

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

Phase 1 Submission Template

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.

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:

Data2Policy

Team Lead:

Beth Ann Fiedler, PhD

City/Province:

Jacksonville, FL

The Healthy Behavior Data Challenge

Phase 1 Submission Template

E-mail:

data2policy@gmail.com

Phone Number:

904 219-1739; 321 277-7393

Subject-matter/domain expertise:

Health Generalist, Health Informatics; Public Policy

Team Member #1

E-mail

Subject-matter/domain expertise

Jacqueline Ortiz Baerga

jortizbaerga@hotmail.com

Population Health

Are all team members residents of the United States?

Yes

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

Organization Name:

Website:

Type of

Organization:

DNA

ucf.academia.edu/BethFiedler
https://www.researchgate.net/profile/Beth_Fiedler

DNA

3) How did you find out about this challenge?

Email subscription to CDC notices to alternate email account ba_fiedler64@hotmail.com.

4) Submission Overview

Project Title: Extending Public Health Surveillance Reporting Through Digital Data Collection of Behavior Patterns and Existing Conditions

Project Overview

The Healthy Behavior Data Challenge

Phase 1 Submission Template

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?

The circular momentum of sleep, physical activity, nutrition and sedentary behavior are important to public health surveillance because disruption in any of these factors breaks the cycle of physical activity-> rest -> strength and healing of the body -> nutrition feeds the body to prepare for physical activity. A break in the repetitive pattern contributes to the problem of obesity in the United States and leads to a variety of preventable health conditions such as diabetes. Key aspects of reporting include whether the survey respondent understands food serving size, ability to read labels (food or medication), their proper weight according to their height, age and gender, the amount of weekly exercise, the amount of daily sedentary behavior, and consumption habits (e.g., food, alcohol, non-prescription drug use). There is an abundance of literature to support this foundation.

Additional information (e.g., employment status, access to affordable insurance, method of transportation) will also be taken into consideration to discern the ability for individuals to obtain healthy and affordable food; health insurance including meeting monthly contributions, co-payments, and prescription medications; and to determine if individuals are subject to additional harmful pollutants using public transportation. Noteworthy is that while other agencies collect data through mapping (i.e., distance to local park or green area), "identification and location of health inequalities and access to recreational spaces" (Fiedler and Cook, 2018, *in press*) is only the first step because this information is only of value if they are collected in relation to current health status or some other measurable outcome.

An alternative solution of incorporating digital device information includes permitting survey respondents to voluntarily link and/or upload data according to fields established in the device. Individuals willing to provide access to wearable technologies should be encouraged to participate in the survey to collect important information that digital devices do not contain. Thus, digital data will be appended to required survey fields affording the opportunity to gain metrics on those who use digital tracking options between and among other digital tracking methods and those who do not. Additional information from other agencies can also deploy this method (i.e., air pollution by zip code). As the learning curve is established for both the individual respondent, digital product manufacturers, and end users of the data, sophisticated methods of mapping will eventually permit data fields to converge 1) permitting insight into perceived activity and actual, 2) maintaining general information for comparison against those who do not utilize digital tracking, and 3) retaining digital user information in the overall survey.

The Healthy Behavior Data Challenge

Phase 1 Submission Template

Respondents who track both physical activity and caloric intake on smartphone or tablets, such as www.myfitnesspal.com and/or <https://supertracker.usda.gov/>, can be encouraged to participate in the voluntary collection process to add depth to information collected by the survey. Social media and the opportunity to engage current users on the websites will offer further validity to the data as a larger portion of the population will be represented in the collection process.

- 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?

Utilization of external data sources will require a system wide convergence of metrics, the convergence of data types and data fields potentially resolved using data mapping, as well as the incorporation of agreed upon outcomes measures that would be developed over time. Therefore, source data will be collected using a digital invitation via text or email for patients to complete a survey at a specific URL with a list of questions prior to non-emergency appointments and after appointments as a reminder. Patients receiving unscheduled medical services, such as emergency services, will receive a request after services are received with a follow-up reminder about 1 week after date of service. Since medical professionals are providing International Classification of Diseases (ICD) diagnosis information on visit reports, survey respondents will also be asked to provide this information on the survey (e.g., 250 diabetes mellitus). The initial invitation will notify patients that they will need the visit report with diagnosis codes to accurately complete the survey.

