | 177.75 | 237 | 317 | 117 | 124 | 39.15 | 33.35 |
| | Height | 65.08 | 67.63 | 71 | 73 | 61 | 59 | 2.87 | 2.36 |
| | BMI | 28.11 | 27.32 | 40.78 | 42.01 | 17.52 | 19.53 | 6.56 | 4.97 |
| | PBF | 33.24 | 21.07 | 45.10 | 37.00 | 9.80 | 10.30 | 7.97 | 6.51 |
| | LBM | 110.04 | 138.58 | 140.22 | 199.73 | 0.00 | 0.00 | 17.23 | 18.54 |
| | Waist | 85.33 | 88.97 | 120.00 | 124.00 | 65.00 | 68.00 | 17.61 | 11.53 |
| | Hip | 109.01 | 101.36 | 169.00 | 128.00 | 90.00 | 88.00 | 14.97 | 7.79 |
| | waist-hip ratio | 0.78 | 0.88 | 0.96 | 0.97 | 0.44 | 0.75 | 0.10 | 0.06 |

Figure (8.3) illustrates the relationships proposed by the study for variables associated with physical activities; these relationships include several variables from the survey, others extracted from the PASTA platform, and the body composition test data.

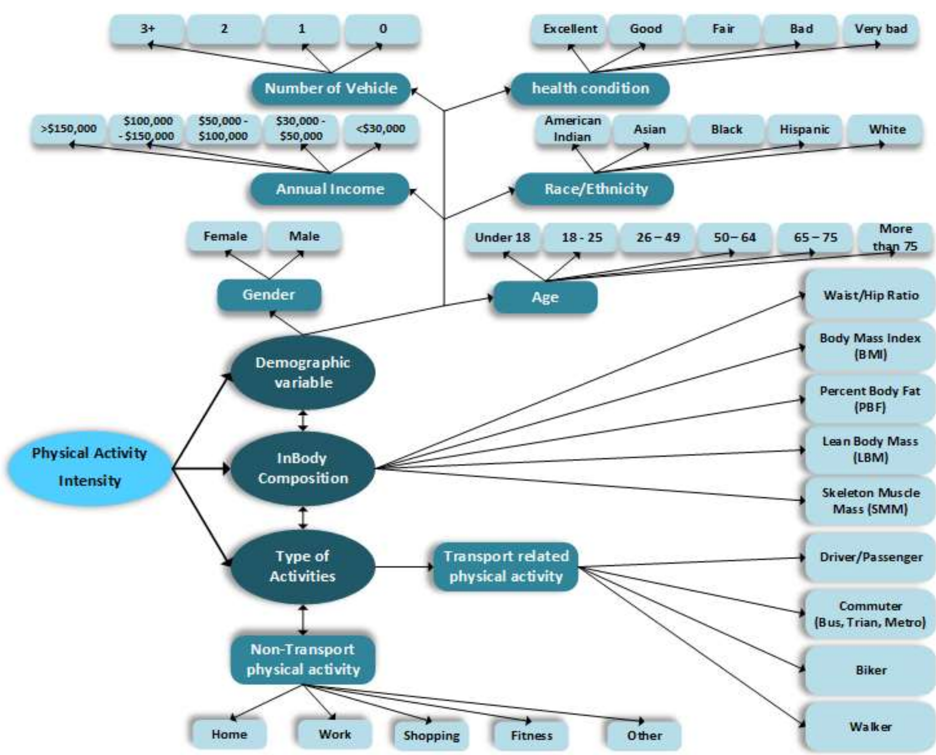



**Figure 8.3 The relationship of transportation with PA and the most important variables that may determine physical activity intensity**

## 8.3 Results and Discussion

This study links the data (PASTA application data, questionnaire data, and body composition test data) together in both Stata and Excel for data analysis.

### 8.3.1 Descriptive Analysis

Figure (8.4) shows the amount of physical activity (average PAM per minute) by gender for each type of transport mode. For a given transportation mode, females exert higher physical activity intensity than males except for the auto and walk modes. The largest difference in physical activity intensity between female and male participants occurs with the bicycle mode.

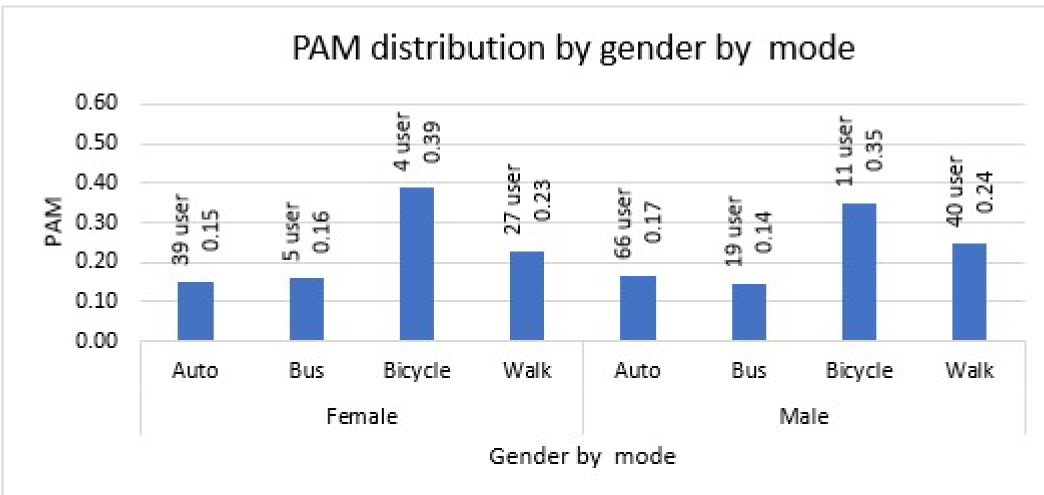
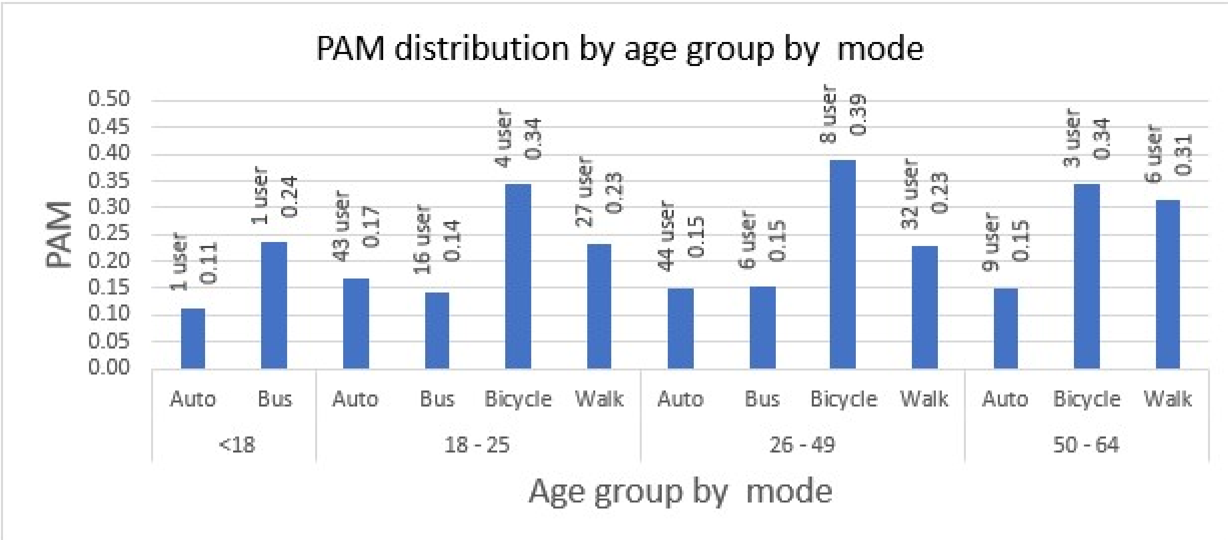



**Figure 8.4 PAM distribution by gender by mode**

Figure (8.5) shows the amount of physical activity (PAM per minute) by age group for each transportation mode. The study does not capture any bus users for the over 50 age group. Those 50-64 appear to achieve greater PA from the walk mode while those 26-49 exert the greatest overall PA for the bike mode.




**Figfure 8.5 PAM distribution by age group by mode**

Figure (8.6) shows the amount of physical activity (PAM per minute) by race/ethnic group for each transportation mode. Whites achieve high levels of physical activity when using bicycles and walking, but they achieve much less when using the bus. None of the Black study participants use bicycles. The Hispanics participants achieve their highest PAM when bicycling. The Asian subjects seem to be within normal measurements of the diversity of transportation modes and good limits for physical activity.

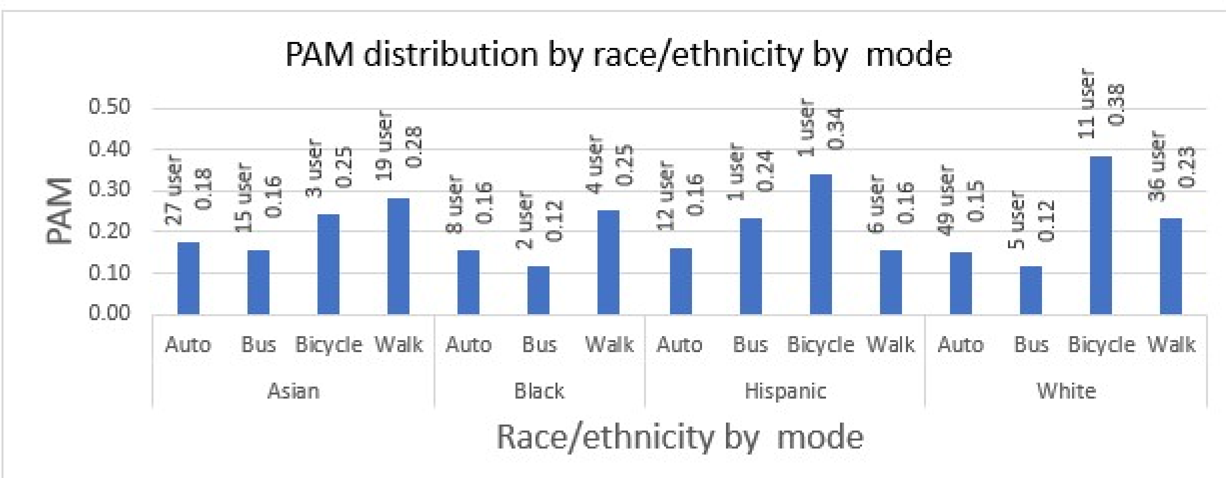



**Figure 8.6 PAM distribution by race/ethnicity by mode**

Figure (8.7) shows the amount of physical activity (PAM per minute) by the level of education with each mode of transportation. The result shows that university students and graduates from different levels of the university achieve higher physical activities levels.

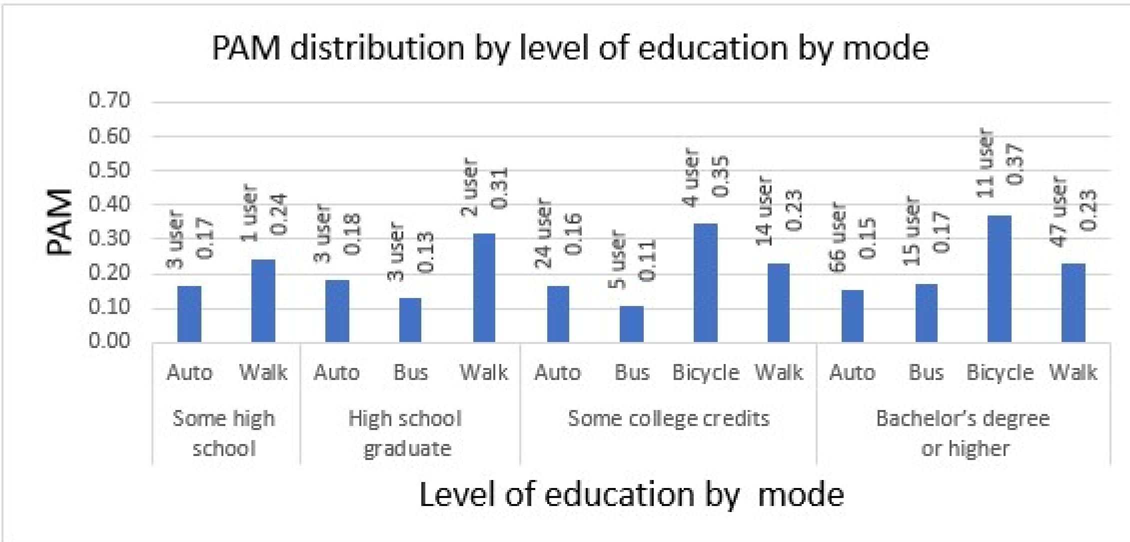



**Figure 8.7 PAM distribution by level of education by mode**

Figure (8.8) shows the amount of physical activity (PAM per minute) according to professional status with each mode of transportation. The result shows that outdoor jobs are the most frequent user of auto with low physical activities. None of the study subjects who are professionals in administrative positions or office workers use buses. Students and university faculty achieve good levels of diversity in transportation modes and physical activities.

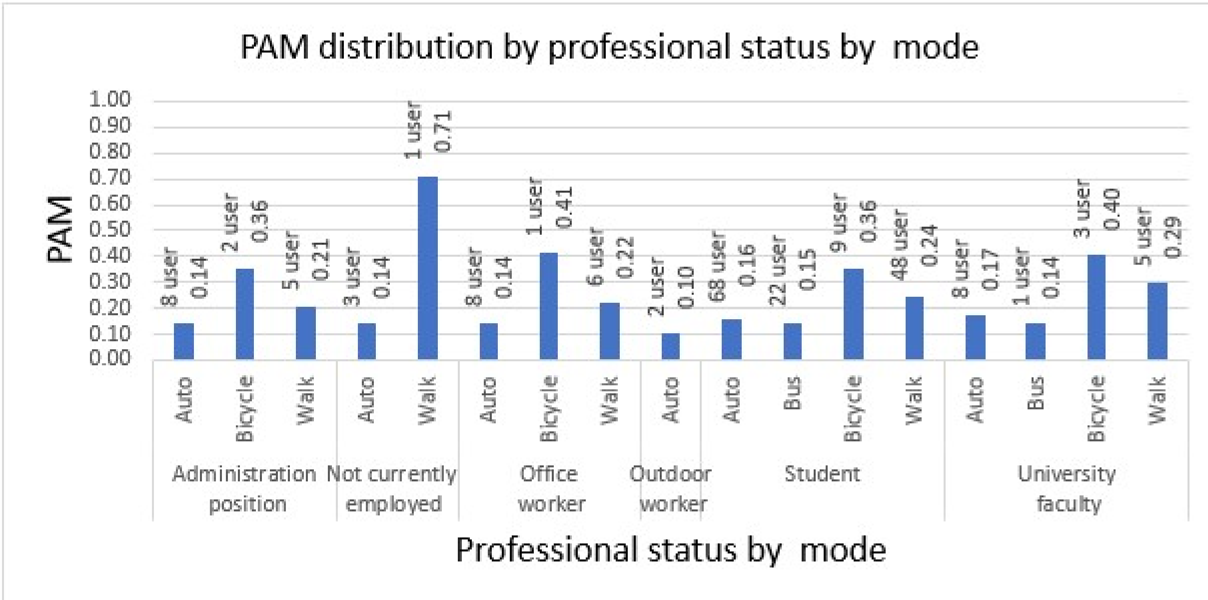



**Figure 8.8 PAM distribution by professional status by  mode**

Figure (8.9) shows the amount of physical activity (PAM per minute) according to the self-perceived health condition of the study participants for each transportation mode. The figure shows that most mode by mode outcomes appear relatively similar; however, those with bad health appear to exert a higher PAM when walking.

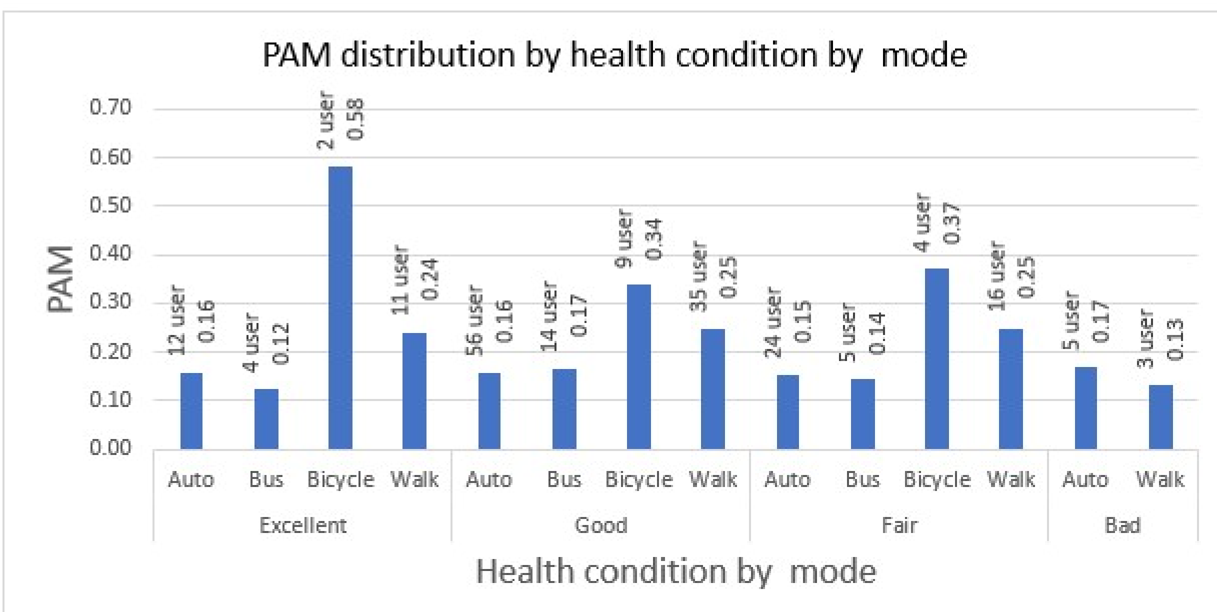



**Figure 8.9 PAM distribution by health condition by  mode**

In Figure (8.10), it was evident that vehicle ownership appears to have little impact on the use of bicycles in American society when one or two vehicles are owned. However, this situation is much less when they have three or more vehicles. This is also reflected in the desire to achieve physical activities. As for walking, it seems clear that people who do not own any vehicle were mostly using walking mode.

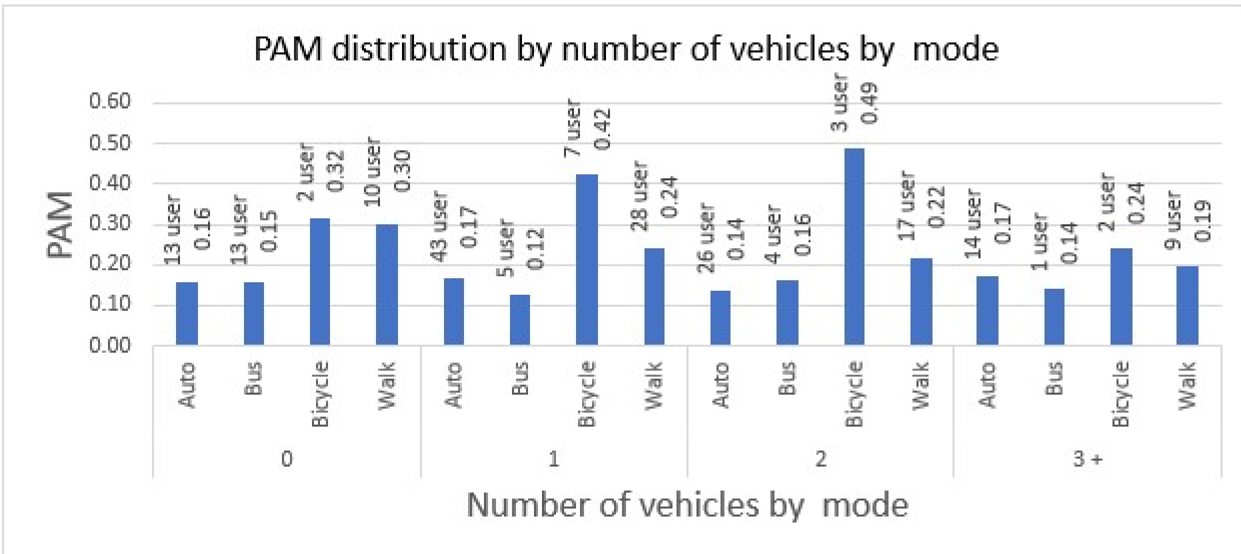



**Figure 8.10 PAM distribution by number of vehicles by mode**

In Figure (8.11), the use of the bus is reduced, with income exceeding $ 50,000 per year significantly. Distribution is also normal in the variety of transport patterns in groups with annual incomes below $ 50,000 per year.

