Form Approved

OMB No. 0990-0390

Expiration Date 04/30/2018

**The Healthy Behavior Data Challenge**

Phase 1 Submission Template

**Introduction**

The Healthy Behavior Data Challenge responds to the call for new ways to address the challenges and limitations of self-reported health surveillance information and tap into the potential of innovative data sources and alternative methodologies for public health surveillance.

The Healthy Behavior Data Challenge will support the development and implementation of prototypes to use these novel methodologies and data sources (e.g., wearable devices, mobile applications, and/or social media) to enhance traditional healthy behaviors surveillance systems in the areas of nutrition, physical activity, sedentary behaviors, and/or sleep among the US adult population aged 18 years and older.

The collection of health data through traditional surveillance modes including telephone and in-person interviewing is becoming increasingly challenging and costly with declines in participation and changes in personal communications. In addition, the self-reported nature of responses particularly in the areas of nutrition, physical activity, sedentary behaviors, and sleep has been a major limitation in these surveillance systems, since self-reported data are subject to under/over reporting and recall bias. Meanwhile, the advent of new technologies and data sources including wearable devices (Fitbit, Garmin, Adidas, Jawbone, Smart Watches, Activity Trackers, etc.), mobile health applications on smartphones or tablets, and data from social media represents an opportunity to enhance the ability to monitor health-related information and potentially adjust for methodological limitations in traditional self-reported data.

The Healthy Behavior Data Challenge will harness this potential and identify feasible alternative options for collecting health-related behaviors in new ways. Conducted in two phases, Phase I (Prototype Development) entails Challenge participants developing a concept proposal for obtaining data collected from wearable devices, mobile applications and/or social media for public health surveillance purposes.

The Healthy Behavior Data Challenge participants will propose data sources and approaches for aggregating data from wearable devices, mobile applications and/or social media in the areas of nutrition, physical activity, sedentary behaviors, and/or sleep. In Phase II (Prototype Implementation), a subset of submissions (up to 3) with promising concepts will be invited to test their proposed approaches for ongoing public health surveillance.

**Website**:

Additional Information:

Information on the Behavioral Risk Factor Surveillance System can be found at [www.cdc.gov/brfss](http://www.cdc.gov/brfss). Details on the HBD Challenge may be found at challenge.gov.

For Further Information Contact: Dr. Machell Town at BRFSSinnovations@cdc.gov.

**Submission Deadline**:

1. Challenge Team Information

|  |  |  |
| --- | --- | --- |
| Team Name |  |  |
| eCaregiving team | | |
| Team Lead |  | City/Province |
| Adam Bouras |  | Columbia/Missouri |
| E-mail |  | Phone Number |
| Bourasm@health.missouri.edu |  | 573-884-9095 |
| Subject-matter/domain expertise |  |  |
| Health Informatics |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Team Member #1** |  | **E-mail** |  | **Subject-matter/domain expertise** |
| **Satvinder Dhingra** |  | **sdhingra@quantextual.co** |  | **Health surveys and multi-sided digial platforms** |
| **Team Member #2** |  | **E-mail** |  | **Subject-matter/domain expertise** |
| **Richard Ellis** |  | **rellis@quantextualized.com** |  | **Project management** |
| **Team Member #3** |  | **E-mail** |  | **Subject-matter/domain expertise** |
|  |  |  |  |  |
| **Team Member #4** |  | **E-mail** |  | **Subject-matter/domain expertise** |
|  |  |  |  |  |
| **Team Member #5** |  | **E-mail** |  | **Subject-matter/domain expertise** |
|  |  |  |  |  |
| **Team Member #6** |  | **E-mail** |  | **Subject-matter/domain expertise** |
|  |  |  |  |  |

|  |
| --- |
| **Are all team members residents of the United States?** |
| **Yes** |

1. Organization (if submitting on behalf or as part of an organization)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Organization Name** |  | **Website** |  | **Type of Organization** |
| Department of Health Management and Informatics (University of Missouri) |  | **http://hmi.missouri.edu/** |  | **Academic Department** |