These measures are important because the personal patterns of behavior embedded in the questions in conjunction with reported diagnosis will provide a general overview of patient physical, mental, and social health. The unique way in which questions are worded will also provide alternative measurements of data providing supplemental information on public health surveillance. For example, while many questions may be written with dichotomous responses (yes/no), the questions will also be written in such a way as to allow transformation to ordinal variables demonstrated in a 3-point Likert Scale. The scale represents 1) Agree indicated by Yes responses, 2) Neither Agree or Disagree (2) is indicated by a option to not answer the question, or 3) Disagree (3) is indicated by a No response.

Transformed and initial data types, (e.g., dichotomous and ordinal), can undergo a variety of advanced statistical analysis and provide the opportunity to retain survey responses for alternate evaluation that otherwise would be cleansed due to missing data. A pop-up field

The Healthy Behavior Data Challenge

Phase 1 Submission Template

where respondents select known diagnosis associated with ICD codes (e.g., diabetes, asthma, etc.) can be added to the survey if respondents do not have actual codes. Further, the coding can also be categorized as non-communicable and communicable diseases where ordinal values can be deployed using the same 3-point Likert Scale. In this case the question would ask, "Do you suffer from at least one communicable disease." Then, "Do you suffer from two or more communicable diseases?" Same for non-communicable diseases. If yes or agree=1, if no answer=2, if no or disagree=3. Utilizing a variety of carefully written question will allow easy transition to other data types that may be more useful in certain types of analysis and contribute to policy decision-making process. Finally, adding the categorical classification of (non)communicable disease will provide the capacity to analyze the categorical demographic data that identifies the study population.

- How do you see your concept improving on current public health surveillance in the areas of sleep, physical activity, nutrition, and sedentary behaviors?

I foresee my proposed method of public health surveillance to overcome the current limitations on study populations based on three primary factors. First, the decreasing number of participants who can be contacted through random land lines; second, the increasing number of potential survey participants who have access to digital modes of communication; and third, the optimization of existing data structures to add value to analysis leading to informed decision-making. By considering the public transition to digital modes of communication, future study populations will not be skewed towards those who are home due to disabilities, lack of employment, or access to modern communication methods. Thus, the method has a greater potential to represent an improved sampling of participants that is more likely to be representative of the general population. Findings, in turn, present the opportunity to address problems from a clinical and non-clinical perspective leading to improved methods of treatment that better represent the actual, not skewed, incidence of certain behaviors and other diagnosis that may have had a higher representation in previously collected samples. Ultimately the transition to this method of data collection coupled with questions designed to elicit various analysis will be a strong foundation for social service development and public administration. The inclusive nature of the method will also account for public health conditions to advise the population regarding specific behavioral patterns related to locality, urban/rural designation, ethnicity and other important factors in the tradition of the CDC. But, offer data fields more amenable to analysis thus providing health informatics to policymakers based on evidence-based analysis methods and research design.

The Healthy Behavior Data Challenge

Phase 1 Submission Template

- 5) Indicators to be measured (the indicators listed below are not comprehensive and innovators are recommended to include other relevant indicators)

A complete list of complimentary questions will be provided in the event this application moves to Phase II. Some sample questions have been embedded in response to specific questions.

A) Physical Activity

- Amount of MVPA¹ 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
- Do you exercise for at least twenty minutes three times a week?
If yes
How many calories do you estimate that you burn?
- Do you exercise for at least twenty minutes less than three times a week?
If yes
How many calories do you estimate you burn?
- Do you exercise for at least twenty minutes more than three times a week?
If yes
How many calories do you estimate you burn?
- Have you recently lost 10% of your body weight? (Example, original weight 190, lost 19 pounds and now weigh 171)
If yes, have you been able to maintain the new weight for more than six months?
- Have you recently lost 20% of your body weight?
If yes, have you been able to maintain the new weight for more than six months?