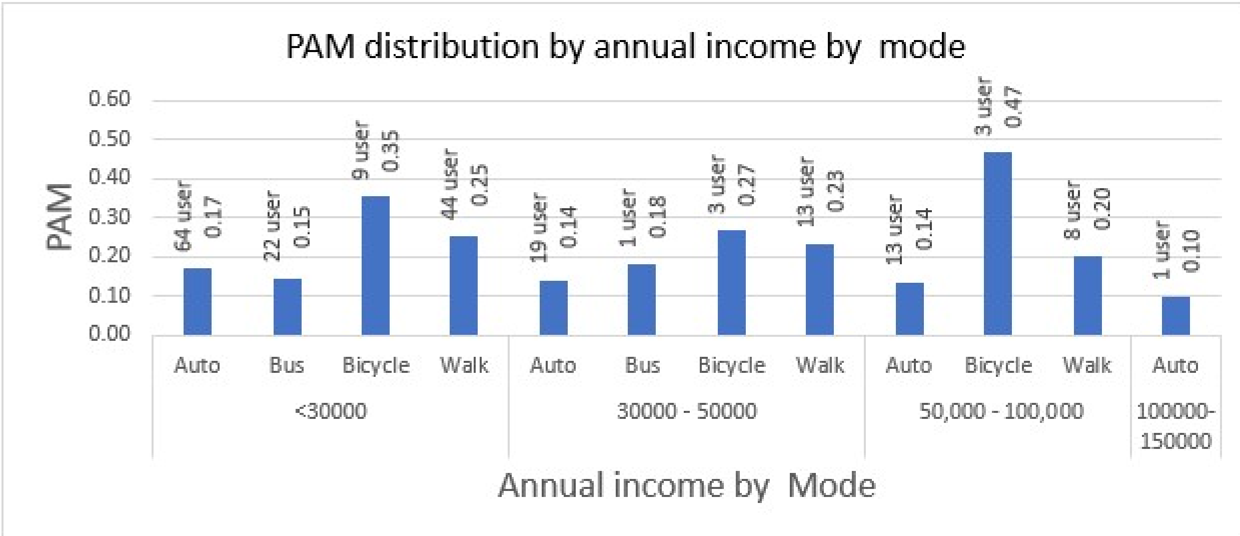
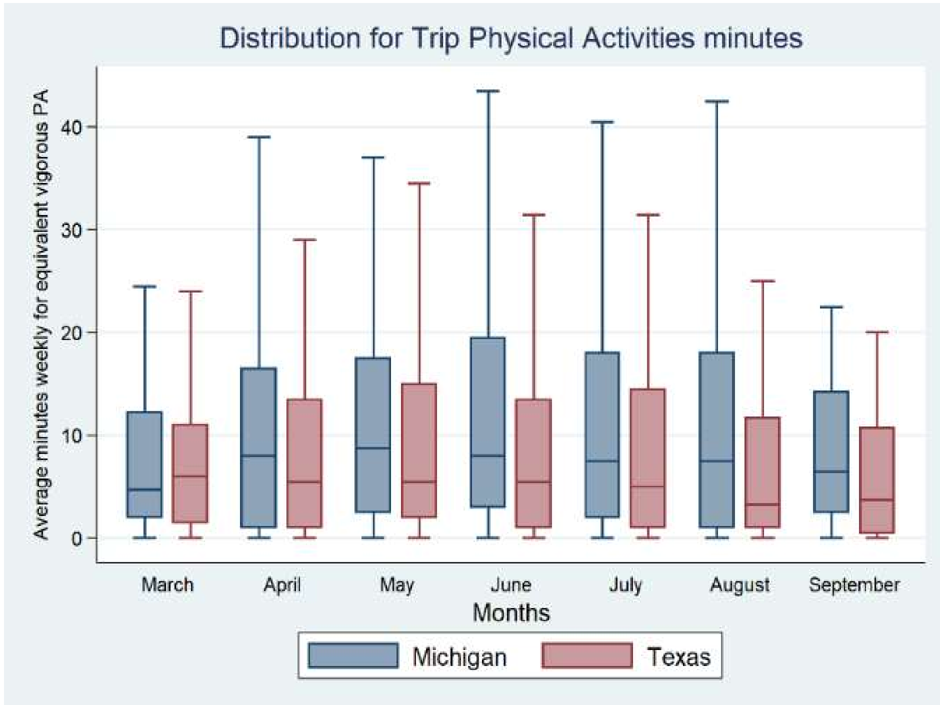
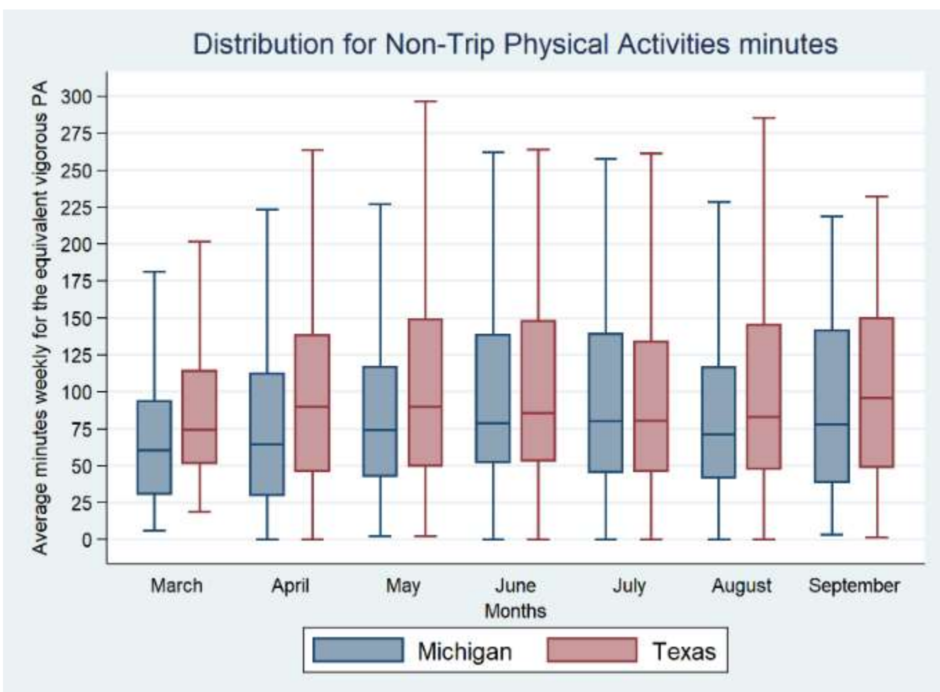



**Figure 8.11 PAM distribution by annual income by mode**

Figure (8.12-A) shows the distribution of the average weekly minutes of equivalent vigorous physical activity for trips for each month of study duration, and for the states of Michigan and Texas. It is obvious from the figure that the physical activities achieved in the state of Michigan were more than those achieved in Texas in all months except March. Typically, Michigan experiences cold weather between November to March. The cold weather may reduce people's desire to use active transportation modes. Figure (8.12-B) relates to the distribution of the average weekly minutes of equivalent vigorous physical activity for non-trips and for each month of the study period. This figure shows the convergence in the distribution of the values of physical activities in both states in the form that may increase one from the other in one month and then decrease in the following month.




A




**B**

**Figure 8.12 Distribution of average weekly minutes to the equivalent of a vigorous PA in trip and non-trip activities for each month of study.**

Figure (8.13) shows the distribution of physical activity for each transportation mode of in each state (Michigan and TexasThe physical activities from the non-motorized transportation clearly exceed those achieved by motorized transportation, as expected. This may be attributed to the fact that the overall atmosphere in Texas tends to be moderate, which helps to achieve comfortable physical activities for its users. The weather in Michigan is colder in the months of the study, which may lead road users to reduce their dependence on walking or cycling.

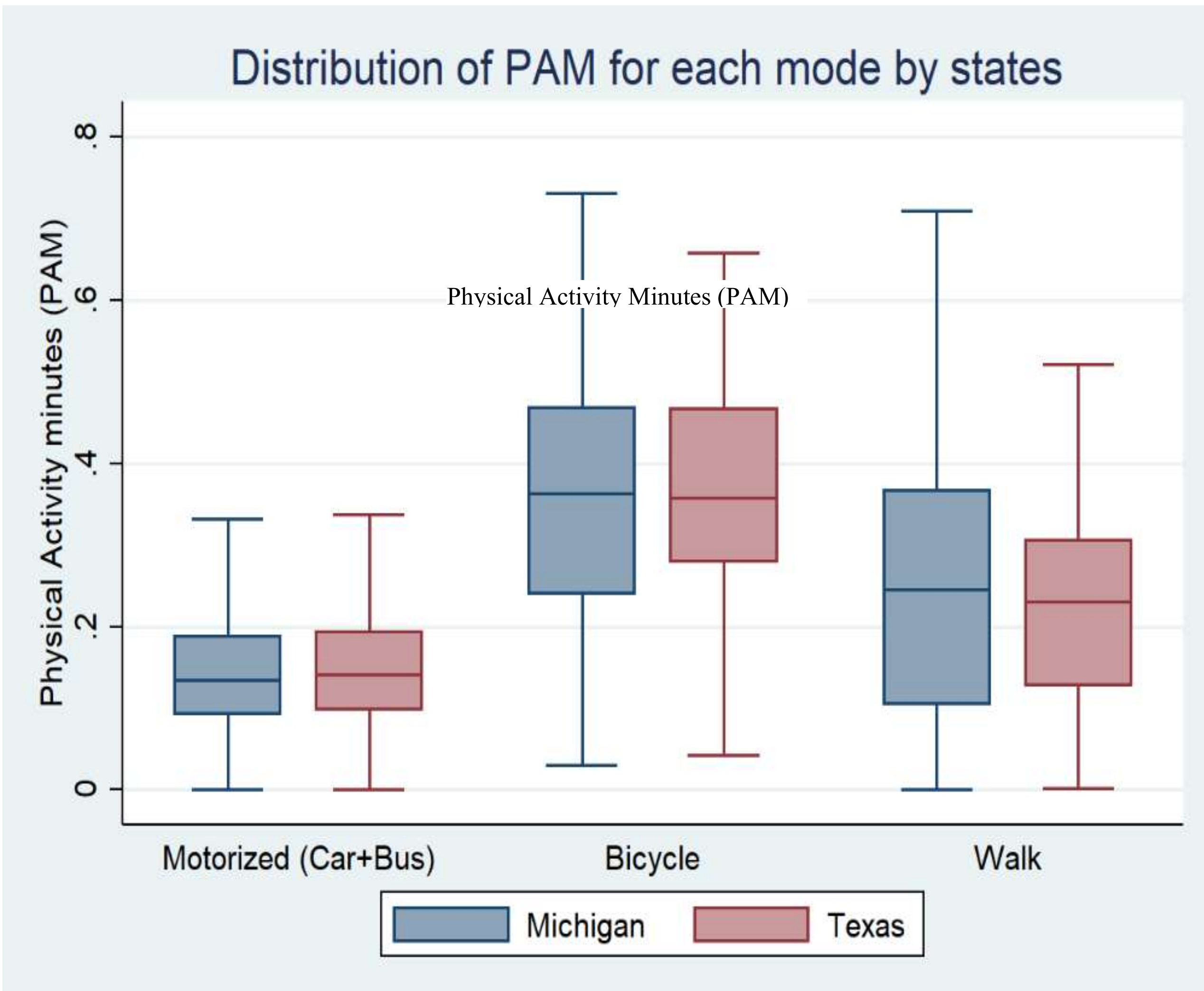



**Figure 8.13 Distribution of average weekly minutes to the equivalent of a vigorous PA in Michigan and Texas.**

### 8.3.2 Path Analysis

Path analysis is a special case of structural equation modeling (SEM) whereby it shows how the set of specified causal and non-causal relationships attributes to the observed relationships among variables. Unlike the simple regression model, path analysis can be used to find the mediation effect among variables, thus decomposing the total effect of given exogenous variables to direct and indirect effect. Path analysis was selected in this study because the predictor variables were assumed to be correlated to one another. An additional advantage of path analysis over a normal regression equations is that it allows for a researcher to run a simultaneous regression equation at once, thus allowing for the estimation of overall model goodness of fit.The path analysis also estimates the proportion of variance that was not accounted for in the model (Hox et al., 2003; Jihye et al., 2015; Savalei et al., 2006).

For this study, the visualization of possible causations among exogeneous and endogeneous variables is stipulated in figure (8.14). The path analysis was used to discern the factors that influence the persons' physical activities. The measure of physical activities was weekly equivalent vigorous PA minutes from transportation collected using PASTA application. The path diagram shows different socioeconomic and demographic variables that were obtained from the survey of participants as already discussed in the data section.

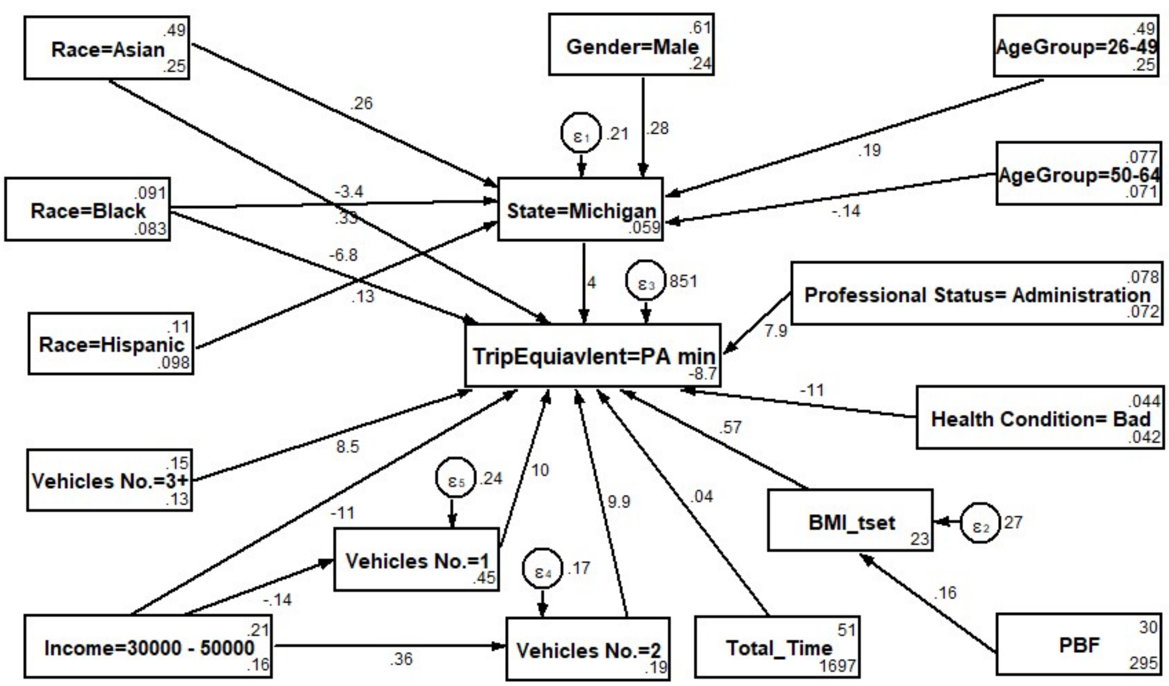



**Figure 8.14 Predictor model of the Physical Activity related to transportation (PA minutes for the**

It includes variables such as race, gender, state that a participant resides, professional status, and health conditions. The body composition test was also included in the model. The direct and indirect effect of different exogenous variables on equivalent PA minutes were specified by drawing an arrow that shows the direction of causation.

Table (12) shows the results of path analysis separated into total, direct, and indirect effects of different variables. Variables that were significant at 95 confidence level (i.e. $p<0.05$) were retained in the model. Variables that had an only direct effect to the weekly equivalent vigorous PA minutes from transportation were the state where the participant is residing, body mass index (BMI), number of vehicles per household and total time spent in different transportation mode as reported by the participants in the survey.. The coefficients that were obtained after running the path analysis were all intuitively correct. On average, participants in Michigan had 4 minutes of weekly equivalent vigorous PA minutes from transportation higher than participants in Texas. A unit increase in total minutes time spent in transportation increased the weekly equivalent vigorous PA minutes from transportation by 0.4. A participants who reported have only one vehicle per household were likely to have higher weekly equivalent vigorous PA minutes from transportation than participants who reported to own more than one vehicles per household. An increase in BMI of a person increased the weekly equivalent vigorous PA minutes from transportation by 0.57. Participants who reported in survey to have bad health condition had 11 minutes decrease of weekly equivalent vigorous PA minutes from transportation compared to participants who had fair to excellent health condition. The professional status of the participants had direct effect on weekly PA minutes. Participants who were working at administrations were likely to have higher weekly equivalent vigorous PA minutes from transportation by 7.9 minutes compared to participants that had another professional status.

Variables that had both direct and indirect effects were race and annual income (in dollars) of the participants. The change in income from earning less than $30,000 a year to $30000-$50,000 a year is likely to decrease the participants' weekly equivalent vigorous PA minutes from transportation by 11 minutes. Income also had an indirect effect moderated by number of vehicles. Higher income was negatively associated with number of vehicles, which in turn reduces the participants' weekly equivalent vigorous PA minutes from transportation.

## Table 8.5 Summary of the significant variables (total, direct, indirect) related to physical activities

| Variables | Total | | | | Direct Effects | | | | Indirect Effects | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Coef. | Std. Err. | z | P>z | Coef. | Std. Err. | z | P>z | Coef. | Std. Err. | z | P>z |
| State =Michigan | | | | | | | | | | | | |
| Gender=Male | 0.285 | 0.024 | 12.010 | **0.000** | 0.285 | 0.024 | 12.010 | **0.000** | 0 | | (no path) | |
| Age Group =26-49 | 0.188 | 0.024 | 8.000 | **0.000** | 0.188 | 0.024 | 8.000 | **0.000** | 0 | | (no path) | |
| Race=Black | 0.133 | 0.039 | 3.430 | **0.001** | 0.133 | 0.039 | 3.430 | **0.001** | 0 | | (no path) | |
| Race=Asian | 0.327 | 0.042 | 7.830 | **0.000** | 0.327 | 0.042 | 7.830 | **0.000** | 0 | | (no path) | |
| Race=Hispanic | 0.255 | 0.028 | 9.160 | **0.000** | 0.255 | 0.028 | 9.160 | **0.000** | 0 | | (no path) | |
| Age Group=50-64 | -0.144 | 0.045 | -3.170 | **0.002** | -0.144 | 0.045 | -3.170 | **0.002** | 0 | | (no path) | |
| Constant | 0.059 | 0.030 | 2.000 | **0.045** | | | | | | | | |
| Trip-equivalent=PA min | | | | | | | | | | | | |
| state=Michigan | 4.038 | 1.483 | 2.720 | **0.006** | 4.038 | 1.483 | 2.720 | **0.006** | 0 | | (no path) | |
| BMI-test | 0.573 | 0.145 | 3.960 | **0.000** | 0.573 | 0.145 | 3.960 | **0.000** | 0 | | (no path) | |
| Vehicles No.=2 | 9.929 | 2.501 | 3.970 | **0.000** | 9.929 | 2.501 | 3.970 | **0.000** | 0 | | (no path) | |
| Vehicles No.=1 | 10.383 | 2.203 | 4.710 | **0.000** | 10.383 | 2.203 | 4.710 | **0.000** | 0 | | (no path) | |
| PBF | | | | | 0 | | (no path) | | 0.092 | 0.024 | 3.900 | **0.000** |
| Total time | 0.040 | 0.017 | 2.300 | **0.022** | 0.040 | 0.017 | 2.300 | **0.022** | 0 | | (no path) | |
| Gender=Male | | | | | 0 | | (no path) | | 1.150 | 0.433 | 2.660 | **0.008** |
| Age Group=26-49 | | | | | 0 | | (no path) | | 0.759 | 0.295 | 2.580 | **0.010** |
| Race=Black | | | | | 0 | | (no path) | | 0.536 | 0.251 | 2.130 | **0.033** |
| Race=Asian | -6.751 | 2.680 | -2.520 | **0.012** | -6.751 | 2.680 | -2.520 | **0.012** | 1.319 | 0.513 | 2.570 | **0.010** |
| Race=Hispanic | -3.371 | 1.731 | -1.950 | **0.051** | -3.371 | 1.731 | -1.950 | **0.051** | 1.031 | 0.395 | 2.610 | **0.009** |
| Vehicles No.=3+ | 8.543 | 2.574 | 3.320 | **0.001** | 8.543 | 2.574 | 3.320 | **0.001** | 0.000 | | (no path) | |
| Income=30000-50000 | -10.815 | 2.194 | -4.930 | **0.000** | -10.815 | 2.194 | -4.930 | **0.000** | 2.108 | 0.808 | 2.610 | **0.009** |
| Health Condition=Bad | -11.386 | 3.846 | -2.960 | **0.003** | -11.386 | 3.846 | -2.960 | **0.003** | 0.000 | | (no path) | |
| Professional Status: Administration | 7.937 | 2.844 | 2.790 | **0.005** | 7.937 | 2.844 | 2.790 | **0.005** | 0.000 | | (no path) | |
| Age Group =50-64 | | | | | 0 | | (no path) | | -0.581 | 0.281 | -2.070 | **0.039** |
| Constant | -8.704 | 4.200 | -2.070 | 0.038 | | | | | | | | |
| BMI Test | | | | | | | | | | | | |
| PBF | 0.161 | 0.007 | 22.760 | **0.000** | 0.161 | 0.007 | 22.760 | **0.000** | 0.000 | | (no path) | |
| Constant | 23.232 | 0.242 | 95.810 | 0.000 | | | | | | | | |
| Vehicles No.=2 | | | | | | | | | | | | |
| Income=30000-50000 | 0.357 | 0.024 | 14.870 | **0.000** | 0.357 | 0.024 | 14.870 | **0.000** | 0.000 | | (no path) | |
| Constant | 0.189 | 0.011 | 17.310 | 0.000 | | | | | | | | |
| Vehicles No.=1 | | | | | | | | | | | | |
| Income=30000-50000 | -0.138 | 0.028 | -4.880 | **0.000** | -0.138 | 0.028 | -4.880 | **0.000** | 0.000 | | (no path) | |
| Constant | 0.452 | 0.013 | 35.060 | 0.000 | | | | | | | | |
| var (State =Michigan) | 0.213 | 0.007 | | | | | | | | | | |
| var(trip equivalent =PA min) | 850.939 | 28.154 | | | | | | | | | | |
| var (BMI Test) | 26.821 | 0.887 | | | | | | | | | | |
| var (Vehicles No.=2) | 0.173 | 0.006 | | | | | | | | | | |
| var (Vehicles No.=1) | 0.241 | 0.008 | | | | | | | | | | |

Variables that had only indirect effects to weekly equivalent vigorous PA minutes from transportation include age and gender which was mediated by participant's state and percent of body fat (PBF) which was mediated by the BMI. As for the age, the results showed that older participants aged 50-64 years had less weekly equivalent vigorous PA minutes from transportation compared to young participants aged 26-49 years. Further, males were found to have more weekly equivalent vigorous PA minutes from transportation compared to females.

## 8.4 Conclusion

This study includes three data sources: questionnaire, InBody test, and daily physical from smartphones and smartwatches. The factors influencing the amount of weekly equivalent vigorous PA minutes from transportation require further investigation because infrastructure and other geospatial factors may play a role as well as trip purpose. This modeling represents a first step in developing more sophisticated mode choice models that seek to include phical activity achieved in the utility function.

# Chapter 9: Integrated Transportation and Health Impacts Model (ITHIM) with a new approach to measuring the relative risk of physical activity related and non-related to travel

## 9.1 Introduction

This section seeks a leading model that measures the health effects associated with transportation. This model may permit more effective coordination public health and transportation planners. A clear vision of a sustainable and integrated transportation system grounded in public health can support health benefits for all.

## 9.2 Research Methodology

This study implements two directions in the research and investigation process. The first reviews the relevant literature, which discusses the integrated transportation and health impact model. The second direction focuses on the development of measurement methods for physical activities related or non-related to transportation. Figure 9.1 shows the mechanism to implement the study and achieve the desired objectives.

The topic of the health effects of transportation includes literature that covers a variety of topics: the environmental impacts of air pollution and noise, the impact of traffic crashes, and the impact of physical activities related to transportation. The literature often uses the disability-adjusted life years (DALYs) as an essential measurement tool for risk or benefit assessment. The research also focuses on studies that involve more than one factor of influence because the research objectives focus on integrated measurement or evaluation methods. The researchers removed many articles by requiring the research articles to include "DALYs" as a robust criterion for examination and review.
For the second direction of the study, the study adopts an assessment process known as the comparative risk assessment (CRA) for the identification of integrated transportation and health effects (ITHIM). In the CRA method, the population attributable fraction (PAF) formula is applied to each of the factors influencing health, that come from transportation (e.g. level of community physical

activity, level of crashes, and level of air pollution and noise). The CRA method identifies changes in the disease burden (DB) when exposed to the risks and benefits of different transportation. Relative risk (RR) determines the value of the PAF, which is defined as "the probability of a situation if a particular variable is exposed to risk if not detected". To complete this part of the study, the research team use one-week data from the Kalamazoo participants.