1. How did you find out about this challenge?

|  |
| --- |
| CDC website |

1. Submission Overview

|  |
| --- |
| **Project Title** |
| **Prospective Cohort Study to Assess Physical Activity, Sedentary Behavior, And Diet Using Passive and Active Data: Tale of Two Recruitment Approaches** |
| **Project Overview** |
| Describe in 500 words or less:   * What aspects of sleep, physical activity, nutrition, and sedentary behavior do you propose to report on and why are they important for public health surveillance? * Provide a brief description of the source(s) of data that will be used to report on these aspects, how your team proposes to access them, and why they are appropriate for use in public health surveillance? * How do you see your concept improving on current public health surveillance in the areas of sleep, physical activity, nutrition, and sedentary behaviors?   **This project focuses on understanding the relationship between physical activity, nutrition, and sedentary behaviors using both passive and active data. The Quantextual.co platform, mobile Health (mHealth apps) and researcher dashboard developed by Public Good Ventures Limited has been licensed to the Department of Health Management and Informatics (University of Missouri). This application is designed to capture and deploy research and behavioral interventions while empowering study participants with data and information to improve their own health. The platform provides capabilities to capture passive data from biosensor devices and active data like patient reported outcomes. For passive data, the application syncs with most wearable devices and collects high frequency data, which allow any stakeholders to have access to biosensor data that are robust and not subject to recall bias. The application also enables users to collect social and location data to administer location-based and event-based surveys. These data will help improve public health by collecting relevant information on the prevalence of health behaviors, which are one of the leading causes of illness in the USA. Furthermore, these prevalences can be used to target most needed areas and evaluate the effectiveness of a health intervention.**  **In addition to the bio data, the application also equips individuals to connect to patient portals and enables collection of medical and genotype data. These data can augment the patient reported outcome and biosensor data for better surveillance.**  **We plan to implement this new concept using two sampling approaches. The first approach is probability-sampling approach and the second one is non-probability sampling approach.**  **To implement the probability sampling approach, we propose to include two additional questions to the Center of Disease and Control and Prevention (CDC) supported, dual-frame, random digit dialed, state-based Behavioral Risk Factors Surveillance System (BRFSS) surveys. The purpose of these questions is to invite BRFSS survey respondents to join a panel for future studies for the Healthy Behavior Data Challenge. By doing so, we can guarantee that we have a random representative sample at the state and the national level. As for the non-probability sampling approach, we propose to recruit these panel members using social media (like Facebook, Google, and Twitter etc….). By using these two different approaches, we will be able to assess prevalence estimates from each approach at the fraction of the cost of the existing methodologies of conducting surveys.** |

1. Indicators to be measured (the indicators listed below are not comprehensive and innovators are recommended to include other relevant indicators)
   1. Physical Activity

* Amount of MVPA[[1]](#footnote-0) time per day
* Amount of MVPA time per day obtained in bouts of 10 minutes or more
* Amount of MVPA time accrued while at work, at home and/or in transit
* Identification of times during the day where MVPA is high
* Daily number of steps
* Miles/km (Distance) on foot or other modes of active transportation
* Frequency of MVPA
* Calories burned
* Type of activity (aerobic, strength, etc.)
* Level of activity (low, moderate, high)
* Time spent in different domains of MVPA (home/occupational, travel and recreational)
* Location of MVPA (recreation facility, at home, at work, on sidewalk/bike lane)
* Perception of safety while active
* Enjoyment level of the MVPA
* Number/flights of stairs climbed
* Average and peak heart rate
* Hours per week adults spent in sports, fitness or recreational physical activities
* Other indicators
  1. Sedentary Behavior[[2]](#footnote-1)
* Amount of time per day spent sedentary, excluding sleep time
* Amount of time per week spent on a computer/screen including watching TV, videos, playing computer games, emailing or using the internet
* Amount of sedentary time accrued while at work, at home and/or in transit
* Sitting time at work/ number and frequency of breaks at work from sedentary time
* # of hours spent in a car or motor-vehicle
* Other indicators
  1. Sleep
* Hours of sleep per night (sleep duration)
* Amount of time awake after sleep onset
* Sleep efficiency
* Amount of time to fall asleep (i.e., sleep latency)
* Consistency of bedtime
* Consistency of wake time
* Amount of time in REM vs. non-REM sleep (duration of sleep stage)
* Type of activity directly before sleep (e.g., screen time, reading, TV)
* Sleep satisfaction in morning
* Daytime sleepiness
* Other indicators
  1. Nutrition
* Total calories consumed per day
* Total calories from fat
* How often fruit (not including juices) was consumed (day, week, or month)
* How many times per day/week/month a green leafy or lettuce salad, with or without other vegetables, was eaten
* How often vegetables (not including lettuce salads and potatoes) was eaten (day, week, or month)
* Number of sugar-sweetened beverages consumed in a week (or per day)
* Number of caffeinated drinks consumed in a week (or per day)