¹ Moderate-to-vigorous physical activity (MVPA) is any activity with an energy expenditure >3 metabolic equivalents

The Healthy Behavior Data Challenge

Phase 1 Submission Template

- Have you recently lost more than 20% of your body weight?
If yes, have you been able to maintain the new weight for more than six months?
- Have you recently gained 10% of your body weight? (Example, original weight 120, gained 12 pounds and now weigh 132)
If yes, due to changes in health? injury? depression? Unemployment? _____
- Have you recently gained 20% of your body weight?
If yes, due to changes in health? injury? depression? Unemployment? _____
- Have you recently gained more than 20% of your body weight?
If yes, due to changes in health? injury? depression? Unemployment? _____

B) Sedentary Behavior²

- ~~• 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
- Do you sit about eight hours per day?
- Do you sit less than eight hours per day?
- Do you sit more than eight hours per day?
- Do you spend less than two hours recreation time on a computer/screen including watching TV, videos, playing computer games, emailing or using the internet per day?
- Do you spend about two hours recreation time on a computer/screen including watching TV, videos, playing computer games, emailing or using the internet per day?
- Do you spend more than two hours recreation time on a computer/screen including watching TV, videos, playing computer games, emailing or using the internet per day?
-

C) 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

² Sedentary behavior is any waking activity characterized by an energy expenditure ≤ 1.5 metabolic equivalents and a sitting or reclining posture

The Healthy Behavior Data Challenge

Phase 1 Submission Template

- Daytime sleepiness
- Other indicators
- Do you sleep less than 8 hours per day?
- Do you sleep more than 8 hours per day?
- Do you sleep about 8 hours a day?
-
-

D) 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)
- Can you digest milk products?
 - If yes
 - a) If you can digest milk products, do you eat about one serving per day?
 - b) If you can digest milk products, do you eat less than about two services per day?
 - c) If you can digest milk products, do you eat three or more servings per day?
 - If no
 - d) If you cannot digest milk products, do you eat about three servings per day of a milk product substitute?
 - e) If you cannot digest milk products, do you eat less than three servings per day of a milk product substitute?
 - f) If you cannot digest milk products, do you eat more than three servings per day of a milk product substitute?
- Do you eat about three servings of fresh or frozen vegetables per day?
- Do you eat less than three servings of fresh or frozen vegetables per day?
- Do you eat more than three servings of fresh or frozen vegetables per day?
- Do you eat about 2-3 servings of red meat per week?
- Do you eat less than 2-3 servings of red meat per week?
- Do you eat more than 2-3 servings of red meat per week?
- Do you include fish about 2-3 times per week in your diet?
- Do you include fish less than 2-3 times per week in your diet?
- Do you include fish more than 2-3 times per week in your diet?
-

The Healthy Behavior Data Challenge

Phase 1 Submission Template

E) Social

- Are you aware of the weight that the medical profession considers to be most healthy for your age, gender, and height?
- Are you within 10 pounds of that prescribed weight?
- Are you within 20 pounds of that prescribed weight?
- Are you more than 20 pounds over that prescribed weight?
- Do you feel you understand the concept of serving size?
- Do you feel that labels on serving sizes are easy to understand?
- Do you feel that the concept of serving size is not used when preparing meals?
- Do you have access to health insurance?
- Can you afford monthly premiums and copayments?
- Are you aware of opportunities to attend community development and planning meetings?
- In the past year, have you attended at least one community development or planning meeting?
- In the past year, have you attended at least two community development or planning meetings?
- In the past year, have you attended three or more community development or planning meetings?
- Are you interested in hearing about projects in your local community?
- Are you interested in hearing about opportunities to volunteer?