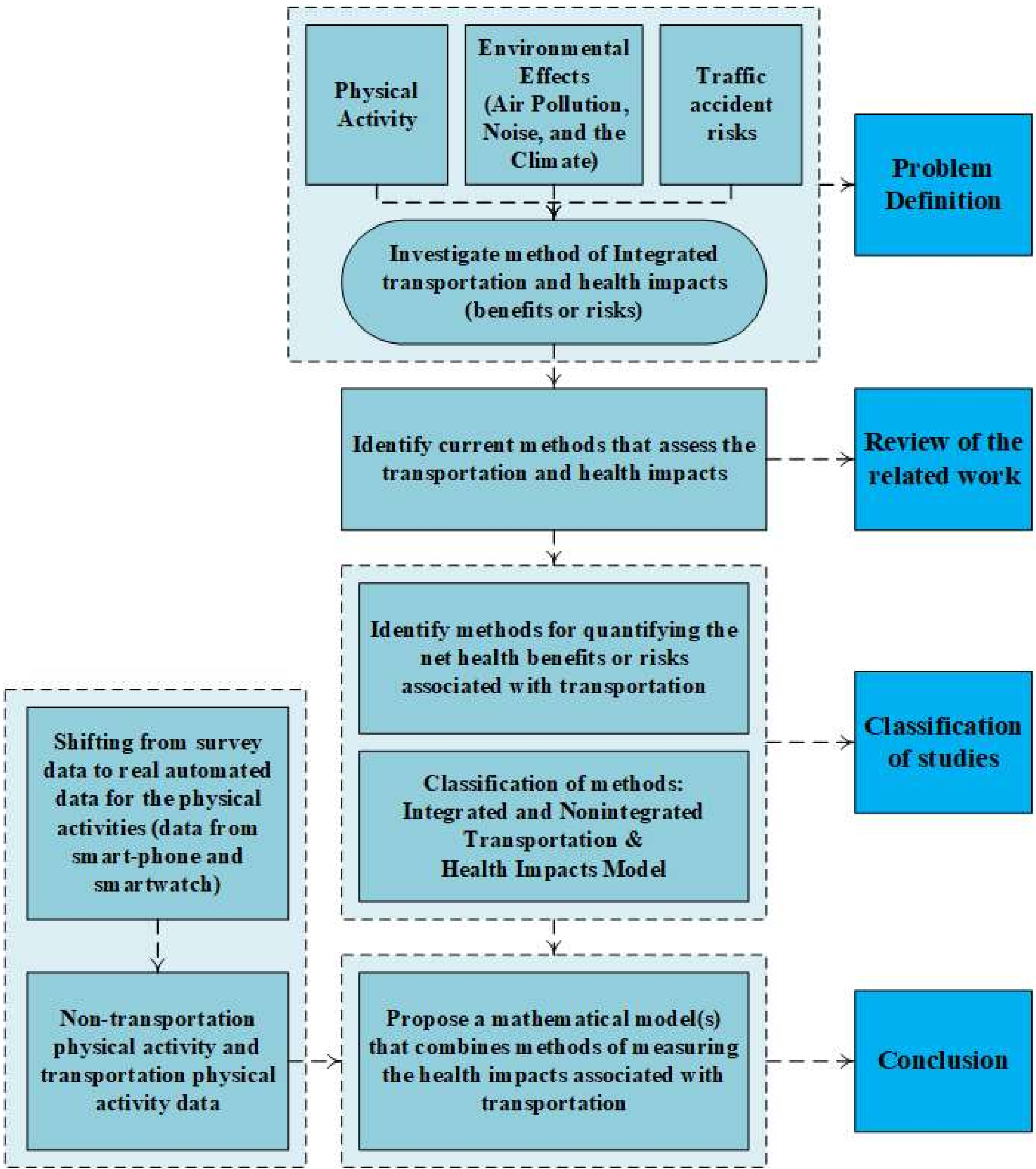



**Figure 9.1 A framework for the Study of Integrated measurement methods for health effects of transportation**

## 9.3 Results and Discussion

### 9.3.1 Identify Literature Relevant to The Impact of Transportation on Health

Figure 9.2 shows the flow of the literature research process. of the review identifies 3373 articles on the health effects of transportation, including the "DALYs" model as a measurement tool, and through the databases of the Engineering Village and the PubMed. The review process excludes 22 duplicate articles, as well as the exclusion of 1,373 articles when determining the "Subject/Title/Abstract." The process excludes 167 articles that do not appear as conference papers, papers, or reports and 1,193 articles during the abstract review due to mismatch with the subject of the study. The procedure also excludes 29 articles written in languages other than English and 545 articles published before 2000.

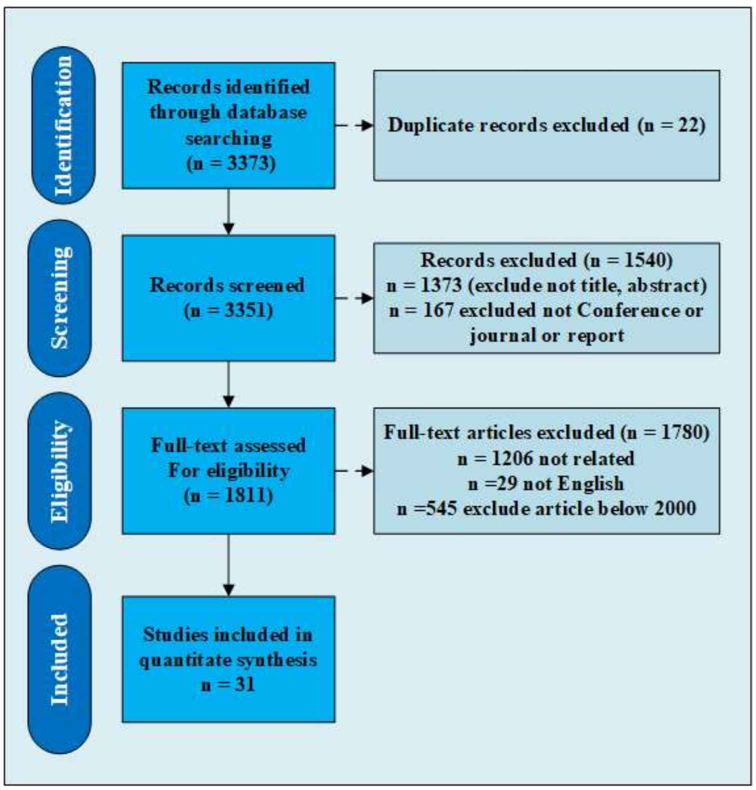



**Figure 9.2 Diagram illustrating the number of articles excluded through the title and abstract analyses.**

Table 9-1 presents a list of the remaining (31) studies that relate to the health effects of transportation. In general, the studies include nineteen on traffic safety risks were identified, fifteen on air pollution, seven n noise pollution, one on climate effects, and twelve on physical activity. These studies may be classified into four categories (benefit assessment, cost-benefit assessment, risk assessment, and comparative risk) for assessing the health effects of transportation.

**Table 9.1 Details of relevant reviews and detailed information for each study.**

| | Author, date, and Location | The title for each study | The focus of the study | | | | | methods used to assess HIT |
|---|---|---|---|---|---|---|---|---|
| | | | Traffic Accident | Air pollution | Noise | Climate | PA | |
| 1 | Ria et al., 2008, Belgium | Environmental burden of disease due to transportation noise in Flanders (Belgium) | | | ● | | | Risk assessment |
| 2 | Suzanne et al., 2015, Netherlands | Burden of road traffic injuries: Disability-adjusted life years in relation to hospitalization and the maximum abbreviated injury scale | ● | | | | | Risk assessment |
| 3 | Stijn et al., 2011, Vietnam | Environmental health impacts of mobility and transport in Hai Phong, Vietnam | ● | ● | ● | | ● | Review without DAYLs as a tool measurement |
| 4 | Ting et al., 2015, Australia | Traffic-related air pollution and health co-benefits of alternative transport in Adelaide, South Australia | | ● | | | ● | Benefit assessment |
| 5 | Natalie et al., 2017, Switzerland | Health impacts related to urban and transport planning: A burden of disease assessment | | ● | ● | ● | ● | Risk assessment |
| 6 | Rodrigues, Rui Calejo, 2018, Portugal | Quiet areas and urban sustainability | | | ● | | | Risk assessment |
| 7 | Tunde O et al., 2017, United States | The health burden and economic costs averted by ambient PM2.5pollution reductions in Nagpur, India | | ● | | | | Risk assessment |
| 8 | Hérick de et al., 2017, Brazil | Health impact modelling of different travel patterns on physical activity, air pollution and road injuries for São Paulo, Brazil | ● | ● | | | ● | Risk assessment |
| 9 | Nilsson et al., 2017, Sweden | Modelling the effect on injuries and fatalities when changing mode of transport from car to bicycle. | ● | | | | | Risk assessment |
| 10 | Marko Tainio 2015, Poland | Burden of disease caused by local transport in Warsaw, Poland. | ● | ● | ● | | ● | Review |
| 11 | Woodcock et al., 2017, United Kingdom | Health effects of the London bicycle sharing system: health impact modelling study. | ● | ● | | | ● | Risk assessment using MET for Physical Activity |
| 12 | Woodcock et al., 2013, United Kingdom | Health impact modelling of active travel visions for England and Wales using an Integrated Transport and Health Impact Modelling Tool (ITHIM). | ● | ● | | | ● | Risk assessment using MET for Physical Activity |
| 13 | Eriksson et al., 2017, Sweden | Burden of disease from road traffic and railway noise - a quantification of healthy life years lost in Sweden. | | | ● | | | Risk assessment |
| 14 | Tetreault et al., 2018, Canada | Estimating the health benefits of planned public transit investments in Montreal. | ● | ● | | | ● | risk assessment framework (CRA) using MET |
| 15 | Furberg et al., 2018, Sweden | Live and Let Die? Life Cycle Human Health Impacts from the Use of Tire Studs. | ● | ● | | | | Risk assessment, but the study not concerned to active transportation |
| 16 | Chapman et al., 2018, New Zealand | A Cost Benefit Analysis of an Active Travel Intervention with Health and Carbon Emission Reduction Benefits. | ● | ● | | | ● | Cost-effectiveness |
| 17 | Paunovic et al., 2014, Belgrade | Burden of myocardial infarction attributable to road-traffic noise: a pilot study in Belgrade. | | | ● | | | Risk assessment |

**Table 9.1 Details of relevant reviews and detailed information for each study. (Cont.)**

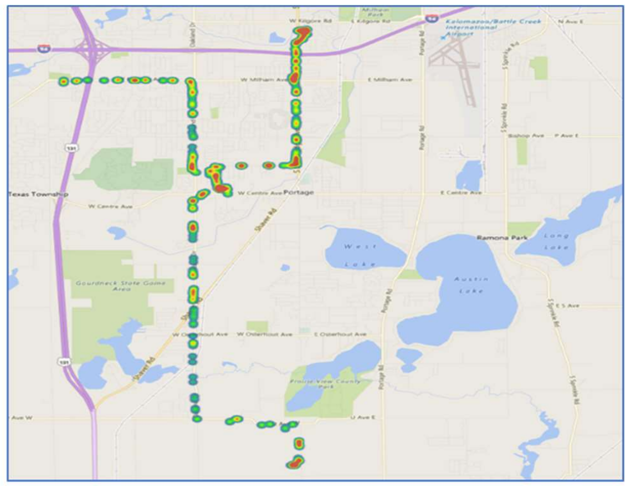
| | Author, date, and Location | The title for each study | The focus of the study | | | | | The focus of the study |
|---|---|---|---|---|---|---|---|---|
| | | | Traffic Accident | Air pollution | Noise | Climate | PA | |
| 18 | Jarjour et al., 2014, United States | Cyclist route choice, traffic-related air pollution, and lung function: a scripted exposure study. | | ● | | | | Risk assessment |
| 19 | Rojas-Rueda et al., 2013, Spain | Health impact assessment of increasing public transport and cycling use in Barcelona: a morbidity and burden of disease approach. | ● | ● | | | ● | Risk assessment |
| 20 | Tainio et al., 2014, Swedish | Severity of injuries in different modes of transport, expressed with disability-adjusted life years (DALYs). | ● | | | | | Risk assessment |
| 21 | Perez et al., 2015, Switzerland | Transport-related measures to mitigate climate change in Basel, Switzerland: A health-effectiveness comparison study. | | ● | | | | Risk assessment |
| 22 | Dhondt et al., 2013, Belgium | Integrated health impact assessment of travel behaviour: model exploration and application to a fuel price increase. | | ● | | | ● | Comparative Risk Assessment (CRA) |
| 23 | GBD 2015 Eastern Mediterranean Region Transportation Injuries Collaborators | Transport injuries and deaths in the Eastern Mediterranean Region: findings from the Global Burden of Disease 2015 Study. | ● | | | | | Overview the burden of Transport injuries |
| 24 | Banstola et al., 2016, Low- and Middle-Income Countries (LMICs) | Cost-effectiveness of interventions to prevent road traffic injuries in low- and middle-income countries: A literature review. | ● | | | | | There is no measurement method for assessment |
| 25 | Stewart et al., 2015, Solomon Islands | Extent, causes and impact of road traffic crashes in the Solomon Islands 1993-2012: data from the orthopaedic department at the National Referral Hospital, Honiara. | ● | | | | | There is no measurement method for assessment |
| 26 | Moodie et al., 2009, Australia | Cost-effectiveness of active transport for primary school children - Walking School Bus program. | | | | | ● | Cost-effectiveness |
| 27 | Chong et al., 2010, Australia | Relative injury severity among vulnerable non-motorized road users: comparative analysis of injury arising from bicycle-motor vehicle and bicycle-pedestrian collisions. | ● | | | | | Risk assessment |
| 28 | Bijkerk et al., 2019, Netherland | Quantitative health impact assessment of transport policies: two simulations related to speed limit reduction and traffic re-allocation in the Netherlands. | ● | | | | | Risk assessment |
| 29 | Margie Peden 2007, Switzerland | Global collaboration on road traffic injury prevention. | ● | | | | | Cost-effectiveness |
| 30 | Chisholm et al., 2012, sub-Saharan Africa and South East Asia | Cost effectiveness of strategies to combat road traffic injuries in sub-Saharan Africa and South East Asia: mathematical modelling study. | ● | | | | | Cost-effectiveness |

Two particular studies discuss the comparative risk assessment (CRA) involving integrated transportation modeling and health impacts (ITHIM). The term integration appears to be expressed through studies that include two or more transportation impacts on health in one study. The ITHIM framework provides a balance against most studies that emphasize transportation risks by focusing on transportation benefits.

### 9.3.2 Physical Activities Related and Non-Related to Transportation

Figure 9.3 shows a sample of the display interfaces in the PASTA platform, which provides reports on the transportation (trip) or non-trip activity of the participants. Previous research demonstrates the importance of monitoring physical activities as a crucial factor in the application of ITHIM, and

| Type | Start | End | Activity Duration | HR Duration | Data Availability |
|---|---|---|---|---|---|
| Home | 12:00 AM | 06:52 AM | 412 | 397 | 96.36 % |
| Trip | 06:53 AM | 07:01 AM | 8 | 8 | 100 % |
| Church | 07:13 AM | 10:27 AM | 194 | 194 | 100 % |
| Trip | 10:27 AM | 10:36 AM | 9 | 9 | 100 % |
| Home | 10:36 AM | 12:11 PM | 95 | 95 | 100 % |
| Trip | 12:11 PM | 12:18 PM | 7 | 7 | 100 % |
| Pharmacy | 12:18 PM | 12:55 PM | 37 | 37 | 100 % |
| Trip | 12:56 PM | 01:02 PM | 6 | 6 | 100 % |
| Supermarket | 01:03 PM | 01:19 PM | 16 | 16 | 100 % |
| Trip | 01:19 PM | 01:31 PM | 12 | 12 | 100 % |
| Home | 01:31 PM | 05:35 PM | 244 | 244 | 100 % |
| Trip | 05:35 PM | 05:49 PM | 14 | 14 | 100 % |
| Home | 05:58 PM | 09:12 PM | 194 | 194 | 100 % |



| Type | Start | End | Activity Duration | HR Duration | Data Availability | Calories | Transportation Mode | Steps | HR Average | GeoHash | Very Light | Light | Moderate | Hard | Very Hard | Maximal | PAM | Total HR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Home | 12:00 AM | 06:52 AM | 412 | 397 | 96.36 % | 526 | | 854 | 62 | dpe05f | 0 % | 93.45 % | 2.27 % | 3.53 % | 0 % | 0.76 % | 6.92 | 24789 |
| Trip | 06:53 AM | 07:01 AM | 8 | 8 | 100 % | 14 | | 58 | 69 | dpe05c | 0 % | 77.78 % | 0 % | 0 % | 0 % | 22.22 % | 0.1 | 623 |
| Church | 07:13 AM | 10:27 AM | 194 | 194 | 100 % | 376 | | 1371 | 74 | dp7psm | 0 % | 42.05 % | 47.69 % | 1.03 % | 0 % | 9.23 % | 12.95 | 14469 |
| Trip | 10:27 AM | 10:36 AM | 9 | 9 | 100 % | 15 | | 13 | 72 | dp7psq | 0 % | 70 % | 20 % | 0 % | 0 % | 10 % | 0.4 | 724 |
| Home | 10:36 AM | 12:11 PM | 95 | 95 | 100 % | 147 | | 325 | 70 | dpe05f | 0 % | 70.83 % | 18.75 % | 1.04 % | 0 % | 9.38 % | 2.55 | 6699 |
| Trip | 12:11 PM | 12:18 PM | 7 | 7 | 100 % | 13 | | 8 | 74 | dpe05g | 0 % | 50 % | 25 % | 12.5 % | 0 % | 12.5 % | 0.56 | 592 |
| Pharmacy | 12:18 PM | 12:55 PM | 37 | 37 | 100 % | 71 | | 268 | 67 | dpe0hr | 0 % | 78.95 % | 15.79 % | 0 % | 0 % | 5.26 % | 0.56 | 2563 |
| Trip | 12:56 PM | 01:02 PM | 6 | 6 | 100 % | 25 | | 284 | 75 | dpe0k2 | 0 % | 57.14 % | 42.86 % | 0 % | 0 % | 0 % | 0.7 | 528 |
| Supermarket | 01:03 PM | 01:19 PM | 16 | 16 | 100 % | 55 | | 470 | 85 | dpe0k6 | 0 % | 0 % | 82.35 % | 5.88 % | 0 % | 11.76 % | 2.83 | 1445 |
| Trip | 01:19 PM | 01:31 PM | 12 | 12 | 100 % | 19 | | 27 | 74 | dpe0k3 | 0 % | 30.77 % | 53.85 % | 0 % | 0 % | 15.38 % | 0.68 | 961 |
| Home | 01:31 PM | 05:35 PM | 244 | 244 | 100 % | 457 | | 1834 | 72 | dpe05f | 0 % | 55.51 % | 28.57 % | 4.49 % | 0 % | 11.43 % | 13.19 | 17704 |
| Trip | 05:35 PM | 05:49 PM | 14 | 14 | 100 % | 17 | | 0 | 62 | dpe05g | 0 % | 93.33 % | 0 % | 0 % | 0 % | 6.67 % | 0.04 | 937 |
| Home | 05:58 PM | 09:12 PM | 194 | 194 | 100 % | 260 | | 302 | 65 | dpe05f | 0 % | 90.77 % | 5.64 % | 0.51 % | 0 % | 3.08 % | 2.39 | 12749 |

**Figure 9.3 Sample of the results from the PASTA application**

Figure (9.4 shows this study's contribution to move the monitoring physical activities from the qualitative data used by Wu et al. (2019) to quantitative data.

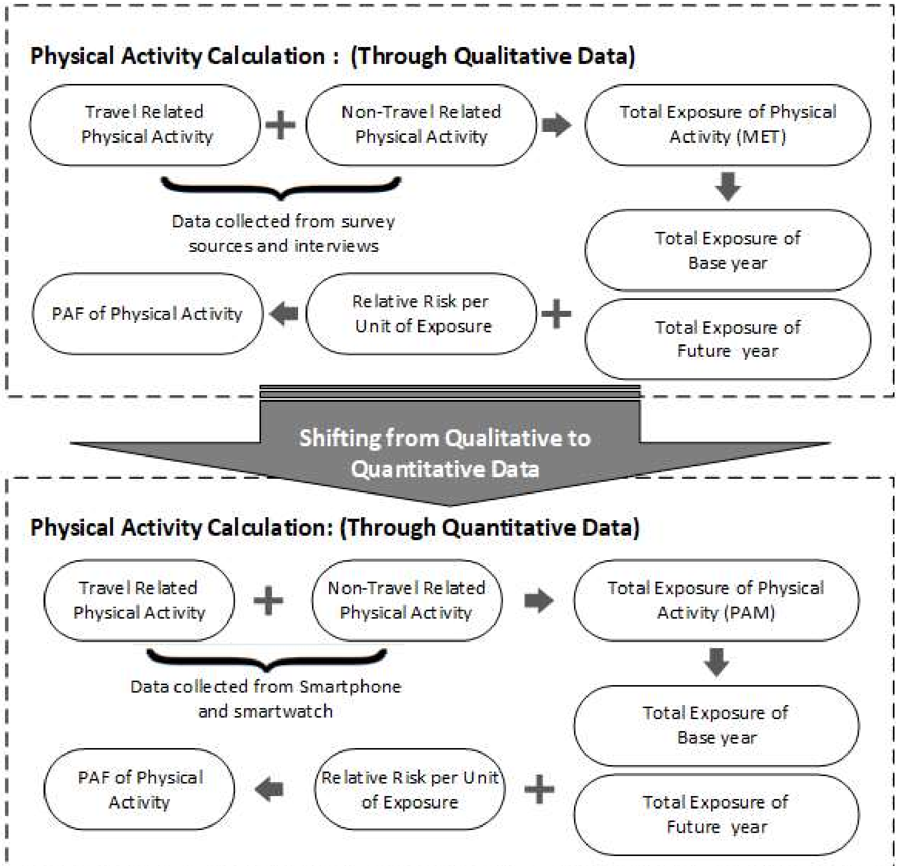



**Figure 9.4 Shifting from qualitative data to quantitative data for the Physical Activities impact factor**

The previous studies represent the starting point for a new concept of automatic PA related or non-related to transportation data assembly. The first part of the diagram, which is adopted in Wu et al. (2019), most of the data derive from questionnaires and interviews. They seek to redesign the mathematical model used in calculating the Relative Risks (RRs) of PA. The main element of the RRs model is METs [equation (9.1)] for both types of physical activities (travels related physical activities and non-travel related physical activities).

$$\boldsymbol{RRs = 0.94^{\sqrt{METs}}} \qquad \textbf{(9.1)}$$

They use recorded data from surveys, questionnaires, and interviews, which depend on the subject's memory to provide the METs data; As a result, the METs data are often characterized by inaccuracy, bias, discontinuity, and high cost. Wu et al. (2019) identify the need to use two types of smart devices that ensure the availability of data that define the features of physical activities, namely smartphones and smartwatches. Therefore, a shift to automated data collection to provide PA data seems to be the most appropriate solution, through continuous, accurate, inexpensive, and geographically unlimited data. This transformation can also be described as a shift from qualitative to quantitative data. A second shift must also appear because the MET has a constant value for each transportation mode (e.g. bus, driving, bike, and walking). The new data collection approach accounts for the PA directly, which is variable within each travel mode, as illustrated in the second part of the Figure 9.4 using the Physical Activity Minutes (PAM) value. The PAM value is calculated by the relationship shown in Equation (9.2), which collects the percentage heart rate reserve (%HRR) shown in equation (9.2).

$$PAM = \int_{Start}^{End} \%HRR \cdot dt \qquad \textbf{(9.2)}$$

Where:

PAM = Physical Activity Minutes

%HRR = Percentage heart rate reserve

Note that, %HRR value was calculated [shown in equation (9.3)] based on (Nakanishi et al. 2018)

$$\%HRR = \frac{HR_{act} - HR_{rest}}{HR_{max} - HR_{rest}} \qquad \textbf{(9.3)}$$

Where:

HRact = Heart rate during activity

HRrest = Resting heart rate

HRmax = Maximum heart rate

Figure (9.5) shows some of the most important results from the overall study, which show the speed, the difference between heart rate activity and heart rate rest for each activity ($HR_{act}$- $HR_{rest}$), calories and PAM, for different transportation modes. The transportation mode choice results in different levels of physical activity, which certainly produces different health outcomes. Physical activity appears highest with cycling, it appears lower with walking, and it decreases further when using vehicles.