1. Summary of proposed data source(s) (complete applicable sections)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Data Source** | | **Data Accessibility** (e.g., API, specialized software, existing data set) | **Data Cost** (i.e., fee for access, open access) | **Data Recency and Update Frequency** (i.e., how recent is the data and how often is it collected) | **Applicable Functional Area(s) and Indicator** (i.e., physical activity, nutrition, sleep, and/or sedentary behavior) | **Existing Users of the Data Source** (i.e., identify examples of organizations or other groups that have or are using the data source) |
| **Organization (e.g., company)** | Method of Collection (e.g., wearable, self-reported) |
| *1* | HMI (Quantextual.co) | self-reported BRFSS | Survey questionnaire | $20~35 per completed survey | Daily | physical activity, nutrition, sleep, and/or sedentary behavior | Any organization or individual that are using BRFSS data set |
| *2* | HMI (Quantextual.co ) | Wearable devices | Number of steps (different wearable devices collect different phsycial activity related measures ) | $25~$150 per participant (depending on how is the study period and the granularity of the data) | Per Minute | physical activity, sedentary behavior | Any organization or individual that are using BRFSS data set |
| *3* | HMI (Quantextual.co ) | Wearable devices | Number of steps (different wearable devices collect different sleep activity related measures ) | $25~$150 per participant (depending on how is the study period and the granularity of the data) | Per Minute | Sleep | Any organization or individual that are using BRFSS data set |
| *4* | HMI (Quantextual.co ) | self-reported | Survey questionnaire | $20~35 per completed survey | Daily/Weekly | Nutrition | Any organization or individual that are using BRFSS data set |
| *5* |  |  |  |  |  |  |  |
| *6* |  |  |  |  |  |  |  |
| *7* |  |  |  |  |  |  |  |
| *8* |  |  |  |  |  |  |  |
| *9* |  |  |  |  |  |  |  |
| *10* |  |  |  |  |  |  |  |

1. Describe how the data that you will use provides information and insight that is complementary to or more novel and innovative than that currently utilized for public health surveillance by CDC? (Novelty/innovation can apply at the level of the individual data source(s) selected, the specific indicators to be measured, tools/solutions that are used to capture the data, or result from newly created linked data sets). (750-word limit)

**The BRFSS questionnaire is administered at the state level. Currently, the questionnaire has three parts: 1) the core component, consisting of the fixed core, rotating core, and emerging core, 2) optional modules, and 3) state-added questions. All health departments must ask the core component questions without modification in wording, however, the modules are optional. Like other comprehensive publicly available health-related data at national and state levels, BRFSS is only available annually with 2 year lags; sub-state level data are sparse, possibly flawed, costly, and not timely. Sub-state level data are available only for the largest ~300 counties while for the remaining ~2,800 counties, 7 years of data must be aggregated to produce stable estimates. Biometric/ and vital signs data available only at the national level (NHANES), are more than 2 years old and released every 2 years. Survey questionnaires are finalized a year or two in advance making it impossible to address emerging health-related issues. The questionnaires are already too long, and sub-state stakeholders, independent and institutional researchers have no mechanism to add questions. These government run survey systems cost tens of millions of tax payer dollars annually. Traditional sampling and data collection are infeasible for most sub-state municipalities and non-government organizations. The confluence of increasing cost of healthcare, austere government budgets, and affordable cloud-based technologies provide an opportunity to do “more with less.”**