F) Mental

- Do you overeat about once a week?
- Do you overeat about twice a week?
- Do you overeat three or more times a week?
- Do you undereat about once a week?
- Do you undereat about twice a week?
- Do you undereat three or more times a week?
- Do you feel that you can afford the type of nutritious food that your body needs to stay healthy?
- Do you feel that the cost of nutritious food is a big factor in your food selection of less nutritious items?
- Do you drink less than one serving of alcohol per week?
- Do you drink about one serving of alcohol per week?
- Do you drink about two servings of alcohol per week?
- Do you drink three or more servings of alcohol per week?
- Do you drink less than one serving of alcohol per week?
- Do you participate in the use of non-prescription or other recreational drug use?
If yes
 - a) Do you participate in the use of non-prescription or other recreational drug use about once a week?
 - b) Do you participate in the use of non-prescription or other recreational drug use about twice a week?

The Healthy Behavior Data Challenge

Phase 1 Submission Template

- c) c) Do you participate in the use of non-prescription drugs or other recreational drug use three or more three times a week?
- Do you stress eat?
If yes
Do you stress eat about once a week?
Do you stress eat about twice a week?
Do you stress eat three or more times a week?
What triggers stress eating? Work, family, school, _____.
- Do you have access to an outdoor area to exercise?
Do you restrict your exercise to indoors?
If yes
Is this restriction based on poor air quality in your area?
Is this restriction based on allergies or other physical respiratory problem such as asthma?
Do you feel safe in that public area where you could exercise?
- Do you track your daily caloric input?
If yes
- What is the average number of calories you consume per day?
- Do you track your daily steps?
If yes
What is the average number of steps you take per day?
Are you tracking your daily steps for weight loss?
- Do you track your caloric intake?
If yes
What is the average calories per day?
Are your tracking your calories for weight loss?
- Are you willing to submit digital information about your caloric intake?
- Are you willing to submit digital information about your exercise?
- Would you be interested in hearing about free mobile phone applications to track your calories?
- Would you be interested in hearing about free mobile phone applications to track your physical activity
- Do you have a reliable source of clean water?
- Do you have a reliable source of electricity?

G) Outcomes

- Do you suffer from a non-communicable disease(s)? Yes/No Enter ICD Code/Describe
- Are you diagnosed with one noncommunicable disease?
- Are you diagnosed with two non-communicable diseases?
- Are you diagnosed with three or more non-communicable diseases?
- Do you suffer from communicable disease(s)? Yes/No Enter ICD Code/Describe
- Are you diagnosed with one communicable disease?
- Are you diagnosed with two communicable diseases?
- Are you diagnosed with three or more communicable diseases?

6) Summary of proposed data source(s) (complete applicable sections)

	Repeat		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)					
1a	Individual Level	Self-reported	Accumulation of data based on survey response via dedicated website	Minimal-dedicated website portal to house data, open access after annual completion and data cleansing	Composite of collection of surveys beginning each January 1	The online survey will cover all four functional areas in a series of questions.	None, the proposed source would be newly created, using existing CDC reported data fields where possible.
1b	Individual Level	Initial Self-Reported Digital Access; Link can be automated after initial upload	Initial downloads will use existing fields tracked in digital devices; development of	Minimal-eventual cost of data mapping to ensure translation of disparate data fields and operating systems from various digital devices	Permission to access averages of caloric intake and physical activity on a quarterly basis along with weight changes and/or changes in medication or diagnosis.	Digital information will focus on physical activity and nutrition and appended to survey responses.	This method will not utilize direct downloads of massive information from digital device sources because the data will not be associated with key predictive measures.

The Healthy Behavior Data Challenge

Phase 1 Submission Template

	Repeat		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)					
2	Provider Level	Healthcare provider assessed diagnosis	Provided to patient at each visit.	Minimal, if any, to Provider since visit information is already provided to patients	None	Provides disease indications that could be used as an outcome index or composite of overall patient health	ICD information is currently collected by the CDC from providers to inform national public health.
3	Healthcare Scheduling Organizations	Notification System	Add invitation to survey through existing portal.	Minimal, if any, simple modification GUI to existing structure	Annual changes to dates to keep GUI up to date	All	AppointmentQuest LLC in Colorado
4	CDC	Compiled website for survey	Size of data sample would require statistical software, such as SPSS or SAS, to handle large quantities of information	Minimal, use existing personnel to create and maintain website portal; cost of enterprise SPSS or SAS software	Self-reported information would require daily back-up and annual offload from data collected between Jan 1 and Dec 31	Utilized individual level responses supported by provider level diagnosis data to analyze four behavior indicators and supplemental digital health device information	Newly created to extend existing data fields making long-term analysis of trends still possible while incorporating new metrics.