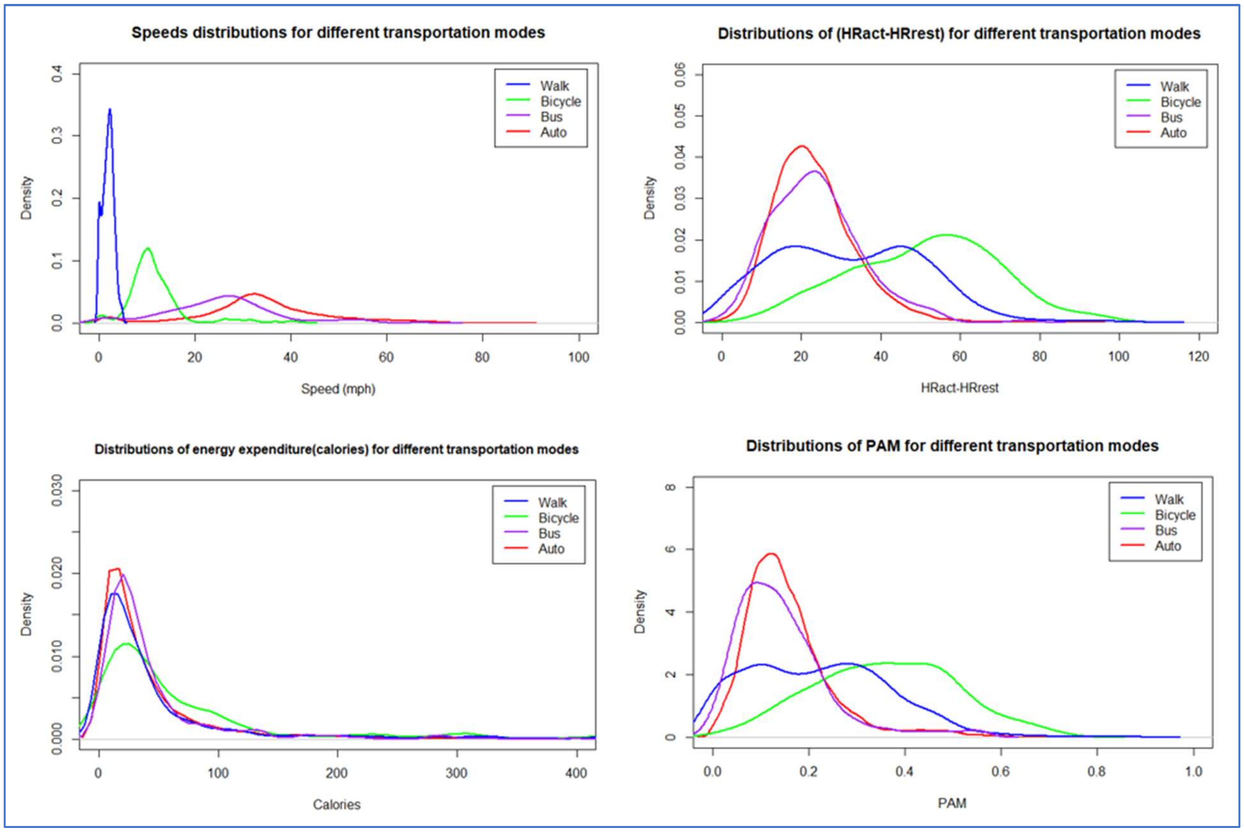



**Figure 9.5 Shifting from qualitative data to quantitative data for the Physical Activities impact factor**

Also, the researchers evaluate the relationship between the METs calculated from calories and the PAM values calculated from the heart rate (both of the data derived from the smartwatches). Table (9.2) shows the linear regression results for the relationship between METs and PAM.

**Table 9.2 Relationship between PAM and METs in (A) and (B) for travel related and non-travel related PA**

| Relationship between PAM and METs for travel related to PA | | | | | | |
|---|---|---|---|---|---|---|
| **Travel Related METs** | Coef. | Std. Err. | t | P>t | [95% Conf. Interval] | |
| **PAM -average** | 6.244 | 0.477 | 13.1 | 0 | 5.306 | 7.182 |
| **Constant** | 0.834 | 0.092 | 9.03 | 0 | 0.653 | 1.016 |
| **Auxiliary Statistics**<br><br>Number of obs = 284<br>F(1, 282) = 171.66<br>Prob > F = 0.0000<br>R-squared = 0.3784<br>Adj R-squared = 0.3762<br>Root MSE = 0.90736 | | | | | | |
| **Relationship between PAM and METs for travel non-travel related to PA** | | | | | | |
| **Non-Travel Related METs** | Coef. | Std. Err. | t | P>t | [95% Conf. Interval] | |
| **PAM -average** | 3.148 | 0.294 | 10.7 | 0 | 2.569 | 3.726 |
| **Constant** | 1.421 | 0.048 | 29.51 | 0 | 1.326 | 1.515 |
| **Auxiliary Statistics**<br><br>Number of obs = 344<br>F(1, 342) = 114.58<br>Prob > F = 0.0000<br>R-squared = 0.2509<br>Adj R-squared = 0.2488<br>Root MSE = 0.56794 | | | | | | |

Figure 9.6 shows a comparative sample of the MET values obtained for real samples using different transportation modes with standard METs values. The study and through the data recorded by the PASTA platform proved the opportunities provided by the study to overcome the limitations faced by previous studies in determining the values of METs. It is difficult to guess the levels of physical activity of people through different modes of transportation by questionnaires, interviews or even self-logging.

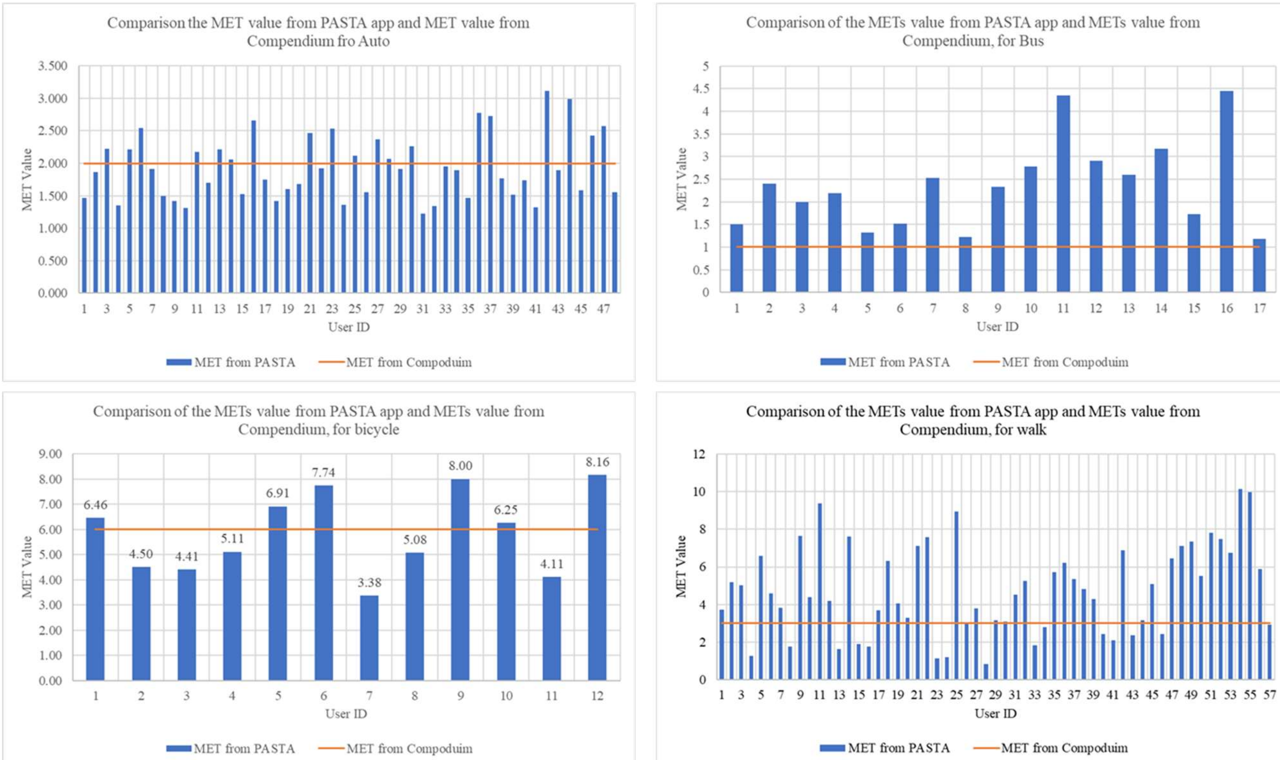



**Figure 9.6. A comparison of the standard MET values for the four common Physical Activities related to transportation with the actual measured MET**

## 9.4 Conclusion

The study reviews research on the assessment of the health impacts of transportation through the assessment of health risks or benefits. In addition, this study tries to identify the measurement methods used in the evaluation processes to produce an integrated assessment system. Considerable emphasis had been placed on review studies, as well as on studies linking more than one source on the impact of transportation on health. The study reveals through the literature the health risks of air pollution and noise pollution and traffic crashes caused by transportation. In addition, the health benefits of active transportation, which stem from physical activity, yield health benefits.

Previous studies identify four ways to assess the health impacts of transportation: benefit assessment, cost-benefit assessment, risk assessment, and comparative risk assessment. The comparative risk assessment method reveals the level of risk or benefit according to the differentiation of the influencing factors, and the ITHIM model adopts this approach. The ITHIM evaluation model produces tools that promote transportation policies to encourage the use of active transportation to increase its health benefits, with the potential to improve the environment, diversify transportation patterns, and reduce crashes. The study also presents an enhanced approach to improving the assessment mechanism for the physical activity of individuals in the ITHIM model. This study reveals the limitations within the previous evaluation mechanisms, which may be overcome by the spread of smart devices.

# Chapter 10: Conclusion and Recommendation

Public health importance of physical activity continues to attract attention, and the lack of physical activities may cause health problems. Travel activities provide a certain amount of physical activity, and active transportation, such as walking and cycling, may represent an essential tool for both transportation planners and public health officials. Active transportation may contribute to improving human health by reducing cardiovascular disease, obesity, and premature death; however, the detailed relationship between transportation mode choices and human health remains poorly understood. Therefore, the need to investigate traveler behaviors and their effect on physical activity and public health seems critical. This study analyzes and quantifies participants' actual physical activity by using wearable devices with sensing and GPS tracking technology. The study seeks to characterize the health outcomes from the physical activity associated with transportation options.

The research team develops a mobile application platform named Physical Activity through Smart Travel Activity (PASTA) to monitor the travel and physical activities of transportation users. The mobile app collects daily travel activities including locations from the GPS in the subject's mobile phone and physical activity data from a wearable device (Fitbit Charge 2 and 3). The PASTA platform includes mobile data communication, big data analysis, activity classification, transportation mode detection, and physical activity quantification on different interfaces, such as smartphones, cloud databases, and computers. The platform provides data to compare physical activities attributable to transportation across different geographical areas.

The study tests the PASTA platform in Texas and Michigan using a total of 120 participants and proved to be useful in apportioning the total physical activity into travel-related physical activities and non-travel related physical activities. The survey gathers participants' demographic, social, economic, travel acitivity pattern, and associated physical activy characteristics. From the survey responses, a similar distribution of active and highly active participants appear in Kalamazoo (45.8%) and Arlington (44.8%); however, the obesity rate remains higher at WMU than UTA (33.9% vs. 25.9%) based on BMI. Based on the travel mode usage, more than 50% of the participants use private vehicles for commuting for both study areas. Arlington subjects use active transportation (38% participants) as their second highest transportation mode usage while Kalamazoo participants use public transit. Since the Arlington participants frequently use bicycle and walking as their transportation mode, they seem more physically active in comparison to Kalamazoo participants. The researchers conduct a cross tabulation analysis to compare the perceived health and their physical activity, where the overall results show that perceived health does not always align with objective health measures such as BMI and physical activity level.

This study develops three different approaches to identify and recognize transportation user activities and trips based on GPS trajectories. The approaches apply different thresholds of spatiotemporal change by developing Geohash clustering, GIS-based approach, and an integrated Geohash-GIS system for activity only, trip only, and sequential activity-trip recognition with GPS data. The Combined Geohash-GIS approach with a dwell time of 5 minutes produces the best accuracy, which could significantly enhance the efficiency and accuracy of GPS travel survey by correctly (about 88%) recognizing user activity and trip patterns. This proposed combined approach could serve as a foundation for a future data collection of full-scale traveler information identification with GPS data. The study develops machine learning models to predict transportation mode from smartphone and smartwatch data. The models distinguish between transportation activities (trips) and non-related to transportation activities (e.g. home, work, and shopping). Among the four machine learning models, Random Forest (RF) outperforms the other models in detecting non-motorized modes: walking (97.2%) and bicycle (90.6%).

This study conducts a comprehensive descriptive and path analysis to explore the relationship between the levels of physical activity of individuals, their socio-economic features and body shapes by using questionnaire, in-body composition tests, and physical activity data. The BMI, baseline active transportation time, age and gender have a direct effect to the weekly equivalent vigorous PA minutes from transportation, while PBF has only an indirect effect. Race and annual income have both direct and indirect effects. This research also develops an enhancement to the existing integrated transportation and health impact model (ITHIM) by adding the quantitative PA data obtained from PASTA, which substantially reduces the previous ITHIM limitations.

The findings of this study help in incorporating human health into transportation planning by addressing health outcomes from the physical activity associated with transportation choices. However, there are some limitations of this study. Although the research team successfully developed and collected data from the mobile application, the GPS accuracy was very poor at inside the building structures. The mobile application was unable to capture data for the iOS-enabled phone users. The research team also faced the problem of determining the in-between walking trips from home to parking-lot or between two buildings. In terms of physical activity computation, there were some error occurred due to the misleading step value calculation while driving/resting (Fitbit considers a step at the time of moving your hand over driving wheel) position. In the study area, majority of the peoople are auto users and therefore, it was really a challenging issue to collect data for other transportation modes which could enhance the comprehensive health analysis outcomes from different transportation choices.

This study provides information that can be used to enhance community awareness of the health benefits that result from different transportation mode choices. This research contributes to integrating human health into transportation planning by addressing health outcomes impacted by the physical and cardiovascular activities associated with transportation options. As the future direction of this research, the resaerch team could upgrade their mobile application by incorporating iOS users and enhance the accuracy of monitoring transportation activity/trip in a greater scale. Different environmental characteristics (e.g. weather, temperature, etc.), and built-environmental characteristics (e.g. landuse, accessibility to physical activity facilities, park accessibility, etc.) could be incorporated to assess the physical activity level or health outcome for using different transportation options. In addition, different attributes of active travel environment (e.g. bikeability, walkability, transit accessibility, etc.) could be incorporated to evaluate the integrated transportation and health outcome modelling. In overall, there is a strong need for further assessment of health outcome modelling based on different emerging active transportation modes and facilities by incorporating other attributes related or non-related to transportation.

# Reference


1. Active Living Research. Moving Toward Active Transportation . (ALRMTAT). (2016). How Policies Can Encourage Walking and Bicycling. 2016.
2. Adesiyun, T.A. and Russell, S.D. (2018). Exercise, Fitness, and Cancer Outcomes. In *Lifestyle in Heart Health and Disease* (pp. 99-114). Academic Press.
3. Ahima, R.S. and Lazar, M.A. (2013). The health risk of obesity—better metrics imperative. *Science*, *341*(6148), pp.856-858.
4. Akinkunmi, M. (2019). Introduction to Statistics Using R.
5. Alanazi, H. O., A. H. Abdullah, and Qureshi, K. (2017). A Critical Review for Developing Accurate and Dynamic Predictive Models Using Machine Learning Methods in Medicine and Health Care. Journal of Medical Systems, Vol. 41, No. 4, 2017. https://doi.org/10.1007/s10916-017-0715-6.
6. Ameri, E., Dehkhoda, M.R. and Hemayattalab, R. (2012). Bone mineral density changes after physical training and calcium intake in students with attention deficit and hyper activity disorders. *Research in developmental disabilities*, *33*(2), pp.594-599.
7. America's Health Rankings Annual Report. (2016). Accessed from https://americashealthrankings.org/explore/annual/measure/Sedentary/state/ALL?edition-year=2016. Access date: 08/20/2019
8. American Cancer Society. (2018). *Global Cancer Facts & Figures 4th Edition*. Atlanta: American Cancer Society.
9. American Heart Association (2018). Heart disease and stroke statistics 2018 at-a-glance. https://www.heart.org/-/media/data-import/downloadables/heart-disease-and-stroke-statistics-2018---at-a-glance-ucm_498848.pdf, Accessed July 5, 2019.
10. Anderson, L.H., Martinson, B.C., Crain, A.L., Pronk, N.P., Whitebird, R.R., O'Connor, P.J. and Fine, L.J. (2005). PEER REVIEWED: Health Care Charges Associated with Physical Inactivity, Overweight, and Obesity. *Preventing chronic disease*, *2*(4).
11. Annual Estimates of the Resident Population. (2018). Accessed from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF. Accessed 08/20/019. Source: U.S. Census Bureau, Population Division. Release Date: April 2019.

12. Ansari, Z., and A. Golroo. Automated Transportation Mode Detection Using Smart Phone Applications via Machine Learning: Case Study Mega City of Tehran. Transportation Research Board 94th Annual Meeting, 2015.
13. Antar, A., and Ahmed, M. (2018). A Comparative Approach to Classification of Locomotion and Transportation Modes Using Smartphone Sensor Data. No. February 2019, 2018. https://doi.org/10.1145/3267305.3267516.
14. Arasheben, A., Barzee, K.A. and Morley, C.P. (2011). A meta-analysis of bone mineral density in collegiate female athletes. *J Am Board Fam Med*, *24*(6), pp.728-734.
15. Asgari, F., and Clemencon, S. (2018). Transport Mode Detection When Fine-Grained and Coarse-Grained Data Meet. 2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE), 2018, pp. 301–307.
16. Ashqar, H. I., M. H. Almannaa, M. Elhenawy, H. A. Rakha, and House, L. (2019). Smartphone Transportation Mode Recognition Using a Hierarchical Machine Learning Classifier and Pooled Features From Time and Frequency Domains. IEEE Transactions on Intelligent Transportation Systems, Vol. 20, No. 1, 2019, pp. 244–252. https://doi.org/10.1109/TITS.2018.2817658.
17. Aune, D., Norat, T., Leitzmann, M., Tonstad, S. and Vatten, L.J. (2015). Physical activity and the risk of type 2 diabetes: a systematic review and dose–response meta-analysis.
18. Axhausen, K. W., Schönfelder, S., Wolf, J., Oliveira, M., & Samaga, U. (2004, January). Eighty weeks of gps traces, approaches to enriching trip information. In Transportation Research Board 83rd Annual Meeting Pre-print CD- ROM.
19. Balli, S., and Sagbas, E. (2018). Diagnosis of Transportation Modes on Mobile Phone Using Logistic Regression Classification. IET Software, Vol. 12 Iss. 2, 2018, pp. 142–151. https://doi.org/10.1049/iet-sen.2017.0035.
20. Barengo, N.C., Hu, G., Lakka, T.A., Pekkarinen, H., Nissinen, A. and Tuomilehto, J. (2004). Low physical activity as a predictor for total and cardiovascular disease mortality in middle-aged men and women in Finland. *European Heart Journal*, *25*(24), pp.2204-2211.
21. Bayarma, A., Kitamura, R. and Susilo, Y.O. 2007. On the recurrence of daily travel patterns: a stochastic-process approach to multi-day travel behavior. Transportation Research Record, 2021: 55–63.

22. Bedogni, L., M. Felice, and Bononi, L. (2016). Context-Aware Android Applications through Transportation Mode Detection Techniques. Wireless communications and mobile computing, No. July, 2016, pp. 2523–2541. https://doi.org/10.1002/wcm.
23. Beltra, H. (2015). Past, Present, and Future of Healthy Life Expectancy. 2015, pp. 1–11.
24. Bhaskaran, K., Douglas, I., Forbes, H., Dos-Santos-Silva, I., Leon, D.A. and Smeeth, L. (2014). Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5· 24 million UK adults. *The Lancet*, *384*(9945), pp.755-765.
25. Bielemann, R.M., Martinez-Mesa, J. and Gigante, D.P. (2013). Physical activity during life course and bone mass: a systematic review of methods and findings from cohort studies with young adults. *BMC musculoskeletal disorders*, *14*(1), p.77.
26. Boccuzzi S.J. (2003) Indirect Health Care Costs. In: Weintraub W.S. (eds) *Cardiovascular Health Care Economics. Contemporary Cardiology*. Humana Press, Totowa, NJ
27. Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. Transportation Research Part C: Emerging Technologies, 17, 285-297.
28. Bonora, E., Kiechl, S., Willeit, J., Oberhollenzer, F., Egger, G., Meigs, J.B., Bonadonna, R.C. and Muggeo, M. (2004). Population-based incidence rates and risk factors for type 2 diabetes in white individuals: The Bruneck study. *Diabetes, 53*(7), pp.1782-1789.
29. Bouchard, C., Blair, S.N. and Haskell, W.L. (2007). Why study physical activity and health. *Physical activity and health*, *1*, pp.3-20.
30. Boyle, T., Keegel, T., Bull, F., Heyworth, J. and Fritschi, L. (2012). Physical activity and risks of proximal and distal colon cancers: a systematic review and meta-analysis. *Journal of the national cancer institute*, *104*(20), pp.1548-1561.
31. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, *68*(6), pp.394-424.
32. Brichetto, G., Pedullà, L., Podda, J., & Tacchino, A. (2019). Beyond center-based testing: Understanding and improving functioning with wearable technology in MS. Multiple Sclerosis Journal, 25(10), 1402–1411. https://doi.org/10.1177/1352458519857075.