**The rise of apps in an increasingly mobile world holds many promises for data collection. Apps can make surveys more accessible and might allow people to provide information that is hard to collect using traditional survey methodologies. Polls have shown that the majority of Americans depend on their smartphones to access the internet and for entertainment.[[3]](#footnote-2),[[4]](#footnote-3) Smart phones offer their users the ability to navigate and access many mobile apps simultaneously and on the go. This access makes mobile users vulnerable to short attention span, which offers new opportunities and challenges to researchers to conduct app surveys. A market research report suggested limiting the number of questions to no more than 15 questions and the survey administrator can send as many surveys frequently as possible.[[5]](#footnote-4) Researchers can therefore achieve high response and completion rate.**

**Similarly, we propose to administer BRFSS questionnaire through our apps, and by doing so the survey questionnaire can be implemented in short time and with little cost. In addition, unlike the traditional BRFSS’ survey methodology, using the Quantextual.co platform; we will be able to collect passive data from each participant’s wearable device (which includes the number of hours of sleeps, number of steps). Researchers will have access to wide range of data of different frequencies (high frequency biosensor data through wearable devices and low frequency data from administering self-reported outcome surveys). Most importantly, researchers can investigate the challenges and opportunities of conducting surveys through mobile apps.**

1. Describe the process you will use to link the data from the different sources you’ve identified. Include a description of feasibility and any considerations that will be made to ensure the privacy, security and confidentiality of the data and data subjects throughout this process. (750-word limit)

**As we mentioned earlier, we are planning to execute the project using two different approaches. The first approach is based on probability sampling and the other one is based on recruiting a convenient sample. For the first approach, we envision to include two questions into the BRFSS survey that pertain to whether an interviewee has a smart phone or not and whether he is willing to participate in BRFSS survey questionnaire implemented through a mobile application downloaded into their smart phone. We will be able to link BRFSS participants to their previous survey responses, and be able to follow them over time using the new concept.**

**Regardless of the recruiting approach, participants who decide to join the study will use double opt-in registration, where the participants create a profile using their email and once their email is validated, they can log in and complete their profile. Once these participants complete their registration, every participant will be assigned a random code that is associated with their data instead of their name. Every participant coded data will be added to the data of other study participants, and all the identified data items will be removed from the data. This will guarantee maximum security and privacy before making the data available for the public. Finally, every participants will have a unique account that they can access and use to review their own data, and they can decide whether to share their data or not**.

1. Describe how the linked data set(s) or individual data source(s) will be used to develop values for your proposed set of metrics in sleep, sedentary behaviors, nutrition, and/or physical activity. (500-word limit)

**This project combines two sort of data. Data collected through biosensors, and data reported by the participants. These two data are directly linked to sleep, sedentary behaviors, nutrition, and physical activity. For instance, using wearable devices, we can get the number of steps, hours of sleep, and nutrition every day and continuously. In addition, we can send survey questionnaires at a given frequencies to assess these participants physical health and nutrition. By doing so, researchers can investigate associations between self-reported data and wearable devices data. For instance, a review study found likely premature definitive conclusions about the efficacy of exercise as a nonpharmacological treatment for poor sleep,[[6]](#footnote-5) and the relationships among different type of physical activity and sleep are not well understood. Physical activity can be used as a treatment for poor sleep, but also research has not fully addressed the need for more sleep as recovery following several days of extended vigorous intensity exercise. With wearable devices, it is now possible to measure 24-hour cycle of physical activity and sleep, which can be used to answer a number of these specific research questions that cannot be addressed with current measurement methods.[[7]](#footnote-6) Furthermore, the data can be used to investigate the relationship between sedentary behavior and physical activity and between sedentary behavior and nutrition. A systematic review have investigated the relationship between physical activity, sedentary behavior, and nutrition, and found that unhealthy dietary is associated with sedentary behavior. [[8]](#footnote-7) The systematic review study reported also that most studies did not control for confounding factors such as diet or physical activity. Smart phones offer an opportunity to develop strong prospective studies by collecting daily passive data on physical activity on the subject participants and their diet. Researchers can thus control for any potential confounding factors since, the app collects both passive and active data, which will help researchers investigate the strength of the association among these outcome variables (e.g., nutrition, physical activity, and sedentary behavior).**