- 7) 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 provision to link other data sources in the short-term is impractical. This position is based on the current data practice of collecting data that represent various indicators using disparate variable types and/or do not use outcome measures limiting the capacity to apply health informatics techniques that inform policymakers regarding expenditure decisions based on a causal link between behaviors and health outcomes. A case in point, a collection of data indicating access to a park within a half mile without collecting data as to whether that person uses the facility or feels safe in the park is inconsequential information. Collecting the same data in the content of safety and usage in conjunction with their current health status and personal characteristics provides more meaningful information that can be supplemented with information from many other agencies (e.g., Agriculture, Defense, Commerce, Health and Human Services, Transportation, Environmental Protection Agency, etc.). Having behavioral predictor information compared to actual health using methods, such as health economics and outcomes research (HEOR), is an optimal and reliable method from which to discern patterns, permit CDC to assess national health, inform clinicians to develop viable treatment protocols, and for collected data to support improvements in public health.

However, I believe that adding health data from data collected as part of physical education for K-12 students could benefit public health analysis. Schools may be categorized in the same way as a provider through school nursing programs that link confidential information with parental consent to the current CDC reporting structure. There is a dismal amount of information on the under-age population due to the constraint of informed consent. Otherwise, parents may be asked to complete the survey for children under age 18 so that a more accurate representation of this public health sector can be obtained.

- 8) 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)

Not applicable.

The data will be collected on a secure website maintained by the CDC. Information will be directly uploaded to data fields based on the responses from survey participants.

- 9) 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)

The Healthy Behavior Data Challenge

Phase 1 Submission Template

Individuals responding to the survey will act as primary data source to generate frequency counts, ratio, and percentage values to also compare against other tracked agency data. These data will then be analyzed using a variety of statistical analysis methods to accurately assess behavioral patterns in sleep/sedentary, nutrition and physical activity behavior patterns. The method in which the survey questions are presented will provide several analysis options outputting total counts, ratio values such as ethnicity or age/(non)communicable diseases, and reporting statistically significant data reflecting the contribution of certain behaviors to health status to determine priority for healthcare budgeting, programs, and community services.

- 10) 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)

General information will be gathered in the form of control variables on age, level of education, geographic location (urban/rural; state), income, race, and ethnicity. This information will be used to account for these differences in the interpretation of the results and final analysis. While zip codes can be collected on urban locations, the same cannot be said for rural locations due to the potential for identifying survey respondents. Thus, the question of zip code will be controlled in the designation of urban/rural and only open to those who designate 'rural' in their response. Gender status will be two-fold. First, the study will request gender at birth indicating male or female. Another question will request gender identity and include options such as male, female, bi-gender, transgender, transsexual, trans male, trans female, gender non-conforming/unspecific, and transvestite. Further, other survey population characteristics can be gathered to inform responses. For examples, the following questions can reveal certain mental health, social limitations, or deficiency in basic education that will supplement analysis.

Sample questions could include:

- 1) Do you feel you understand the concept of serving size?
- 2) Do you feel that labels on serving sizes are easy to understand?
- 3) Do you feel that the concept of serving size is not used when you are preparing meals for yourself or your family?
- 4) Do you know the recommended number of calories you should be eating daily to maintain normal weight?
- 5) Do you know the recommended number of calories you should be eating daily to lose weight?
- 6) Do you sometimes overeat (snacking between meals, eating more than serving size portions, eating more than your recommended daily calories)?
- 7) Do you overeat at least once a week?

The Healthy Behavior Data Challenge

Phase 1 Submission Template

- 8) Do you overeat at least twice a week?
- 9) Do you overeat three or more times a week?