33. Buliung, R.N., Roorda, M.J. and Remmel, T.K. 2008. Exploring spatial variety in patterns of activity-travel behaviour: initial results from the Toronto travel-activity panel survey (TTAPS). Transportation, 35: 697–722.
34. Burke, V., Zhao, Y., Lee, A.H., Hunter, E., Spargo, R.M., Gracey, M., Smith, R.M., Beilin, L.J. and Puddey, I.B. (2007). Predictors of type 2 diabetes and diabetes-related hospitalisation in an Australian Aboriginal cohort. *Diabetes research and clinical practice*, *78*(3), pp.360-368.
35. Cadilhac, D.A., Cumming, T.B., Sheppard, L., Pearce, D.C., Carter, R. and Magnus, A. (2011). The economic benefits of reducing physical inactivity: An Australian example. *International journal of behavioral nutrition and physical activity*, *8*(1), p.99.
36. Carlson, S.A., Fulton, J.E., Pratt, M., Yang, Z. and Adams, E.K. (2015). Inadequate physical activity and health care expenditures in the United States. *Progress in cardiovascular diseases*, *57*(4), pp.315-323.
37. Centers for Disease Control and Prevention (CDC). (2018). Physical Activity and Health, Physical Activity, CDC. https://www.cdc.gov/physicalactivity/basics/pa-health/index.htm. Accessed Oct. 17, 2018.
38. Centers for Disease Control and Prevention. (2017). National Diabetes Statistics Report, 2017. Atlanta, GA: Centers for Disease Control and Prevention, U.S. Dept. of Health and Human Services; 2017.
39. Chapleau, R., Gaudette, P., and Spurr, T. (2019). Application of Machine Learning to Two Large-Sample Household Travel Surveys : A Characterization of Travel Modes. TRB, 2019, pp. 1–19.
40. Chastin, S.F., Mandrichenko, O., Helbostadt, J.L. and Skelton, D.A. (2014). Associations between objectively-measured sedentary behaviour and physical activity with bone mineral density in adults and older adults, the NHANES study. *Bone*, *64*, pp.254-262.
41. Chen, T., and Carlos, G. (2016). XGBoost: A Scalable Tree Boosting System. Il Friuli medico, Vol. 19, No. 6, 2016.
42. Cho, W., and E. Choi. (2015). A GPS Trajectory Map-Matching Mechanism with DTG Big Data on the HBase System. Proceedings of the 2015 International Conference on Big Data Applications and Services - BigDAS '15, No. October, 2015, pp. 22–29. https://doi.org/10.1145/2837060.2837062.

43. Christie, N. (2010). EU-O. A Review of Accidents and Injuries to Road Transport Drivers. 2010.
44. Clarke TC, Norris T, Schiller JS. (2017). Early release of selected estimates based on data from 2016 National Health Interview Survey. National Center for Health Statistics. May 2017. Available from: http://www.cdc.gov/nchs/nhis.htm.
45. Dabiri, S., and Heaslip, K. (2017). Inferring Transportation Modes from GPS Trajectories Using a Convolutional Neural Network. Transportation Research Part C, Vol. 86, No. December 2017, 2018, pp. 360–371. https://doi.org/10.1016/j.trc.2017.11.021.
46. Das, R. D., and S. Winter. (2016). Automated Urban Travel Interpretation: A Bottom-up Approach for Trajectory Segmentation. Sensors (Switzerland), Vol. 16, No. 11, 2016. https://doi.org/10.3390/s16111962.
47. Das, R. D., and Winter, S. (2018). A Fuzzy Logic Based Transport Mode Detection Framework in Urban Environment. Journal of Intelligent Transportation Systems, Vol. 22, No. 6, 2018, pp. 478–489. https://doi.org/10.1080/15472450.2018.1436968.
48. Deiotte, R., and R. La Valley. Comparison of Spatiotemporal Mapping Techniques for Enormous Etl and Exploitation Patterns. Vol. Iv, 2017, p. 5194.
49. Dhondt, S., Q. Le, and Hieu, X. (2011). Environmental Health Impacts of Mobility and Transport in Hai Phong , Vietnam. 2011, pp. 363–376. https://doi.org/10.1007/s00477-010-0374-3.
50. Ding, D., Kolbe-Alexander, T., Nguyen, B., Katzmarzyk, P.T., Pratt, M. and Lawson, K.D. (2017). The economic burden of physical inactivity: a systematic review and critical appraisal. *Br J Sports Med*, *51*(19), pp.1392-1409.
51. Ding, D., Lawson, K.D., Kolbe-Alexander, T.L., Finkelstein, E.A., Katzmarzyk, P.T., Van Mechelen, W., Pratt, M. and Lancet Physical Activity Series 2 Executive Committee (2016). The economic burden of physical inactivity: a global analysis of major non-communicable diseases. *The Lancet*, *388*(10051), pp.1311-1324.
52. Domènech, A., and A. Gutiérrez. A GIS-Based Evaluation of the Effectiveness and Spatial Coverage of Public Transport Networks in Tourist Destinations. ISPRS International Journal of Geo-Information, Vol. 6, No. 3, 2017, p. 83. https://doi.org/10.3390/ijgi6030083.
53. Edwards, P., Peggy, A., and Tsouros, (2008). A. A Healthy City Is an Active City: A Physical Activity Planning Guide. World Health Organization, 2008.

54. Efthymiou, A., Barmpounakis, E., Efthymiou, D., and Vlahogianni, E. (2018). Identifying Transportation Mode of Unimodal Trips Using Smartphone Data and Machine Learning Algorithms. 2018, pp. 1–6.
55. Elhenawy, M. (2017). Random Forest / Hidden Markov Transportation Mode Recognition Model Using Smartphone Sensor Data. TRB, Vol. 250, 2017.
56. Endo, Y., H. Toda, K. Nishida, and Kawanobe, A. (2016). Deep Feature Extraction from Trajectories for Transportation Mode Estimation. Advances in Knowledge Discovery and Data Mining, 2016, pp. 54–66.
57. Environments, S. (2017). Sensing Environments. 2017, pp. 1–20. https://doi.org/10.3390/s17061427.
58. Etemad, M., A. S. Junior, and Matwin, S. (2018). Predicting Transportation Modes of GPS Trajectories Using Feature Engineering And. Advances in Artificial Intelligence, 2018, pp. 259–264.
59. Etemad, M., A. Soares, S. Matwin, and Torgo, L. (2019). On Feature Selection and Evaluation of Transportation Mode Prediction Strategies. CEUR-WS, Vol. 2322/BMDA, 2019.
60. Facchinetti, G. Small Satellites: Economic Trends. Vol. 130, No. December, 2016, pp. 1–102.
61. Fan, S., Chen, J., Huang, J., Li, Y., Zhao, L., Liu, X., Li, J., Cao, J., Yu, L., Deng, Y. and Chen, N. (2015). Physical activity level and incident type 2 diabetes among Chinese adults. *Medicine and science in sports and exercise*, *47*(4), pp.751-756.
62. Fang, S. H., H. H. Liao, Y. X. Fei, K. H. Chen, J. W. Huang, Y. D. Lu, and Tsao, Y. (2016). Transportation Modes Classification Using Sensors on Smartphones. Sensors (Switzerland), Vol. 16, No. 8, 2016, pp. 1–15. https://doi.org/10.3390/s16081324.
63. Fang, S., Y. Fei, Z. Xu, and Tsao, Y. (2017). Learning Transportation Modes from Smartphone Sensors Based on Deep Neural Network. IEE, Vol. 1748, No. c, 2017, pp. 1–8. https://doi.org/10.1109/JSEN.2017.2737825.
64. Feng, T., and Timmermans, H. (2016). Comparison of Advanced Imputation Algorithms for Detection of Transportation Mode and Activity Episode Using GPS Data. Transportation Planning and Technology ISSN:, Vol. 1060, 2016. https://doi.org/10.1080/03081060.2015.1127540.
65. Frank, L., Sallis, J., Conway, T., Chapman, J., Saelens, B., and Bachman, W. (2006). Many pathways from land use to health: Associations between neighborhood walkability and active

transportation, body mass index, and air quality. Journal of the American Planning Association, 72 (1) (2006), pp. 75-87.

66. Flint, E. and Cummins, S. (2016). Active commuting and obesity in mid-life: cross-sectional, observational evidence from UK Biobank. *The lancet Diabetes & endocrinology*, *4*(5), pp.420-435.
67. Flint, E., Cummins, S. and Sacker, A. (2014). Associations between active commuting, body fat, and body mass index: population based, cross sectional study in the United Kingdom. *Bmj*, *349*, p. 4887.
68. Frank, LD., Greenwald, MJ., Winkelman, S., Chapman, J., and Kavage, S. (2010). Carbonless footprints: promoting health and climate stabilization through active transportation. Prev Med. Suppl 1: S99-105. doi: 10.1016/j.ypmed.2009.09.025. Epub 2009 Oct 20.
69. Fretts, A.M., Howard, B.V., Kriska, A.M., Smith, N.L., Lumley, T., Lee, E.T., Russell, M. and Siscovick, D. (2009). Physical activity and incident diabetes in American Indians: The Strong Heart Study. *American journal of epidemiology*, *170*(5), pp.632-639.
70. Friedenreich, C.M., McGregor, S.E., Courneya, K.S., Angyalfi, S.J. and Elliott, F.G. (2004). Case-control study of lifetime total physical activity and prostate cancer risk. *American journal of epidemiology*, *159*(8), pp.740-749.
71. Frost, H.M. (2003). Bone's mechanostat: a 2003 update. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology: An Official Publication of the American Association of Anatomists*, *275*(2), pp.1081-1101.
72. Ganti, R., F. Ye, and H. Hei. (2011). Mobile Crowdsensing: Current State and Future Challenges. IEEE Communications Magazine, Vol. 49, No. 11, 2011, pp. 32–39. https://doi.org/10.1109/MCOM.2011.6069707.
73. Garrett, N.A., Brasure, M., Schmitz, K.H., Schultz, M.M. and Huber, M.R. (2004). Physical inactivity: direct cost to a health plan. *American journal of preventive medicine*, *27*(4), pp.304-309.
74. Giovannucci, E.L., Liu, Y., Leitzmann, M.F., Stampfer, M.J. and Willett, W.C. (2005). A prospective study of physical activity and incident and fatal prostate cancer. *Archives of internal medicine*, *165*(9), pp.1005-1010.

75. Gordon-Larsen, P., Boone-Heinonen, J., Sidney, S., Sternfeld, B., Jacobs, D.R. and Lewis, C.E. (2009). Active commuting and cardiovascular disease risk: the CARDIA study. *Archives of internal medicine*, *169*(13), pp.1216-1223.
76. Gotschi, T., and Mills, K. (2008). Active Transportation for America: The Case for Increased Federal Investment in Bicycling and Walking. TRR Journal. Record URL: http://www.railstotrails.org/resources/documents/whatwedo/atfa/ATFA_20081020.pdf
77. Gregov, C. and Šalaj, S. (2014). The Effects of Different training modalities on bone mass: A Review. *Kinesiology: International journal of fundamental and applied kinesiology*, *46*(Supplement 1), pp.10-29.
78. Guvensan, M., B. Dusun, B. Can, and Trkmen, H. (2017). A Novel Segment-Based Approach for Improving Classification Performance of Transport Mode Detection. Sensors, Vol. 18, No. 2, 2017, p. 87. https://doi.org/10.3390/s18010087.
79. Hagenauer, J., and Helbich, M. (2017). A Comparative Study of Machine Learning Classifiers for Modeling Travel Mode Choice. Expert Systems With Applications, Vol. 78, 2017, pp. 273–282. https://doi.org/10.1016/j.eswa.2017.01.057.
80. Haines, A., and Dora, C. (2012). How the Low Carbon Economy Can Improve Health? BMJ (Clinical research ed.), Vol. 344, No. March, 2012, pp. 1–6. https://doi.org/10.1136/bmj.e1018.
81. Hales CM, Carroll MD, Fryar CD, Ogden CL. (2017). Prevalence of obesity among adults and youth: United States, 2015–2016. NCHS data brief, no 288. Hyattsville, MD: National Center for Health Statistics. 2017.
82. Hallal, P. C., L. B. Andersen, F. C. Bull, R. Guthold, W. Haskell, U. Ekelund, J. R. Alkandari, et. al. and Wells, J. (2012). Global Physical Activity Levels: Surveillance Progress, Pitfalls, and Prospects. The Lancet, Vol. 380, No. 9838, 2012, pp. 247–257. https://doi.org/10.1016/S0140-6736(12)60646-1.
83. Hamer, M. and Chida, Y. (2008). Active commuting and cardiovascular risk: a meta-analytic review. *Preventive medicine*, *46*(1), pp.9-13.
84. Hansen, D., Bazell, C., Pelizzari, P., and Pyenson, B. (2019). Medicare Cost of Osteoporotic Fractures, The Clinical and Cost Burden of an Important Consequence of Osteoporosis., National Osteoporosis Foundation (NOF).

85. Hatano, Y. (1993). Use of the pedometer for promoting daily walking exercise. *ICHPER*, *29*, pp.4-8.
86. He, S. (2017). Spatial Query Processing for Location Based Application on Hbase. 2017, pp. 110–114.
87. Health and Human Services (HHS). (2018). Physical Activity Guidelines for Americans. Accessed from https://www.hhs.gov/fitness/be-active/physical-activity-guidelines-for-mericans/index.html.
88. Hox, J. J., and T. M. Bechger. Introduction to Structural Equation Modeling. Applied Quantitative Analysis in Education and the Social Sciences, 2013, pp. 245–264. https://doi.org/10.4324/9780203108550.
89. Hu, G., Qiao, Q., Silventoinen, K., Eriksson, J.G., Jousilahti, P., Lindström, J., Valle, T.T., Nissinen, A. and Tuomilehto, J. (2003). Occupational, commuting, and leisure-time physical activity in relation to risk for Type 2 diabetes in middle-aged Finnish men and women. *Diabetologia*, *46*(3), pp.322-329.
90. Hu, G., Sarti, C., Jousilahti, P., Silventoinen, K., Barengo, N.C. and Tuomilehto, J. (2005). Leisure time, occupational, and commuting physical activity and the risk of stroke. *Stroke*, *36*(9), pp.1994-1999.
91. Hu, G., Tuomilehto, J., Borodulin, K. and Jousilahti, P. (2007). The joint associations of occupational, commuting, and leisure-time physical activity, and the Framingham risk score on the 10-year risk of coronary heart disease. *European Heart Journal*, *28*(4), pp.492-498.
92. Humphreys, D.K., Goodman, A. and Ogilvie, D. (2013). Associations between active commuting and physical and mental wellbeing. *Preventive medicine*, *57*(2), pp.135-139.
93. Jahangiri, A., and H. Rakha, H. (2015). Applying Machine Learning Techniques to Transportation Mode Recognition Using Mobile Phone Sensor Data. IEEE Transactions on Intelligent Transportation Systems, Vol. 16, No. 5, 2015, pp. 2406–2417. https://doi.org/10.1109/TITS.2015.2405759.
94. Ji, Y., Gao, L., Chen, D., Zhou, Y., and Zhang, Y. (2017). Functional Analysis of Public Transport Network in Trip Mode Detection from Personal Smartphone Trajectory Data. TRB, 2017, pp. 1–17.
95. Jihye Jeon. The Strengths and Limitations of the Statistical Modeling of Complex Social Phenomenon: Focusing on SEM, Path Analysis, or Multiple Regression Models. International

Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, Vol. 9, No. 5, 2015, pp. 1634–1642.

96. Jonker, J.T., De Laet, C., Franco, O.H., Peeters, A., Mackenbach, J. and Nusselder, W.J. (2006). Physical activity and life expectancy with and without diabetes: life table analysis of the Framingham Heart Study. *Diabetes Care*, *29*(1), pp.38-43.
97. Kamruzzaman, M., J. Hine, B. Gunay, and Blair, N. (2011). Using GIS to Visualise and Evaluate Student Travel Behaviour. Journal of Transport Geography, Vol. 19, No. 1, 2011, pp. 13–32. https://doi.org/10.1016/j.jtrangeo.2009.09.004.
98. Katzmarzyk, P.T. and Janssen, I. (2004). The economic costs associated with physical inactivity and obesity in Canada: an update. *Canadian journal of applied physiology*, *29*(1), pp.90-115.
99. Katzmarzyk, P.T., Gledhill, N. and Shephard, R.J. (2000). The economic burden of physical inactivity in Canada. *Cmaj*, *163*(11), pp.1435-1440.
100. Kirch, W. (2008). *Encyclopedia of Public Health: Volume 1: A-H Volume 2: I-Z*. Springer Science & Business Media.
101. KMetro. (2019). Accessed from https://www.kmetro.com, Accessed on 08/25/2019
102. Koh-Banerjee, P., Chu, N.F., Spiegelman, D., Rosner, B., Colditz, G., Willett, W. and Rimm, E. (2003). Prospective study of the association of changes in dietary intake, physical activity, alcohol consumption, and smoking with 9-y gain in waist circumference among 16 587 US men. *The American journal of clinical nutrition*, *78*(4), pp.719-727.
103. Koopmanschap, M.A. and Rutten, F.F. (1996). A practical guide for calculating indirect costs of disease. *Pharmacoeconomics*, *10*(5), pp.460-466.
104. Koopmanschap, M.A., Rutten, F.F., van Ineveld, B.M. and Van Roijen, L. (1995). The friction cost method for measuring indirect costs of disease. *Journal of health economics*, *14*(2), pp.171-189.
105. Kowalski, K., R. Rhodes, P. J. Naylor, H. Tuokko, and MacDonald, S. (2012). Direct and Indirect Measurement of Physical Activity in Older Adults: A Systematic Review of the Literature. International Journal of Behavioral Nutrition and Physical Activity, Vol. 9, 2012. https://doi.org/10.1186/1479-5868-9-148.

106. Kriska, A.M., Saremi, A., Hanson, R.L., Bennett, P.H., Kobes, S., Williams, D.E. and Knowler, W.C. (2003). Physical activity, obesity, and the incidence of type 2 diabetes in a high-risk population. *American journal of epidemiology*, *158*(7), pp.669-675.

107. Krueger, H., Krueger, J. and Koot, J. (2015). Variation across Canada in the economic burden attributable to excess weight, tobacco smoking and physical inactivity. *Canadian Journal of Public Health*, *106*(4), pp.e171-e177.

108. Kruk, J. (2014). MINI-REVIEW Physical Activity and Health. No. January, 2014.

109. Kruk, J. and Czerniak, U. (2013). Physical activity and its relation to cancer risk: updating the evidence. *Asian Pac J Cancer Prev*, *14*(7), pp.3993-4003.

110. Kyu, H.H., Bachman, V.F., Alexander, L.T., Mumford, J.E., Afshin, A., Estep, K., Veerman, J.L., Delwiche, K., Iannarone, M.L., Moyer, M.L. and Cercy, K. (2016). Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. *bmj*, *354*, p.i3857.

111. Langerudi, M.F., Abolfazl, M. and Sriraj, P.S. (2015). Health and transportation: small scale area association. *Journal of Transport & Health*, *2*(2), pp.127-134.

112. Lari, Z. A., and Golroo. A. (2015). Automated Transportation Mode Detection Using Smart Phone Applications via Machine Learning: Case Study Mega City of Tehran. Transportation Research Board 94th Annual Meeting, Vol. 6147, 2015.

113. Leskovec, J., A. Rajaraman, and J. D. Ullman. (2014). Mining of Massive Datasets. Mining of Massive Datasets: Second Edition, 2014, pp. 1–458. https://doi.org/10.1017/CBO9781139924801.

114. Litman, T. (2003). Economic value of walkability. Paper presented at the annual meeting of Transportation Research Board, Washington, DC.

115. Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A., Hugo, G. (2007) Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes. Health Place 13(1):111–122. DOI: 10.1016/j.healthplace.2005. 11.00.

116. Laverty, A. A., E. Webb, E. P. Vamos, and Millett, C. (2018). Associations of Increases in Public Transport Use with Physical Activity and Adiposity in Older Adults. 2018, pp. 1–10.

117. Lee, D., R. Moussalli, S. Asaad, and Srivatsa, M. (2016). Spatial Predicates Evaluation in the Geohash Domain Using Reconfigurable Hardware. Proceedings - 24th IEEE International Symposium on Field-Programmable Custom Computing Machines, FCCM 2016, 2016, pp. 176–183. https://doi.org/10.1109/FCCM.2016.51.
118. Lee, I.M., Shiroma, E.J., Lobelo, F., Puska, P., Blair, S.N., Katzmarzyk, P.T. and Lancet Physical Activity Series Working Group (2012). Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The lancet*, *380*(9838), pp.219-229.
119. Lee, I.M., Shiroma, E.J., Lobelo, F., Puska, P., Blair, S.N., Katzmarzyk, P.T. and Lancet Physical Activity Series Working Group. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy.
120. Lee, K.J., Inoue, M., Otani, T., Iwasaki, M., Sasazuki, S., Tsugane, S. and JPHC Study Group (2007). Physical activity and risk of colorectal cancer in Japanese men and women: the Japan Public Health Center-based prospective study. *Cancer Causes & Control*, *18*(2), pp.199-209.
121. Le-Khac NA., Bue M., Whelan M., Kechadi MT. (2010) A Clustering-Based Data Reduction for Very Large Spatio-Temporal Datasets. In: Cao L., Zhong J., Feng Y. (eds) Advanced Data Mining and Applications. ADMA 2010. Lecture Notes in Computer Science, vol 6441. Springer, Berlin, Heidelberg
122. Li, F., K. J. Fisher, A. Bauman, M. G. Ory, W. Chodzko-zajko, and Harmer, P. (2006). Neighborhood Influences on Physical Activity in Middle-Aged and Older Adults: A Multilevel Perspective. 2006, pp. 87–114.
123. Li, J. and Siegrist, J. (2012). Physical activity and risk of cardiovascular disease—a meta-analysis of prospective cohort studies. *International journal of environmental research and public health*, *9*(2), pp.391-407.
124. Liang, X., and Wang, G. (2017). A Convolutional Neural Network for Transportation Mode Detection Based on Smartphone Platform. 2017. https://doi.org/10.1109/MASS.2017.81.
125. Linkages, E., and Heli, I. (2009). Health and Environment Linkages Policy Series Healthy Transport in Developing Cities. World, 2009.
126. Littman, A.J., Kristal, A.R. and White, E. (2006). Recreational physical activity and prostate cancer risk (United States). *Cancer Causes & Control*, *17*(6), pp.831-841.