1. Describe the representativeness of your data set for public health surveillance (e.g., to what population groups or sub-groups can you meaningfully extrapolate the results of your data set?). How amenable will this data set be to disaggregation by age, gender, education, geography, or other demographic characteristics? (750-word limit)

**We are targeting to recruit individual who are 18 years and older from different socioeconomic strata (SES). In addition, we are planning to use different sampling approach. The golden standard approach (e.g., probability sampling), and nonprobability sampling. By doing so, we can assess how the public health surveillance compares using these two different methodologies. Besides the SES variables, we will be collecting a wide range of data from participants’ sensors available through their wearables. We can, for instance, collect accelerometer, heart rate, GPS, Gyroscope, compass, microphone, ambient light, barometer etc.…[[9]](#footnote-8) These data can be used to compute the prevalence rate of physical activity, sleep, and nutrition that are similar to those estimated using BRFSS survey questionnaire.**

1. How useful will your data set be for public health surveillance, how significant/relevant and generalizable are the results that you expect to obtain? (500-word limit)

**The wide spread of the internet has increased the number of public and commercial initiatives to recruit individuals to participate in scientific research. These initiatives collect personal medical history, physical traits and measurements , ethnicity/ancestry, lifestyle and environmental exposures, and to donate biological material, generally saliva or blood, for DNA analysis. For instance, PatientsLikeMe is an online community where patients can share information on symptoms, health state, and treatments to learn from each other’s experiences. 23andMe sells personal genome tests to individuals who want to learn their genetic risks of common diseases, carrier status of rare diseases, response to drug treatment, and ancestry. Data are collected predominantly through self-report online questionnaires and these initiatives offer the opportunity to make data accessible for the public.**

**Unlike existing initiatives to collect only patient reported outcome, this study (hereafter Quantextual.co ) combines several data sources within one system. It offers a new way to collaborate in designing and organizing surveys and research studies. Quantextual.co is a public-private-academic non-profit organization to be the steward of active and passive volunteered health data of sufficient size and geographic granularity for the study of public health surveillance.**

**CDC has funded similar concept research study through The equipped, enabled, engaged and empowered (e4) to Health and Willingness to Share Health Data Study, which is a capacity building cooperative grant to the Association of University Centers on Disabilities (AUCD). This capacity building collaborative grant is in its third year with the Health Management and Informatics (University of Missouri, School of Medicine) and the Interactive Autism Network (Kennedy Krieger Institute, Johns Hopkins University) as collaborators. Health-ePeople.com, a Public Good Venture has been contracted to coordinate and manage the study.**

1. Will the proposed project’s data and data sets contain information of relevance to other areas of public health surveillance (e.g., chronic or infectious disease)? If yes, please specify and describe any additional work that would be required in order to expand applicability. (500-word limit)

**We propose using the Quantextual.co platform for collecting and disseminating a wide range of consumer and clinical data. The platform equips participants to collect medical record data from their patient portals, and genomic data from 23andMe. The Quantextual.co platform is already functioning and there is no need for additional work except applying for human subject IRB.**

1. Please describe a 3.5-month plan to develop a working prototype during the second phase of this challenge. This should include:
2. Details on how you will gain access to and link data from the source(s) you’ve identified.