127. Liu, Y., and Li, Z. (2017). A Novel Algorithm of Low Sampling Rate GPS Trajectories on Map-Matching. Eurasip Journal on Wireless Communications and Networking, Vol. 2017, No. 1, 2017. https://doi.org/10.1186/s13638-017-0814-6.

128. Loidl, M., G. Wallentin, R. Cyganski, A. Graser, J. Scholz, and Haslauer, E. (2016). GIS and Transport Modeling—Strengthening the Spatial Perspective. ISPRS International Journal of Geo-Information, Vol. 5, No. 6, 2016, p. 84. https://doi.org/10.3390/ijgi5060084.

129. Lynch, B.M., Neilson, H.K. and Friedenreich, C.M. (2010). Physical activity and breast cancer prevention. In *Physical activity and cancer* (pp. 13-42). Springer, Berlin, Heidelberg.

130. Machek, E., Frazier, J., Ingles, A., and Morton, T. (2016). Wearable Sensors in Transportation - Exploratory Advanced Research Program Initial Stage Investigation. Final report. Accessed from https://www.fhwa.dot.gov/publications/research/ear/16034/index.cfm.

131. Maizlish, N., Woodcock, J., Co, S., Ostro, B., Fanai. A., and Fairley, D. (2013). Health cobenefits and transportation-related reductions in greenhouse gas emissions in the San Francisco Bay area. Am J Public Health. 2013 Apr;103(4):703-9. doi: 10.2105/AJPH.2012.300939. Epub 2013 Feb 14.

132. Maresova, K. (2014). The costs of physical inactivity in the Czech Republic in 2008. *Journal of Physical Activity and Health*, *11*(3), pp.489-494.

133. Marillac, A. (1999). Integrating GPS Data within Embedded Internet GIS. Access, 1999, pp. 134–139. https://doi.org/10.1145/320134.320168.

134. Marques, E.A., Mota, J. and Carvalho, J. (2012). Exercise effects on bone mineral density in older adults: a meta-analysis of randomized controlled trials. *Age*, *34*(6), pp.1493-1515.

135. Martin, B. D., V. Addona, J. Wolfson, G. Adomavicius, and Fan, Y. (2017). Methods for Real-Time Prediction of the Mode of Travel Using Smartphone-Based GPS and Accelerometer Data. Sensors (Switzerland), Vol. 17, No. 9, 2017, pp. 1–20. https://doi.org/10.3390/s17092058.

136. Miles, L. (2007). Physical activity and health. *Nutrition bulletin*, *32*(4), pp.314-363.

137. Milne, D., and Watling, D. (2019). Big Data and Understanding Change in the Context of Planning Transport Systems. Journal of Transport Geography, Vol. 76, No. October 2017, 2019, pp. 235–244. https://doi.org/10.1016/j.jtrangeo.2017.11.004.

138. Min, J.Y. and Min, K.B. (2016). Excess medical care costs associated with physical inactivity among korean adults: retrospective cohort study. *International journal of environmental research and public health*, *13*(1), p.136.

139. Mondiale, O. (2009). Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks. *Genève, Suisse: Publications de l'Organisation mondiale de la Santé*.
140. Moore, M. (2016). Tech Giants and Civic Power. Centre for the Study of Media, Communication and Power." King's College London, No. April, 2016, pp. 1–85.
141. Morris, J.N., Heady, J.A., Raffle, P.A.B., Roberts, C.G. and Parks, J.W. (1953). Coronary heart-disease and physical activity of work. *The Lancet, 262*(6796), pp.1111-1120.
142. Murphy, M.H., Nevill, A.M., Murtagh, E.M. and Holder, R.L. (2007). The effect of walking on fitness, fatness and resting blood pressure: a meta-analysis of randomised, controlled trials. *Preventive medicine*, *44*(5), pp.377-385.
143. Nakanishi, M., S. Izumi, S. Nagayoshi, H. Kawaguchi, M. Yoshimoto, T. Shiga, T. Ando, S. Nakae, C. Usui, T. Aoyama, and S. Tanaka. Estimating Metabolic Equivalents for Activities in Daily Life Using Acceleration and Heart Rate in Wearable Devices. BioMedical Engineering Online, Vol. 17, No. 1, 2018, pp. 1–18. https://doi.org/10.1186/s12938-018-0532-2.
144. Nam, D., Kim, H., Cho, J., and Jayakrishnan. R. (2017). A Model Based on Deep Learning for Predicting Travel Mode Choice. TRB, 2017.
145. National Cancer Institute. (2018). Cancer Statistics, https://www.cancer.gov/about-cancer/understanding/statistics. Accessed on 09/17/2019.
146. National Health Interview Survey (NHIS). (2017). Leisure-time Physical Activity Recodes. Accessed from https://www.cdc.gov/nchs/nhis/physical_activity/pa_recodes.htm.
147. National Institutes of Health Osteoporosis and Related Bone Diseases National Resource Center. (2018). Osteoporosis Overview. https://www.bones.nih.gov/sites/bones/files/pdfs/osteopoverview-508.pdf. Published 2018. Accessed September 20, 2019.
148. National Osteoporosis Foundation. (2019). "What is Osteoporosis and What Causes it?", Accessed https://www.nof.org/patients/what-is-osteoporosis/. September 20, 2019.
149. Ndahimana, D., and Kim, E. (2017). Measurement Methods for Physical Activity and Energy Expenditure: A Review. Clinical nutrition research, Vol. 6, No. 2, 2017, pp. 68–80. https://doi.org/10.7762/cnr.2017.6.2.68.
150. Netz, Y., Zeev, A., Arnon, M. and Tenenbaum, G. (2008). Reasons attributed to omitting exercising: A population-based study. *International Journal of Sport and Exercise Psychology*, *6*(1), pp.9-23.
151. Nilsen, T.I., Romundstad, P.R. and Vatten, L.J. (2006). Recreational physical activity and risk of prostate cancer: A prospective population-based study in Norway (the HUNT study). *International journal of cancer*, *119*(12), pp.2943-2947.

152. Nin, J., D. Carrera, and Villatoro, D. (2014). On the Use of Social Trajectory-Based Clustering Methods for Public Transport Optimization. 2014, pp. 59–70. https://doi.org/10.1007/978-3-319-04178-0_6.
153. North Central Texas Council of Governments (NCTCOG). (2019). Economic Benefits of Active Transportation. Accessed from https://www.nctcog.org/trans/plan/bikeped/resources/ebat.
154. Oh, k., Kim, K., Kim, A., Lee, Y., and Jung, J. (2017). Spatial Movement Pattern Analysis in Public Transportation Networks in Seoul. Accessed from https://agile-online.org/conference_paper/cds/agile_2016/posters/154_Paper_in_PDF.pdf
155. Oja, P., and Titze, S. (2011). Physical Activity Recommendations for Public Health: Development and Policy Context. 2011, pp. 253–259. https://doi.org/10.1007/s13167-011-0090-1.
156. Onat, A., Hergenç, G., Küçükdurmaz, Z., Bulur, S., Kaya, Z. and Can, G. (2007). Prospective evidence for physical activity protecting Turkish adults from metabolic disorders. *Archives of the Turkish Society of Cardiology*, *35*(8), pp.467-474.
157. Patel, A.V., Rodriguez, C., Jacobs, E.J., Solomon, L., Thun, M.J. and Calle, E.E. (2005). Recreational physical activity and risk of prostate cancer in a large cohort of US men. *Cancer Epidemiology and Prevention Biomarkers*, *14*(1), pp.275-279.
158. Patire, A. D., M. Wright, B. Prodhomme, and Bayen, A. (2015). How Much GPS Data Do We Need? Transportation Research Part C: Emerging Technologies, Vol. 58, 2015, pp. 325–342. https://doi.org/10.1016/j.trc.2015.02.011.
159. Sitlington, J. (1999). The Relationship Between Transport and Health. Moving to Healthier People and Healthier Places. November, 1999, Accessed from https://www.vichealth.vic.gov.au/~/media/ProgramsandProjects/PlanningHealthyEnvironments/Attachments/vhtransch2.ashx.
160. Parida, P. K. (2012). Artificial Neural Network Based Numerical Solution of Ordinary Differential Equations.
161. Pojani, D., and Stead, D. (2015). Sustainable Urban Transport in the Developing World: Beyond Megacities. No. 2, 2015, pp. 7784–7805. https://doi.org/10.3390/su7067784.
162. Pratt, M., Macera, C.A. and Wang, G. (2000). Higher direct medical costs associated with physical inactivity. *The Physician and sportsmedicine*, *28*(10), pp.63-70.
163. Prelipcean, A., Gidófalvi, G., Susilo, Y. (2016). Measures of transport mode segmentation of trajectories. International Journal of Geographical Information Science. 30. 1-22. 10.1080/13658816.2015.1137297.

164. Procyk, A., L. Frank, J. van Loon, A. Frank, G. Poulos, and Bohle, S. (2013). Transportation & Health: Context Report. No. February, 2013, p. 75.

165. Pronk, N.P., Goodman, M.J., O'Connor, P.J. and Martinson, B.C. (1999). Relationship between modifiable health risks and short-term health care charges. *Jama*, *282*(23), pp.2235-2239.

166. Qiao, Q. and Nyamdorj, R. (2010). Is the association of type II diabetes with waist circumference or waist-to-hip ratio stronger than that with body mass index?. *European journal of clinical nutrition*, *64*(1), p.30.

167. Qin, Y., H. Luo, F. Zhao, Z. Zhao, and Jiang, M. (2018).. A Traffic Pattern Detection Algorithm Based on Multimodal Sensing. International Journal ofDistributed Sensor Networks, Vol. 14, No. 6, 2018. https://doi.org/10.1177/1550147718807832.

168. Quintella, C., L. Andrade, and Campos, C. (2016). Detecting the Transportation Mode for Context-Aware Systems Using Smartphones. 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), 2016, pp. 2261–2266. https://doi.org/10.1109/ITSC.2016.7795921.

169. Quiterio, A.L.D., Carnero, E.A., Baptista, F.M. and Sardinha, L.B. (2011). Skeletal mass in adolescent male athletes and nonathletes: relationships with high-impact sports. *The Journal of Strength & Conditioning Research*, *25*(12), pp.3439-3447.

170. Reis, J.P., Loria, C.M., Sorlie, P.D., Park, Y., Hollenbeck, A. and Schatzkin, A. (2011). Lifestyle factors and risk for new-onset diabetes in a large population-based prospective cohort study. *Annals of internal medicine*, *155*(5), p.292.

171. Report, D. (2016). Links between Noise and Air Pollution and Socioeconomic Status. 2016.

172. Rezaie, M., Patterson, Z., Yu, J., and Yazdizadeh, A. (2018). Travel Mode Detection from Smartphone Data: Semi-Supervised vs. Supervised Learning. 2018.

173. Rezaie, M., Z. Patterson, J. YU, and Yazdizadeh, A. (2017). Semi-Supervised Travel Mode Detection from Smartphone Data. 2017.

174. Rojas, M. B., E. Sadeghvaziri, and Jin, X. (2017). Comprehensive Review of Travel Behavior and Mobility Pattern Studies That Used Mobile Phone Data. Transportation Research Record: Journal of the file:///C:/Users/engra/Dropbox/Paper of Geohash/Reference/Papers/General/Satellite_navigation_technology_secure.pdfTransportation Research Board, Vol. 2563, No. October 2017, 2016, pp. 71–79. https://doi.org/10.3141/2563-11.

175. Romero, J., Fukuda, A., Morisugi, H., & Zusman, E. (2011). Mainstreaming Transport Co-Benefits Approach. Transportation Research Board 90th Annual Meeting, Location: Washington DC, United States, Date: 2011-1-23 to 2011-1-27.

176. Rose, W. (2009). Satellite Navigation Technology Applications for Intelligent Transport Systems: A European Perspective. Proceedings of The Institution of Civil Engineers, Vol. 162, 2009, pp. 75–82. https://doi.org/10.1111/j.1741-3729.2004.00008.x.
177. Samimi, A., Mohammadian, A.K. and Madanizadeh, S. (2009). Effects of transportation and built environment on general health and obesity. *Transportation Research Part D: transport and environment, 14*(1), pp.67-71.
178. Sato, K.K., Hayashi, T., Kambe, H., Nakamura, Y., Harita, N., Endo, G. and Yoneda, T. (2007). Walking to work is an independent predictor of incidence of type 2 diabetes in Japanese men: The Kansai Healthcare Study. *Diabetes Care*, *30*(9), pp.2296-2298.
179. Savalei, V., and Bentler, P. (2006). STRUCTURALSTRUCTURAL EQUATION MODELING. Accessed from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.608.573&rep=rep1&type=pdf.
180. Scarborough, P., Bhatnagar, P., Wickramasinghe, K.K., Allender, S., Foster, C. and Rayner, M. (2011). The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006–07 NHS costs. *Journal of public health*, *33*(4), pp.527-535.
181. Scheepers, C.E., Wendel-Vos, G.C.W., van Wesemael, P.J.V., den Hertog, F.R.J., Stipdonk, H.L., Panis, L.I., van Kempen, E.E.M.M. and Schuit, A.J. (2015). Perceived health status associated with transport choice for short distance trips. *Preventive medicine reports*, *2*, pp.839-844.
182. Schmid, D., Steindorf, K., Leitzmann, M.F. (2014). Epidemiologic studies of physical activity and primary prevention of cancer. *Dtsch Z Sportmed*, *65*, pp. 5-10.
183. Schuessler, N. & Axhausen, K. W. (2009). Processing GPS raw data without additional Information. Proceedings of the 88th Annual Meeting of the Transportation Research Board, January 2009, Washington D.C.
184. Scott, A., Khan, K.M., Duronio, V. and Hart, D.A. (2008). Mechanotransduction in human bone. *Sports Medicine*, *38*(2), pp.139-160.
185. Sdaulton, A. (2018). NYC Taxi Data Prediction. NYC Taxi Data Prediction maintained by sdaulton, published with GitHub Pages. https://sdaulton.github.io/TaxiPrediction/. Accessed Nov. 4, 2018.
186. Shafique, M., and Hato, E. (2015). A Comparison among Various Classification Algorithms for Travel Mode Detection Using Sensors' Data Collected by Smartphones. 2015, pp. 1–17.
187. Shafique, M., and Hato, E. (2016). Travel Mode Detection with Varying Smartphone Data Collection Frequencies. Sensors, MDPI, 2016. https://doi.org/10.3390/s16050716.

188. She, Z., King, D.M. and Jacobson, S.H. (2019). Is promoting public transit an effective intervention for obesity? A longitudinal study of the relation between public transit usage and obesity. *Transportation Research Part A: Policy and Practice*, *119*, pp.162-169.
189. Sheridan, R. P., W. M. Wang, A. Liaw, J. Ma, and E. M. Gifford. (2016). Extreme Gradient Boosting as a Method for Quantitative Structure-Activity Relationships. Journal of Chemical Information and Modeling. 12. Volume 56, 2353–2360.
190. Shi, Y., Li, T., Wang, Y., Zhou, L., Qin, Q., Yin, J., Wei, S., Liu, L. and Nie, S. (2015). Household physical activity and cancer risk: a systematic review and dose-response meta-analysis of epidemiological studies. *Scientific reports*, *5*, p.14901.
191. Shmerling, R. (2016). How Useful is the Body Mass Index (BMI)? Obtained from https://www.health.harvard.edu/blog/how-useful-is-the-body-mass-index-bmi-201603309339, accessed on 5/7/2019.
192. Sifferlin, A. (2013). Why BMI Isn't the Best Measure for Weight (or Health), obtained from http://healthland.time.com/2013/08/26/why-bmi-isnt-the-best-measure-for-weight-or-health/, accessed on 5/7/2019.
193. Singh, P., Oh, H. and Jung, J. (2017). Flow Orientation Analysis for Major Activity Regions Based on Smart Card Transit Data. ISPRS Int. J. Geo-Inf. 2017, 6(10), 318; https://doi.org/10.3390/ijgi6100318.
194. Sirven, J. I., and Varrato, J. (1999). Physical Activity and Health, A Report of the Surgeon General. The Physician and Sportsmedicine, Vol. 27, No. 3, 1999, pp. 63–70. https://doi.org/10.3810/psm.1999.03.723.
195. Smith, J., E. Clayton, and Hanson, D. (2017). Building Sustainable, Inclusive Transportation Systems: A Framework for the Future. Price Waterhouse Coopers, 2017.
196. Song, X., H. Kanasugi, and Shibasaki, R. (2013). DeepTransport : Prediction and Simulation of Human Mobility and Transportation Mode at a Citywide Level. 2013, pp. 2618–2624.
197. Stenneth, L., O. Wolfson, P. S. Yu, B. Xu, and Morgan, S. (2011). Transportation Mode Detection Using Mobile Phones and GIS Information. 2011.
198. Stephenson, J., Bauman, A., Armstrong, T., Smith, B., Bellew, B. (2000). The cost of illness attributable to physical inactivity in Australia. Canberra, ACT, Australia: The Commonwealth Department of Health and Aged Care and the Australian Sports Commission.
199. Stopher, P. R, Jiang, Q. & FitzGerald, C. (2005). Processing GPS data from travel surveys. Proceedings of 2nd Int. Colloquium on the Behavioral Foundations of Integrated Land-use and Transportation Models: Frameworks, Models and Applications. June 2005. Toronto, Canada.

200. Stopher, P., Bullock, P. & Jiang, Q. (2002). GPS, GIS and personal travel surveys: an exercise in visualisation, 25th Australasian Transport Research Forum Incorporating the BTRE Transport Policy Colloquium, October 2002, Canberra.

201. Stopher, P., Clifford, E., Zhang, J. & FitzGerald, C. (2008). Deducing mode and purpose from GPS data. Working paper ITLS-WP-08-0st6it.uIten of Transport and Logistic Studies, the Australian Key Center in Transport and Logistic Management, the Univiersity of Sydney.

202. Suwardi, I. S., D. Dharma, D. P. Satya, and Lestari, D. (2015). Geohash Index Based Spatial Data Model for Corporate. 2015, pp. 478–483.

203. Sylvia, L. G. (2015). A Practical Guide to Measuring Physical Activity. Growth (Lakeland), Vol. 23, No. 1, 2015, pp. 199–208.

204. Tajalli, M. and Hajbabaie, A. (2017). On the relationships between commuting mode choice and public health. *Journal of Transport & Health*, *4*, pp.267-277.

205. Tambi, R., and Li, P. (2018). An Efficient CNN Model for Transportation Mode Sensing. pp. 315–316. https://doi.org/10.3390/s16081324.

206. Tang, L., C. Xiong, and Zhang, L. (2015). Decision Tree Method for Modeling Travel Mode Switching in a Dynamic Behavioral Process. Transportation Planning and Technology ISSN:, Vol. 1060, 2015. https://doi.org/10.1080/03081060.2015.1079385.

207. Tinker A. How to Improve Patient Outcomes for Chronic Diseases and Comorbidities. [(accessed on 30 September 2019)]; Available online: http://www.healthcatalyst.com/wp-content/uploads/2014/04/How-to-Improve-Patient-Outcomes.pdf.

208. Tonstad, S., Stewart, K., Oda, K., Batech, M., Herring, R.P. and Fraser, G.E. (2013). Vegetarian diets and incidence of diabetes in the Adventist Health Study-2. *Nutrition, Metabolism and Cardiovascular Diseases*, *23*(4), pp.292-299.

209. Transportation, City of Arlington, Texas. (2019).
[https://www.arlingtontx.gov/residents/city_services/transportation](https://www.arlingtontx.gov/residents/city_services/transportation).
Accessed on 08/25/2019

210. Tsui, S. Y. A, Shalaby, A. S. (2006). An Enhanced System for Link and Mode Identification for GPS-based Personal Travel Surveys. Proceedings of the 85th Annual Meeting of the Transportation Research Board, January 2006, Washington D.C.

211. Tudor-Locke, C. and Bassett, D.R. (2004). How many steps/day are enough? *Sports medicine*, *34*(1), pp.1-8.

212. Turner, S. M., W. L. Eisele, R. J. Benz, and J. Douglas, J. (1998). Travel Time Data Collection Handbook. Federal Highway Administration, USA., Vol. 3, No. 5, 1998, p. 293.

213. U.S. Department of Health and Human Services. (2018). *Physical Activity Guidelines for Americans, 2nd edition*. Washington, DC: U.S. Department of Health and Human Services; 2018.
214. United States Census Bureau QuickFacts, Arlington city, Texas. (2019). https://www.census.gov/quickfacts/fact/table/arlingtoncitytexas/PST045218. Accessed 08/20/019.
215. United States Census Bureau QuickFacts, Kalamazoo city, Michigan. (2019). https://www.census.gov/quickfacts/fact/table/kalamazoocitymichigan/PST045218. Accessed 08/20/019.
216. US Department of Health and Human Services. (1996). Physical activity and health: a report of the Surgeon General. Atlanta (GA): US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Promotion, 1996.
217. Varo, J. J., M. A. Martínez-gonzález, J. De Irala-estévez, J. Kearney, M. Gibney, and Martínez, J. (2003). Distribution and Determinants of Sedentary Lifestyles in the European Union. No. February, 2003. https://doi.org/10.1093/ije/dyg116.
218. Vazquez, G., Duval, S., Jacobs Jr, D.R. and Silventoinen, K. (2007). Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis. *Epidemiologic reviews*, *29*(1), pp.115-128.
219. Via, C. E., P. Hall, V. Tech, U. States, and Author, C. (2018). Transport-Domain Applications of Widely Used Data Sources in the Smart Transportation: A Survey Sina Dabiri. pp. 1–52. 2018.
220. Vicente-Rodríguez, G. (2006). How does exercise affect bone development during growth? *Sports Medicine*, *36*(7), pp.561-569.
221. Viscusi, W.K. and Aldy, J.E. (2003). The value of a statistical life: a critical review of market estimates throughout the world. *Journal of risk and uncertainty*, *27*(1), pp.5-76.
222. Voskuil, D.W., Monninkhof, E.M., Elias, S.G., Vlems, F.A. and van Leeuwen, F.E. (2007). Physical activity and endometrial cancer risk, a systematic review of current evidence. *Cancer Epidemiology and Prevention Biomarkers*, *16*(4), pp.639-648.
223. Wagner, A., Simon, C., Evans, A., Ferrières, J., Montaye, M., Ducimetière, P. and Arveiler, D. (2002). Physical activity and coronary event incidence in Northern Ireland and France: The Prospective Epidemiological Study of Myocardial Infarction (PRIME). *Circulation*, *105*(19), pp.2247-2252.
224. Waki, K., Noda, M., Sasaki, S., Matsumura, Y., Takahashi, Y., Isogawa, A., Ohashi, Y., Kadowaki, T., Tsugane, S. and JPHC Study Group (2005). Alcohol consumption and other risk factors for self-reported diabetes among middle-aged Japanese: a population-based prospective study in the JPHC study cohort I. *Diabetic medicine*, *22*(3), pp.323-331.