**Quantextual.co platform collects data from participants’ wearable devices and survey questionnaires. The platform uses the OAuth 2.0 specification, which defines a delegation protocol that is useful for conveying authorization decisions across a network of web-enabled applications and APIs. Doing so, we are able to collect data from different wearable devices used by the participants.**

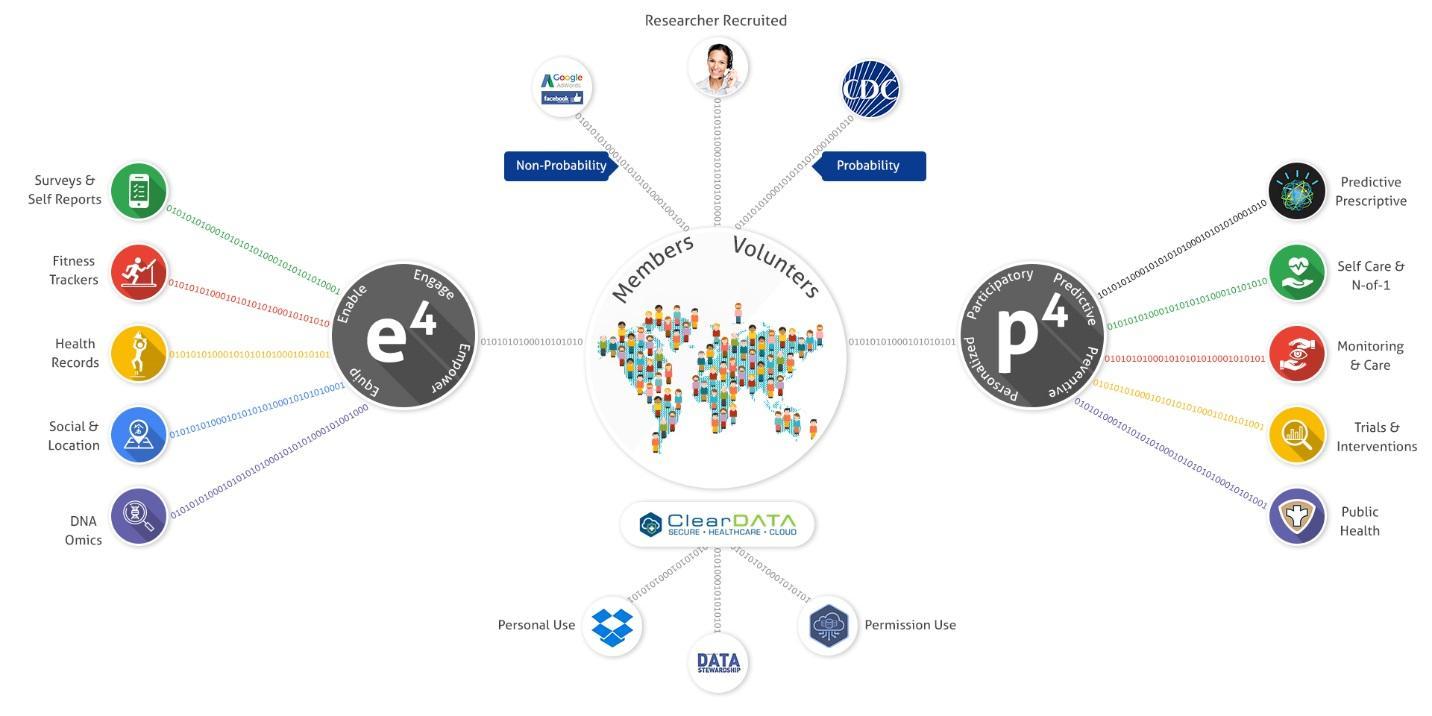
1. Approaches/strategies that will be taken to ensure privacy/confidentiality of data before and after linkage.

**We use double opt-in approach to register eligible participants into the applications. By doing so, we guarantee maximum privacy and security for the participants data. In addition, all data collected are saved into Cleardata.com cloud server, which is a HIPAA compliant cloud hosting, backup, disaster recovery and information security services exclusively for the healthcare industry.**

1. Your approach to comparing results from your prototype to that generated from existing public health surveillance programs.