225. Waller, K., Kaprio, J., Lehtovirta, M., Silventoinen, K., Koskenvuo, M. and Kujala, U.M. (2010). Leisure-time physical activity and type 2 diabetes during a 28-year follow-up in twins. *Diabetologia*, *53*(12), pp.2531-2537.
226. Wang, B., L. Gao, and Juan, Z. (2018). Travel Mode Detection Using GPS Data and Socioeconomic Attributes Based on a Random Forest Classifier. IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, Vol. 19, No. 5, 2018, pp. 1547–1558.
227. Wang, B., Y. Wang, K. Qin, and Xiao, Q. (2018). Detecting Transportation Modes Based on LightGBM Classifier from GPS Trajectory Data. 2018 26th International Conference on Geoinformatics, No. 41471326, pp. 1–7.
228. Wang, F., Street, S., Ross, C., and Ph, D. (2017). Predicting Travel Mode Choices in the Delaware Valley Region with Multinomial Logit Model and Extreme Gradient Boost Model. 2017.
229. Wang, H., H. Luo, F. Zhao, Y. Qin, Z. Z., and Chen. Y. (2018). Detecting Transportation Modes with Low-Power- Consumption Sensors Using Recurrent Neural Network. 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2018, pp. 1098–1105. https://doi.org/10.1109/SmartWorld.2018.00191.
230. Wang, L. I. N., H. Gjoreski, M. Ciliberto, S. MEKKI, S. Valentin, and Roggen, D. (2019). Enabling Reproducible Research in Sensor-Based Transportation Mode Recognition with the Sussex-Huawei Dataset. IEEE Access, Vol. 7, 2019, pp. 10870–10891. https://doi.org/10.1109/ACCESS.2019.2890793.
231. Watkins, S. J. (2014). The Transport and Health Study Group. Journal of Transport and Health, Vol. 1, No. 1, 2014, pp. 3–4. https://doi.org/10.1016/j.jth.2013.09.003.
232. Wennberg, P., Lindahl, B., Hallmans, G., Messner, T., Weinehall, L., Johansson, L., Boman, K. and Jansson, J.H. (2006). The effects of commuting activity and occupational and leisure time physical activity on risk of myocardial infarction. *European Journal of Cardiovascular Prevention & Rehabilitation*, *13*(6), pp.924-930.
233. WHO. (2010). Physical Inactivity a Leading Cause of Disease and Disability, Warns WHO. WHO, 2010.
234. Widener, M. J., and Hatzopoulou, M. (2016). Contextualizing Research on Transportation and Health: A Systems Perspective. Journal of Transport and Health, Vol. 3, No. 3, 2016, pp. 232–239. https://doi.org/10.1016/j.jth.2016.01.008.

235. Wiehe, S., Carroll, A., Liu, g., haberkorn, K., Hoch, S., Wilson, J. and Fortenberry, J. (2008). Using GPS-enabled cell phones to track the travel patterns of adolescents. International Journal of Health Geographics2008, https://doi.org/10.1186/1476-072X-7-22
236. Wikipedia contributors. (2018). Geohash. In Wikipedia, The Free Encyclopedia. Retrieved November 2, 2018, from https://en.wikipedia.org/w/index.php?title=Geohash&oldid=864446668
237. WILL'S NOISE. (2016). Pygeohash 1.0.1: Fast GIS with Geohash and Python - Will's Noise. WILL'S NOISE. http://www.willmcginnis.com/2016/01/16/pygeohash-1-0-1-fast-gis-geohash-python/. Accessed Nov. 8, 2018.
238. Wolf, J., Guensler, R., &Bachman, W. (2001). Elimination of the Travel Diary: An Experiment to Derive Trip Purpose from GPS Travel Data. Proceedings of the 80th Annual Meeting of the Transportation Research Board, January 2001, Washington D.C.
239. World Health Organization. (2006). Health Effects and Risks of Transport Systems: The HEARTS Project. Copenhagen: WHO Regional Office for Europe, 2006, pp. v–73.
240. World Health Organization. (2017). Towards More Physical Activity in Cities. Towards More Physical Activity in Cities, 2017.
241. World Health Organization. (2019). Global Strategy on Diet, Physical Activity, and Health. https://www.who.int/dietphysicalactivity/pa/en/ accessed on 09/23/2019.
242. Wu, Y., D. Rowangould, J. K. London, and Karner, A. (2019). Modeling Health Equity in Active Transportation Planning. Transportation Research Part D, Vol. 67, 2019, pp. 528–540. https://doi.org/10.1016/j.trd.2019.01.011.
243. Xiao, G., Q. Cheng, and Zhang, C. (2019). Detecting Travel Modes from Smartphone-Based Travel Surveys with Continuous Hidden Markov Models. Vol. 15, No. 4, 2019. https://doi.org/10.1177/1550147719844156.
244. Xiao, G., Z. Juan, and Zhang, C. (2015). Travel Mode Detection Based on GPS Track Data and Bayesian Networks. Computers, Environment and Urban Systems, Vol. 54, 2015, pp. 14–22. https://doi.org/10.1016/j.compenvurbsys.2015.05.005.
245. Xu, W. L., X. Feng, C. Luo, J. Li, and Ming, Z. (2019). Energy Harvesting-Based Smart Transportation Mode Detection System via Attention-Based LSTM. IEEE Access, Vol. 7, 2019, pp. 66423–66434. https://doi.org/10.1109/ACCESS.2019.2918555.
246. Yang, D., Chenfeng, X., Tang, L., and Zhang, L. (2019). Travel Mode Detection Using Smartphone GPS Data: A Comparison between Random Forest and Wide-and-Deep Learning. TRB, 2019, pp. 1–5.

247. Yang, F., Z. Yao, Y. Cheng, B. Ran, and Yang, D. (2016). Multimode Trip Information Detection Using Personal Trajectory Data. Journal of Intelligent Transportation Systems, Vol. 20, No. 5, 2016, pp. 449–460. https://doi.org/10.1080/15472450.2016.1151791.
248. Yang, X., K. Stewart, L. Tang, Z. Xie, and Li, Q. (2018). A Review of GPS Trajectories Classification Based on Transportation Mode. Sensors (Switzerland), Vol. 18, No. 11, 2018, pp. 1–20. https://doi.org/10.3390/s18113741.
249. Yusak, O. Susilo, & Martin D. (2010) Behavioural decisions of travel-time ratios for work, maintenance and leisure activities in the Netherlands, Transportation Planning and Technology, 33:1, 19-34, DOI: 10.1080/03081060903429280.
250. Zeegers, M.P., Dirx, M.J. and van den Brandt, P.A. (2005). Physical activity and the risk of prostate cancer in the Netherlands cohort study, results after 9.3 years of follow-up. *Cancer Epidemiology and Prevention Biomarkers*, *14*(6), pp.1490-1495.
251. Zhang, J. and Chaaban, J. (2013). The economic cost of physical inactivity in China. *Preventive medicine*, *56*(1), pp.75-78.
252. Zhang, J., Li, Z. Pu, and Xu, C. (2018). Comparing Prediction Performance for Crash Injury Severity among Various Machine Learning and Statistical Methods. IEEE Access, Vol. 6, 2018, pp. 60079–60087. https://doi.org/10.1109/ACCESS.2018.2874979.
253. Zhang, L., L. Liu, S. Bao, M. Qiang, and Zou, X. (2015). Transportation Mode Detection Based on Permutation Entropy and Extreme Learning Machine. Hindawi Publishing Corporation, Vol. 2015, 2015.
254. Zhou, C., H. Jia, J. Gao, L. Yang, Y. Feng, and Tian, G. (2017). Travel Mode Detection Method Based on Big Smartphone Global Positioning System Tracking Data. Vol. 9, No. 6, 2017, pp. 1–10. https://doi.org/10.1177/1687814017708134.
255. Zhou, X., W. Yu, and Sullivan, W. (2016). Making Pervasive Sensing Possible : Effective Travel Mode Sensing Based on Smartphones. Computers, Environment and Urban Systems, Vol. 58, 2016, pp. 52–59. https://doi.org/10.1016/j.compenvurbsys.2016.03.001.
256. Zhu, Q., M. Li, M. Fu, Z. Huang, Q. Gan, and Zhou, Z. (2016). Identifying Transportation Modes from Raw GPS Data. Social Computing, 2016, p. 395 410.
257. Zhu, Q., M. Zhu, M. Li, M. Fu, Z. Huang, Q. Gan, and Zhou, Z. (2018). Transportation Modes Behaviour Analysis Based on Raw GPS Dataset. Accessed from https://www.inderscience.com/info/inarticle.php?artid=90569.
258. Zhu, W., J. Ash, Z. Li, Y. Wang, and Lowry, M. (2015). Applying Semi-Supervised Learning Method for Cellphone-Based Travel Mode Classification. IEE, 2015.

https://doi.org/10.1109/ISC2.2015.7366148.

259. Zong, F., Y. Bai, X. Wang, Y. Yuan, and He, Y. (2015). Identifying Travel Mode with GPS Data Using Support Vector Machines and Genetic Algorithm. Information, Vol. 6, No. 2, 2015, pp. 212–227. https://doi.org/10.3390/info6020212.

# APPENDIX

## Appendix A. HSIRB Protocol

Western Michigan University
HSIRB Application
Monitoring Daily Activities and Linking Physical Activity Levels Attributed to Transportation Mobility Choices and Built Environment (Mobile App)


| | | |
|---|---|---|
| Principal Investigator: | Jun-Seok Oh | Civil and Construction Engineering |
| Co- Principal Investigator: | Ala I Al-Fuqaha | Computer Science |
| Co- Principal Investigator: | Sangwoo Lee | Human Performance and Health Education |
| Student Investigator: | Raed Hasan | Civil and Construction Engineering |
| Student Investigator: | Hafez Irshaid | Computer Science |
| Student Investigator: | Md Mehedi hasan | Civil and Construction Engineering |


1. ABSTRACT

Physical activities become an importance part of human lives for healthy living. Research has shown that increase in physical and cardiovascular activities tends to decrease in diseases. Although there are several types of physical activities, non-motorized transportation options like walking, running and cycling provide natural ways of being physically active. Accordingly, non-motorized transportation options began being attracted thanks to their natural health benefits. The health benefit of the active transportation comes from participants' physical activities; however, there has been very limited effort in analyzing and quantifying participants' actual physical activities. This study proposes to identify and categorize health outcomes impacted by daily physical activity and quantify the amount of physical activities by different transportation mode users in different areas associated with their daily travel activities. By employing recent wearable devices with sensing and GPS tracking technology, the amount of physical and cardiovascular activities will be quantified by travel activities and transportation mode used. This research will help in incorporating human health into transportation planning by addressing health outcomes impacted by physical and cardiovascular activities associated with transportation options.

2. BACKGROUND INFORMATION

Transportation decisions impact human health at least in three ways, such as traffic crashes, environmental impact, and physical fitness. While there have been ample efforts in reducing traffic crashes and environmental impacts, less attention has been paid to their impacts on physical fitness. Recent efforts on the relationship between transportation and physical fitness were mostly from the context of active transportation. Potential benefits of active transportation include saving in mobility costs, benefits from related businesses, community savings in costs associated with health and environmental benefits.

The health benefit of the active transportation comes from participants' physical activities. Increase in physical activities tends to decrease in diseases. Studies have shown that persons with moderate to high levels of physical activity or cardiorespiratory fitness have a lower mortality rate than those with sedentary habits or low cardiorespiratory fitness. Furthermore, there was a significant trend of decreasing risk of death across increasing categories of distance walked, trips of stairs climbed, and degree of intensity of sports play. Physical activities for cardiorespiratory endurance reduces the risk of developing or dying from cardiovascular diseases (CVD), hypertension, colon cancer, and non-insulin-dependent diabetes mellitus (NIDDM) and improves mental health while endurance-type physical activity may reduce the risk of developing obesity, osteoporosis, and depression and may improve psychological well-being and quality of life. According to a report by the Centers for Disease Control and Prevention (CDC), benefits of physical activities are known 1) to help build and maintain healthy bones, muscles, and joints; 2) help control weight, build lean muscle, and reduce fat; and to prevents or delay the development of high blood pressure and helps reduce blood pressure in some adolescents with hypertension. However, it is still difficult to observe how transportation decisions affect physical fitness. Physical activity is difficult to measure directly. Three types of physical activity measures have been used in observational studies over the last 40 years. Most studies have relied on self- reported level of physical activity, as recalled by people prompted by a questionnaire or interview. A more objectively measured characteristic is cardiorespiratory fitness (also referred to as cardiorespiratory endurance) which is measured by aerobic power. Some studies have relied on occupation to classify people according to how likely they were to be physically active at work.

Although it is difficult to predict how a particular transportation planning decision affects physical fitness, total impacts are likely to be large. Diseases associated with inadequate physical fitness cause an order of magnitude of more deaths, and more than road crashes. Even modest reductions in these illnesses could provide significant health benefits. Therefore, there is a strong need for investigating how transportation options affect the physical activities.

## 3. PROPOSED RESEARCH GOALS AND OBJECTIVES

The primary goal of the research is to explore the factors impacting the amount of physical activity an individual engages in and the proportion of an individual's daily activity attributable to transportation activities. Specific research objectives include:

a. To develop a strategy for monitoring and recording the daily physical activity of a representative sample of 200 individuals in a small urban area and a large metropolitan area.
b. To develop data fusion strategy to combine data from Fitbit (including heart rate) with smart phone data.
c. To identify and categorize health outcomes impacted by daily physical activity.
d. To develop techniques for categorizing the location types by considering their relationship with transportation-related physical activity.
e. To use the fused data to classify a physical activity as recreational, activity-related (e.g. employment or shopping), or transportation-related
f. To create a preliminary scheme for using speed patterns/profiles to classify mode choice.
g. To test the statistical association between the physical activity levels of individuals, their socioeconomic and employment profiles and the nearby land use and site design based on a set of predetermined hypotheses.
h. To develop performance measures that include both physical activity (based on potential health impact) and travel time for evaluating transportation system, land use, and site design decisions

## 4. SUBJECT RECRUITMENT

Subjects will be recruited among those whose expressed their interests in a previous questionnaire survey (HSIRB 17-10-16). Only in Michigan, there are more than 200 subjects expressed their interests in participating in this main data collection. Subjects will be selected after evaluating individual's physical and transportation activities obtained from the previous survey. The criteria for selecting participants are the diversity of daily physical and transportation activities to cover wide range of population.

In this research, we will select a total of 200 subjects: 100 subjects from Michigan and the other 100 from Texas. The research team will contact selected participants and inform detailed tasks and commitments.

## 5. INFORMED CONSENT PROCESS

Once the interested participants are successfully recruited, they will be informed again of this project, and an informed consent form must be filled out, instructions on the experience at the Community Transport Research Center (TRCLC). After signing the consent form, the participant will be given a wearable Fitbit device and a unique identification ID to be used for further data collection and analysis. Finally, the participant will be asked to install the application prepared by the study team on the participant's phone. The application contains the terms and conditions which need to approval by the user. (See Appendix 1 for the Informed Consent Form)

## 6. RESEARCH PROCEDURE

### 6.1. Identify health outcomes impacted by physical activities

The research team will identify and classify health outcomes impacted by physical activities. The effort will allow the research team to incorporate those outcomes into the measurement system in our data collection devices.

### 6.2. Development of mobile application for data collection from wearable devices

The research team will develop a mobile app and its associated server-side infrastructure to collect the raw data about the daily activities of the participants from wearable devices. The app will include an algorithm that automatically classifies the transportation mode to one of the following: walk, run, bike, car, bus. Speed, frequency of stops, accelerometer data, and weather conditions, identity of the source and destination locations, and residence time at intermediate locations can help to identify the transportation mode. We hope the algorithm allows us to classify purpose of trip by mining the raw collected data. Regularity (repeatability), length, time of the day, day of the week, and weather conditions will help us to identify the purpose of the trip. We believe that machine learning techniques and potentially deep learning can help in creating better purpose of trip classifiers. The following details show the Information Technology (IT) infrastructure that will be utilized in this project to collect and analyze the raw data.

Underlying Software Technologies:

The mobile application will utilize the following Application Programming Interfaces (APIs) to implement its core functionality:

a. Cordova: To build a cross-platform mobile app using HTML5, JavaScript, and CSS3. The use of Cordova will allow to easily port the app to Android and iOS.
b. Fitbit Web APIs: To retrieve details about the physical activities of participating subjects including: a time series of their activities, heart rate, steps, calories, METs, and sleep logs.
c. Google Maps, Open Streets, or foursquare APIs: To retrieve details about visited locations.
d. Accelerometer and GPS APIs: To retrieve acceleration, estimate the number of steps, speed, and get details about users' traveled paths.
e. Web Services: To interface the mobile app with the back-end database and the data analysis and reporting services. These services will also allow for raw data access.

6.3. Recruitment of subjects

The research team will recruit subjects from two different areas, Michigan, Chicago, and Texas, in order to investigate seasonal variations and locational characteristics in their physical activities. Based on the survey, subjects for further data collection will be chosen by their characteristics. A stratified sampling technique will be employed to cover range of differences among subjects.

6.4. Data collection and analysis

1) Pre-survey

A pre-survey will be implanted at the first meeting with participants. The survey is to understand general travel and physical activities (Pre-survey Questionnaire available in Appendix 1). In additional to pre-survey, individual subjects will be asked to fill in the device registration form that includes their contact information to assure continuation of the data collection during the study period as well as locations they typically visit. The contact information will only be used when the research team needs to contact the subject because of the system failure in collecting data. The location information will be used for converting their locations into an aggregated format (type of location).

2) Body Composition Data Collection

The following body composition data will be assessed three times (at the beginning, six month later, and one year later) using a non-invasive bioelectrical impedance analyzer (InBody 570; InBody Co., Ltd., Seoul, Korea).

- Muscle-fat analysis: weight (lbs.), skeletal muscle mass (lbs.), and body fat mass (lbs.)
- Obesity analysis: body mass index (kg/m2) and percent body fat (%)
- Segmental lean analysis: right & left arms and legs (lbs.) and trunk (lbs.)

3) Travel and Physical Activity Data Collection

Travel and physical activity data will be collected using a wearable device (Fitbit Charge 2/3) and a mobile app to be installed in the subject's mobile phone.

Travel Activity Data

The mobile app developed will collect daily travel activities including locations from GPS in the subject's mobile phone.

Physical Activity Data from wearables

The physical activity data will be collected by the wearable device (Fitbit Charge 2/3) and the data will be accessed by the mobile app developed by the research team. The subjects will be required to install the mobile app developed to transmit the data to a cloud server. Wearable devices will be distributed to the subjects as an appreciation of participation.

Physical activity data from typical wearables includes:

- Distance traveled (km): distance is calculated by multiplying walking (running) steps by walking (running) stride lengths. The stride lengths are estimated using height and gender.
- Heart rate (beats/min): both resting heart rate and heart rate with physical activities are estimated using a heart rate monitor with photo plethysmography.
- Activity minutes (min): active minutes are estimated using metabolic equivalents (MET). MET is an indication of how much harder than set a particular activity is. For example, 1- MET indicates a body at rest,

therefore, 3-MET means three times harder than rest, such as stationary cycling or walking at a rate of 4 km/h. MET is estimated in any given minutes by calculating the intensity of physical activity. Active minutes are then earned at or above 3- MET.

- Total calories (Kcal): total calories are estimated by taking into account basal metabolic rate (BMR) and calories consumed during physical activities in a day.
  - BMR: BMR is calculated based on gender, age, height, and weight.
    - For men: BMR = 10 × body mass (kg) + 6.25 × height (cm) – 5 age (years) +5
    - For women: BMR = 10 × body mass (kg) + 6.25 × height (cm) – 5 age (years) - 161
  - Calories consumed during physical activities (total calories – BMR): these calories are estimated using the above-mentioned heart rate monitor and a three-axis accelerometer.

4) Data Analysis

The recorded data will be analyzed anonymously. Data collected in this research will be sorted out into following five categories and analyzed their relationship.

- Individual characteristics – age, gender, employment, body composition, fitness exercise, amount of physical and cardiovascular activities, etc.
- Environment characteristics – weather, temperature
- Built Environment – area type, land use, accessibility to physical activity facilities, park accessibility
- Transportation Environment – public transit accessibility, automobile availability, walkability, Bikeability
- Travel activity – trip purpose, transportation mode, travel time
- Physical activity – amount of physical activity

Through the data analysis, the research team aims to develop models quantifying the amount of physical activities associated with various conditions listed above.

## 7. LOCATION OF DATA COLLECTION

The data will be collect from all locations where subjects are traveling. Subjects are recruited from Michigan and Texas, but the data could be collected wherever the subjects are traveling. The data from all participants will be stored in the TRCLC Laboratory (G-208 and/or F-212).

## 8. DURATION OF THE STUDY

Participants' daily activity data will be collected for one year from the dissemination of wearable devices. The data collection period is expected to be from May 1, 2018 to April 30, 2019. After the data collection, the research team will analyze the data for another two months. Therefore, the total duration of the study is 15 months from May 1, 2018 to June 31, 2019.

## 9. METHODOLOGY

The research design uses a mixed-methods approach that utilizes quantitative techniques to analyze qualitative, coded data related to socioeconomic factors, physical characteristics, employment characteristics, seasonal effects, land use, and site design. This study has four key dependent variables to consider: total physical activity, total physical activity related to transportation, total cardiovascular activity (measured as time spent at age specific levels of exertion), and total cardiovascular activity related to transportation. Differentiating between recreational physical activity and transportation related physical activity requires an app that fuses activity data from a wearable with the GPS tracking from a smartphone. The research team will need to select and code a collection of independent variables related to employment type (e.g. service, office, construction, and manufacturing), socioeconomic factors (e.g. race, gender, and income), physical characteristics (e.g. age, body mass index, resting heart rate), seasonal effects (e.g. mean temperature), land use (e.g. walkability, density, urban vs. suburban), and site design (e.g. walking time to parking, walking time to transit). In particular, this project seeks to examine the roles that seasonal effects, land use and site design play in activity levels while controlling for employment, physical characteristics, and socioeconomic effects. Regression and logistic regression analyses will be used to test several hypotheses about the relationships between total physical activity, total physical activity related to transportation, total cardiovascular activity (measured as time spent at age specific levels of exertion), and total cardiovascular activity related to transportation with the independent variables.

The researchers expect that increasing the role of transportation in achieving physical and cardiovascular activity will have a positive impact on health outcomes. For this set of hypotheses, the activity levels become independent variables and health outcomes becomes the dependent variable; pertinent control variables from the first set of analyses will be included as independent control variables. The conclusions from the modeling and consideration of the hypotheses presented below will enable the research team to develop performance measures to evaluate health and travel time tradeoffs.