**Our concept will allow us to compare the current BRFSS estimates with the proposed methodology. The two well-known competing approaches to conduct surveys are probability sampling and non-probability sampling technique. One of the shortcomings of both methodologies is the potential issues of coverage error. This error is more pronounced in non-probability sampling. Our methodology of recruiting participants from probability sampling is a technique to assess the reliability of the non-probability sample estimates compared to those of probability sample. This type of methodology has been applied in building The LISS Panel in Holland. Where panel members were randomly recruited using their address and an invitation letter were sent to them, and as well landline phone numbers were used to recruit those whose phone number were matched to their address.**

1. A description of the format your prototype will take (e.g., visualization, online data tool, etc.)



*Figure 1: Quantextual.co platform (source; Quantextual.co)*

**The figure above summarizes how Quantextual.co platform functions. At the center of the platform are the participating members or volunteers. These participants can share different data with researchers. All the data are saved in secure HIPAA compliant cloud server. The application provides different analytical tools to summarize the data, which can help members to monitor their care. At the same provides researchers power to conduct citizens research.**

1. Costs you expect to incur during this prototyping phase

**The licensing cost for implementing this project is about $35,000.This cost covers platform licensing ($20,000) and participants’ reimbursement for their participation ($15,000).**

1. Significance and Relevance Summary

|  |
| --- |
| In 200 words or less, provide a brief summary of your project using language that is easily understood by the general public. Note: this description will be shared with a broad audience and should not include any information you would not want shared widely.  **The objective of this project is to investigate the difference of using two different recruitment methodologies (individual that are randomly selected to participate versus individuals that have been recruited using social media) and apply them to BRFSS survey questionnaire. At the same, augment this data with data from wearable devices to assess the relationship between physical activity, sleep, and nutrition.** |

Public reporting burden of this collection of information is estimated to average 60 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to CDC/ATSDR Reports Clearance Officer; 1600 Clifton Road NE, MS D-74, Atlanta, Georgia 30333; ATTN: PRA (0990-0390).

1. Moderate-to-vigorous physical activity (MVPA) is any activity with an energy expenditure >3 metabolic equivalents [↑](#footnote-ref-0)
2. Sedentary behavior is any waking activity characterized by an energy expenditure ≤ 1.5 metabolic equivalents and a sitting or reclining posture [↑](#footnote-ref-1)
3. Source: <http://www.pewinternet.org/fact-sheet/mobile/> [↑](#footnote-ref-2)
4. Source: <http://www.pewinternet.org/2015/11/10/the-majority-of-smartphone-owners-download-apps/> [↑](#footnote-ref-3)
5. Source: <https://ondeviceresearch.com/blog/optimal-length-mobile-survey> [↑](#footnote-ref-4)
6. Buman, M.P. & King, A.C., 2010. Exercise as a Treatment to Enhance Sleep. American Journal of Lifestyle Medicine, 4(6), pp.500–514. Available at: http://journals.sagepub.com/doi/10.1177/1559827610375532 [Accessed July 25, 2017]. [↑](#footnote-ref-5)
7. Rosenberger, M.E. et al., 2016. 24 Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices HHS Public Access. Med Sci Sports Exerc, 48(3), pp.457–465. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4760880/pdf/nihms724316.pdf [Accessed July 25, 2017]. [↑](#footnote-ref-6)
8. Pearson, Natalie, and Stuart JH Biddle. "Sedentary behavior and dietary intake in children, adolescents, and adults: a systematic review." American journal of preventive medicine 41.2 (2011): 178-188. [↑](#footnote-ref-7)
9. De Arriba-Pérez F, Caeiro-Rodríguez M, Santos-Gago JM. Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios. Reindl LM, ed. Sensors (Basel, Switzerland). 2016;16(9):1538. doi:10.3390/s16091538. [↑](#footnote-ref-8)