10. OUTCOME DISSEMINATION PLAN

The outcome dissemination plan aims to target three types described below:

a. Report submission: the final research report will be published on the TRCLC website
b. The report will be also disseminated to followings:
    a. Transportation Research Board through the TRB's Transportation Research International Documentation Database (TRID)
    b. National Transportation Library
    c. U.S. Department of Transportation Research Hub
    d. Transportation Library at Northwestern University
    e. Volpe National Transportation Systems Center
    f. FHWA Research Library, Turner-Fairbank Highway Research Center
    g. U.S. Department of Commerce, National Technical Information Service
c. Academic and Research Community: Conference outlets include the Transportation Research Board Annual meeting, the International Conference for Transport and Health, the American Society of Public Administration (ASPA) Annual Conference; Journals include Transportation Research Record, Public Administration Review, Public Works and Management Policy, Transportation Research Record and Transport Policy.

11. RISKS AND COST TO AND PROTECTIONS FOR SUBJECTS

There are no known risks in this study. All participant information, as well as collected physical activity data, will remain confidential and available only to the research team. In order to protect privacy, person information and activity data will be separately stored.

12. BENEFITS OF RESEARCH

Successful completion of this research may enable to help in incorporating human health into transportation planning by addressing health outcomes impacted by physical and cardiovascular activities associated with transportation options.

13. CONFIDENTIALITY DATA

Data for participants who consent to participate will be anonymized since participants will be assigned unique IDs that do not contain or link to personal information. Since no personal information is needed nor will be collected, confidentiality is not an issue. The data will be retained in the Principal Investigator's office for at least three years after experiment completion.

## A.1 Informed Consent Form

Western Michigan University
Department of Civil and Construction Engineering

| | | |
|---|---|---|
| Principal Investigator: | Jun-Seok Oh | Civil and Construction Engineering |
| Co- Principal Investigator: | Ala I Al-Fuqaha | Computer Science |
| Co- Principal Investigator: | Sangwoo Lee | Human Performance and Health Education |
| Student Investigator: | Raed Hasan | Civil and Construction Engineering |
| Student Investigator: | Hafez Irshaid | Computer Science |
| Student Investigator: | Md Mehedi Hasan | Civil and Construction Engineering |
| Title of Study: | Monitoring Daily Activities and Linking Physical Activity Levels Attributed to Transportation Mobility Choices and Built Environment | |

You have been invited to participate in an experiment for a research project, "Monitoring Daily Activities and Linking Physical Activity Levels Attributed to Transportation Mobility Choices and Built Environment" funded by the U.S. Department of Transportation through the Transportation Research Center for Livable Communities (TRCLC) at Western Michigan University. This consent document will explain the purpose of this research project and will go over your commitments, the procedures used in the study, and the risks and benefits of participating in this research project. Please read this consent form carefully and completely and please ask any questions if you need more clarification.

<u>What are we trying to find out in this study?</u>

This research is to explore the factors impacting the amount of physical activity an individual engages in and the proportion of an individual's daily activity attributable to transportation activities. This research will be performed data analysis for body composition data, physical activity data, and travel data using wearable device. This research will help in incorporating human health into transportation planning by addressing health outcomes impacted by physical and cardiovascular activities associated with transportation options.

<u>Who can participate in this study?</u>

Anyone can participate in this study.

<u>Where will this study take place?</u>

There is no specific place, as a participant, be free in your movements and habits. All you need to do is live as usual with Fitbit Charge 2/3 during the duration of the project. Data will be transmitted to the server in the lab and monitored by the research team at the WMU's Transportation Research Center for Livable Communities (TRCLC), located at Parkview campus room number G-208/F- 212.

<u>What is the time commitment for participating in this study?</u>

The data collection duration is 12 months from the reception of Fitbit Charge 2/3

<u>What will you be asked to do if you choose to participate in this study?</u>

After being introduced to the experiment, you will be asked to 1) sign the informed consent form after reading the terms and conditions, 2) install a mobile app developed, 3) activate the wearable device, and 4) perform a trial run using wearable device. After the trial run is successful, you simply turn on the mobile app and wear the wearable device during your daily life. In case you do not receive the data, we will contact you to verify your wearable devices and smart phone. In addition, we will measure your body composition in our lab every three months to see the changes.

<u>What information is being measured during the study?</u>

Data collection will be conducted in two ways: measuring body composition and measuring travel and physical activity. The body composition data will be assessed every six months (at the beginning, six month later, and one year later) using a non-invasive bioelectrical impedance analyzer (InBody 770; InBody Co., Ltd., Seoul, Korea), such as muscle-fat data (weight, skeletal muscle mass and bidy fat mass), obesity-related data (body mass index and percent body fat) and segmental lean analysis (two arms, two legs and trunk).

The travel and physical activity data will be collected by the wearable device and the mobile app developed in this project. Data will be transmitted to the server in our lab and analyzed by the research team. The travel and physical activity data include activity locations, activity duration, travel distance, heart rate, and total calories during travel and physical activities. Although individuals' data are collected, individuals' privacy will be protected by processing all data in an aggregated format.

<u>What are the risks of participating in this study and how will these risks be minimized?</u>
There is no risk in this research other than a potential that your travel locations are identified. The research team will not disclose the information and data collected in this research to any party. The data will only be used for the purposes of the research and securely stored. As data will be used in an aggregated format, it will not reveal the identity of participants or information that affect their privacy.
<u>What are the benefits of participating in this study?</u>
The participants will own the Fitbit Charge 2/3 after the end of this study and will be able to receive her/his body composition data upon requested.
<u>Are there any costs associated with participating in this study?</u>
There are no costs of participating in this study.
<u>Is there any compensation for participating in this study?</u>
The participants will own the Fitbit Charge 2/3 after the end of this study.

<u>Who will have access to the information collected during this study?</u>
The data collected will be analyzed only by the research team members. No others will have access to the data collected. The results of the study are expected to be disseminated on an aggregate basis through a report to the US Department of Transportation as well as possible journal/conference publications.
<u>What if you want to stop participating in this study?</u>
You can stop participating in the study at any stage if you feel uncomfortable. There are no legal or financial consequences as a result of this decision other than an obligation of returning the wearable device provided. The research team may also decide to suspend your participation without your consent if deemed necessary. If your participation is cancelled for any reason within a six-month period, you must return the wearable device to the research team in a week.

Should you have any questions prior to or during the study, you can contact the principal investigator, Dr. Jun-Seok Oh, by e-mail at jun.oh@wmich.edu, You may also contact the Chair, Human Subjects Institutional Review Board at (269) 387-8293 or the Vice President for Research at 269387-8298 if questions arise during the course of the study.

This consent document has been approved for use for one year by the Human Subjects Institutional Review Board (HSIRB) as indicated by the stamped date and signature of the board chair in the upper right corner. Do not participate in this study if the stamped date is older than one year.

----------------------------------------------------------------------------------------------------------------

I have read this informed consent document. The risks and benefits have been explained to me. I agree to take part in this study.

______________________________________________________________________________
Please Print Your Name

______________________________ ______________________________
Participant's signature Date

# Appendix B. Initial Survey

You are invited to a survey intended to ask your general travel and physical activities in order to select research participants who will receive a wearable device, possibly Fitbit Charge 2/3, for physical activity data collection. This research titled, “Monitoring Daily Activities and Linking Physical Activity Levels Attributed to Transportation Mobility Choices and Built Environment” is sponsored by the U.S. Department of Transportation through the Transportation Research Center for Livable Communities (TRCC) at Western Michigan University. We would like to understand individuals’ physical activities associated with their travel patterns for a sampling purpose. This survey will take less than 5 minutes to complete. Your participation in this study is voluntary. There are no known risks associated with this project. However, if you feel uncomfortable answering any question(s), you can withdraw from the survey at any point. It is very important for us to learn your opinions.
The data collected will be used in an aggregated format and your personal information in case you provide for the next data collection will only be used only when contacting you as a research participant. If you have questions about the survey or the research, you may contact the Principal Investigator, Dr. Stephen Mattingly, by calling at (817) 272-2630 or emailing at mattingly@uta.edu at any time.  Thank you again for your time and support. You can now start answering the survey questions by click on the button below.

1. What is your age?

- 18 - 29
- 30 - 39
- 40 - 49
- 50 - 59
- 60 - 69
- 70+

2. What is your gender?

- Female
- Male

3. What are your transportation modes and approximate duration (minutes) for each transportation mode in a typical day?

(exclude unusual long-distance trips)?

- Motor Vehicle (driving) Duration __________ minutes
- Motor Vehicle ( as a passenger) Duration__________ minutes
- Bus Duration__________ minutes
- Rail/Subway Duration__________ minutes
- Bicycle Duration__________ minutes
- Walk Duration__________ minutes
- Other__________ Duration__________ minutes

4. Would you list all physical activities you did during last seven days? Physical activities include aerobic activities (walking, running, swimming, bicycling, tennis, soccer, and etc), muscle-strengthening (lifting weight, push-ups, etc), and Toning/Stretching (yoga, etc)

| | Activity | Number of time per week | Average duration (minutes) |
|---|---|---|---|
| 1 | | | |
| 2 | | | |

| | | | |
|---|---|---|---|
| 3 | ☐ | ☐ | ☐ |
| 4 | ☐ | ☐ | ☐ |
| 5 | ☐ | ☐ | ☐ |

* 5. Are you using a smartphone?

- ○ No
- ● Yes, I use iPhone
- ○ Yes, I use an Android Phone
- ○ Yes, I use other types of smart phone

* 6. If we provide a wearable device (Fitbit Charge 2/3), would you be willing to share your activity data from the device?

- ○ No
- ○ Yes (Please provide your contact information)

7. If you answered yes to the previous question, please provide your information.

First name : ______

Last name : ______

Phone : ______

Email Address : ______

# Appendix C: Main Survey

This research titled, "Monitoring Daily Activities and Linking Physical Activity Levels Attributed to Transportation Mobility Choices and Built Environment" is sponsored by the U.S. Department of Transportation through the Transportation Research Center for Livable Communities (TRCLC) at Western Michigan University. As a part of this research, we would like to know your general travel and physical activities. There are no known risks associated with this project. However, if you feel uncomfortable answering any question(s), you can withdraw from the survey at any point.

This survey will take approximately five minutes to complete. Your participation in this study is completely voluntary. However, in order to participate in the data collection with Fitbit Charge, you must participate in this survey. Even though we are asking your contact information in this survey, your personal information will neither be used other than contacting you nor be shared with anyone. Your contact information will be strictly confidential and stored securely. Thank you so much for your participation in this survey.

Please answer the following questions by filling or circling as required.

1. What is your gender?
   a) Female
   b) Male
   c) Other
2. What is your age group?
   a) Under 18
   b) 18 - 25
   c) 26 – 49
   d) 50 – 64
   e) 65 – 75
   f) More than 75
3. Which race/ethnicity best describes you? (Please choose only one)
   a) American Indian or Alaskan Native
   b) Asian / Pacific Islander
   c) Black or African American
   d) Hispanic American
   e) White / Caucasian
   f) Other (please specify): (                    )
4. What is your highest level of education?
   a) Some high school education, but no diploma
   b) High school graduate with a diploma or equivalent (for example: GED)
   c) Some college credits, but no Bachelor's degree
   d) Bachelor's degree or higher
5. What is your currently professional status?
   g) Student
   h) Administration position
   i) University faculty
   j) Office worker
   k) Outdoor worker
   l) Not currently employed / home with family
6. What is your annual income?
   a) Less than $30,000
   b) $30,000 - $50,000
   c) $50,000 - $100,000
   d) $100,000 - $150,000
   e) More than $150,000
7. How would you rate your general health condition?
   a) Excellent
   b) Good
   c) Fair

   d) Bad
   e) Very bad
8. Do you have a driver's license?
   a) No
   b) Yes
9. How many motorized vehicles (including motorcycles, mopeds, cars, vans and trucks) does your household have? It can include owned, leased or any available vehicle for regular use.
   a) 0
   b) 1
   c) 2
   d) 3+
10. Is any public transit system available in your living area?
   a) No
   b) Yes
11. Would you tell us your fitness activity (activity only for fitness purpose not for travel purpose) during past week?

| | Type of fitness activities (list all applicable)[1)] | Average estimated level of intensity (1 – 10)[2)] | Total amount of time in a day (in minutes) |
|---|---|---|---|
| Sunday | | | |
| Monday | | | |
| Tuesday | | | |
| Wednesday | | | |
| Thursday | | | |
| Friday | | | |
| Saturday | | | |

1) Example of fitness activities: Slow walking, Brisk (fast) walking, Jogging, Running, Cycling, Weight training, Group exercise, Basketball, Soccer, Tennis, etc.
2) Estimated level of intensity: 1 being very light, 10 being extremely hard

12. How long do you typically spend your time for your travel from home to work / school? (Please provide all modes that you use for the commute.)

| Mode you use | Time (in minute) | Check a primary mode |
|---|---|---|
| Walk | | |
| Bicycle | | |
| Drive | | |
| Passenger of auto | | |
| Wait/transfer | | |
| Bus | | |
| Rail | | |
| Taxi | | |
| Motorcycle | | |
| Other ( ) | | |
| Total | | |

# Appendix D: Device Registration Form

| Name | Registration Code | Serial Number (Fitbit Charge 2/3) |
|---|---|---|
| | | |

| Cell Phone Number | Secondary Phone Number | Email | Secondary Email Address |
|---|---|---|---|
| | | | |

List five places you visit frequently (home, work, ……, etc.)

| | Location | Address |
|---|---|---|
| 1 | Home | |
| 2 | Work | |
| 3 | | |
| 4 | | |
| 5 | | |

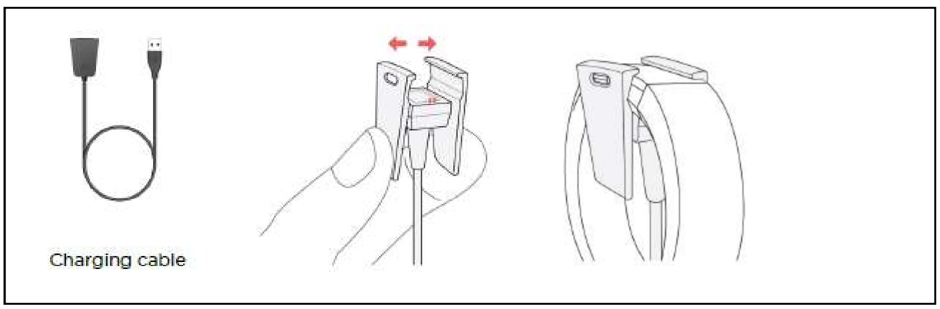
- This information will only be used for aggregating your activity data.

# Appendix E: User Manual

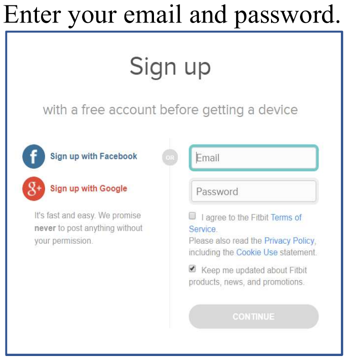
## 1. Installing Fitbit Charge 2/3

fitbit charge 2

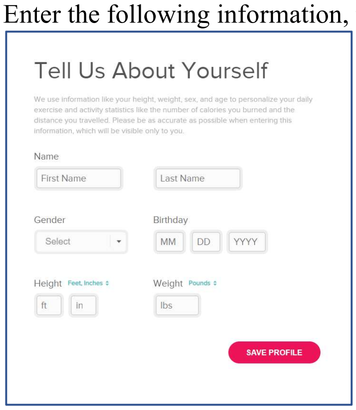
After you receive the Fitbit Charge 2/3, open the box and make sure that the Fitbit works properly and make sure it's fully charged by pressing the side button of the smart-watch for several seconds. If the watch is not charged, it must be connected to the charger which exists in the same box for the smart-watch




Create Account in Fitbit using this link. [https://www.fitbit.com/signup](https://www.fitbit.com/signup)
Enter your email and password. Then click Continue

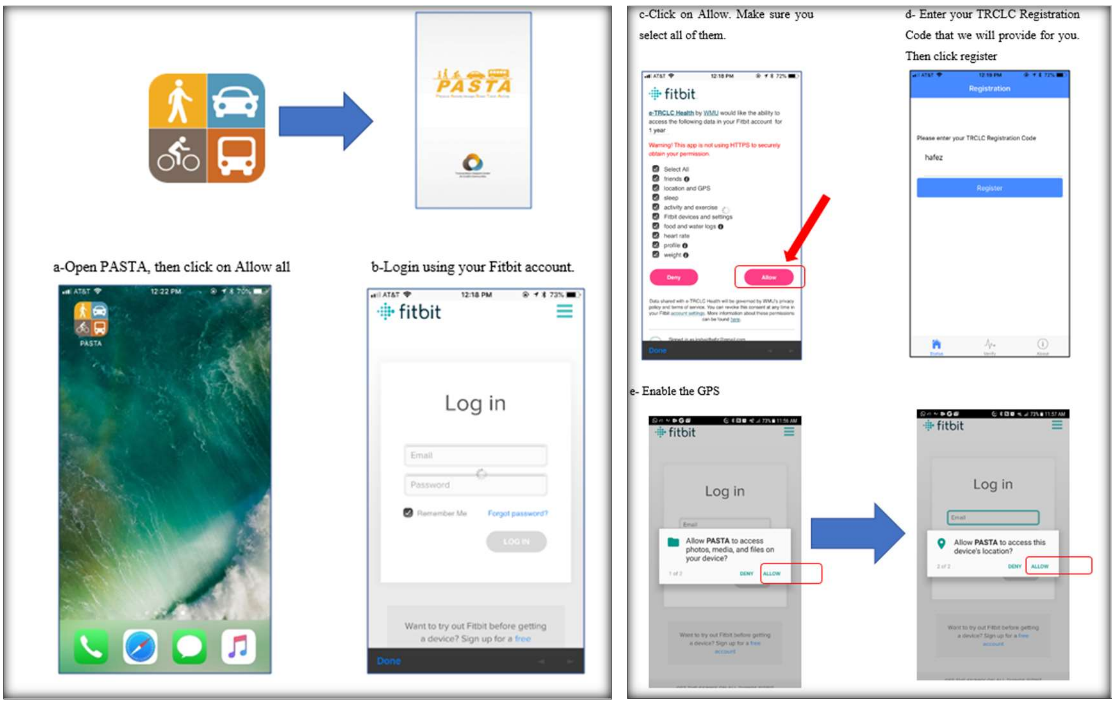



Enter the following information, then click save profile.



- Install Fitbit App in your phone using Play store, then login using your Fitbit account that you created.
- Follow this video to connect Fitbit watch with your phone. (when you press to start the video, wait few second till you see the video is working) Or you can use the like: https://www.youtube.com/watch?v=mH4KHlNuKWQ
- When the registration information for the Fitbit application installed on your phone is complete, you must make sure that you have completed the verification of your email and the password. If you do not complete the verification step, PASTA application will not be able to enter your email and password to activate it.

## 2. Install PASTA application in your phone. Search for *"WMU PASTA"*

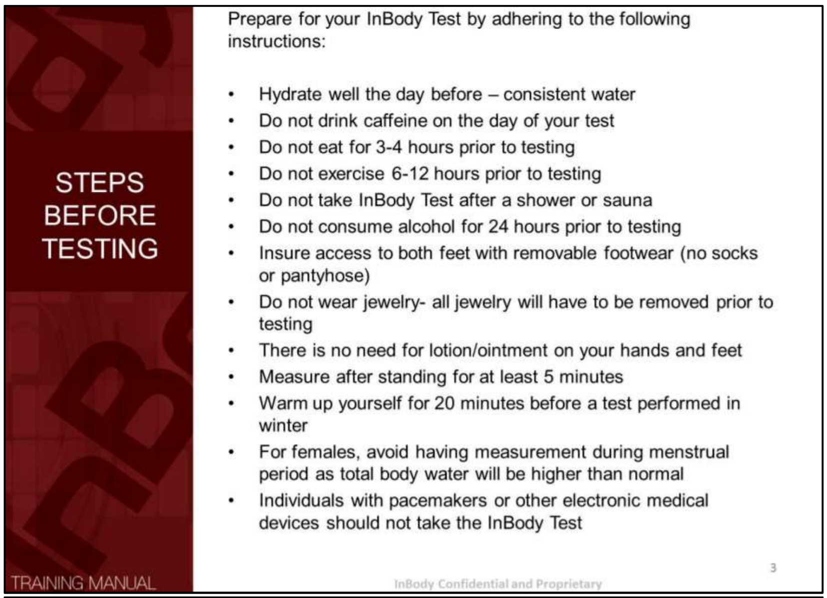


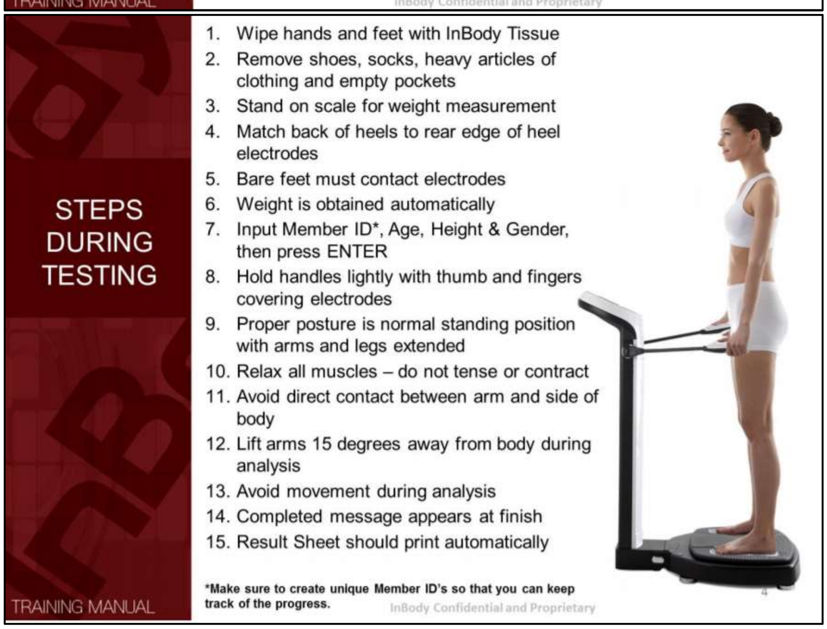

## 3. User's responsibilities

1) Wear a Fitbit watch daily
2) Check the charge of the Fitbit watch battery periodically
3) Check periodically that the PASTA application is running
4) Verify that the activities when requested by the application
5) Respond to the investigation team when receiving a call or email (to resolve any study problem)
6) If the participant wants to change the phone, he or she must repeat the steps again as described in this manual or contact the research team

The following body composition assessments will be conducted twice (before and after):

- BMI ($kg/m^2$): body mass/height$^2$; Percent body fat (%); Lean body mass (kg)
  - (This information can get from the InBody device directly)




- Waist-to-hip girth ratio
  - Divide the waist circumference by the hip circumference