These questions represent the strategic change to adaptable data measurement scales and metrics. For example, questions 1-6, and 10 will be listed as dichotomous (Yes/No) questions while questions 7-9 add more depth to question 6. Ordinal questions represented in questions 7-9 can also undergo advanced statistical analysis and reliability testing of survey questions.

- 11) 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 response potential within each state is high leading to the probability of randomly selecting datasets for comparison to determine the statistical generalizability of each data analysis result. Internal validity will be supported using peer-reviewed literature identifying important constructs and theoretical premises for survey questions. Further, survey questions converted to ordinal questions will also undergo a reliability analysis using the statistical method of Cronbach Alpha. The dataset will also increase the proportion of representation from children under the age of 18 using data acquired from the annual physical requirements.

This data collection method will expand on the type of analysis available to researchers and therefore have a greater capacity to inform policymakers. For example, the data sets will include data fields consistent with long-term data collection that has been performed by the CDC. The addition of data fields will support a wider variety of analysis and allow ratio comparisons that could be customized to localities serving the information needs of community planners in cities. Further, data can be utilized in a multi-tiered format to inform regions, states, and national agencies of public health problems and potential agency interventions for social services to leverage funding across these various levels of government.

- 12) 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)

Reports from provider visits will indicate the ICD9 or 10 designations so that behavioral health predictors in the questions can be directly measured against actual diagnosis. I have previously suggested data sharing methods to optimize the use of existing ICD codes published in literature, presented in national conferences and to legislative liaisons. Further, I have encapsulated the important role of managing medical devices within the context of healthcare and bring forth the notion of global data convergence based on the intentional design to incorporate data from devices within the medical community. Finally, identifying environmental hazards will also become increasingly relevant to the final composition of the CDC database.

The Healthy Behavior Data Challenge

Phase 1 Submission Template

- Fiedler, B.A. (Ed.) (2018). Translating National Policy to Improve Environmental Conditions Impacting Public Health Through Community Planning. Springer International Publishing, *in press*.
- Fiedler, B.A. (Ed.) (2016). Managing Medical Devices within a Regulatory Framework. Elsevier Publishing.
- Fiedler, B.A. & Yadin, D. Reframing product life cycle for medical devices, pp.3-16.
 - Fiedler, B.A. & Ferguson, M. Overview of medical device clinical trials, pp. 17-32.
 - Fiedler, B.A. Review regulatory guidelines by device classification type, pp. 33-50.
 - Fiedler, B.A. Manufacturing/distribution considerations, pp. 51-69.
 - Fiedler, B.A. Defining and meeting regulatory challenges in clinical engineering, pp. 73-90.
 - Fiedler, B.A. Roles of biocompatibility, pp. 91-108.
 - Fiedler, B.A. & Farid, A. Risk management, pp. 109-128.
 - Fiedler, B.A. Sterility and reusability, pp. 129-144.
 - Fiedler, B.A. European Union national differences and potential impact on CE Marking, pp. 147-168.
 - Fiedler, B.A. Understanding the transitioning regulatory EU market, pp. 169-176.
 - Fiedler, B.A. Evaluating new medical device purchases, pp. 179-201.
 - Fiedler, B.A. & Greathouse, J. Evaluating reimbursement strategies in the United States, pp. 203-224.
 - Fiedler, B.A. Clinical and biomedical engineering evidence strategy, pp. 245-261.
 - Fiedler, B.A. Device failure tracking and response to manufacturing recalls, pp. 263-275.
 - Fiedler, B.A. The future of health technology management, pp.299-314.
 - Fiedler, B.A. Challenges of new technology: Securing medical devices and their software for HIPPA compliance, pp. 315-329.
 - Fiedler, B.A. Managing smartphone and tablet applications, Chapter 19, pp. 331-342.
- Fiedler, B.A. (2015). Constructing legal authority to facilitate multi-level interagency health data sharing in the United States. *International Journal of Pharmaceutical and Healthcare Marketing*, 9(2), 175-194.
- Kula, S., & Fiedler, B.A. (2014). Improving health literacy through governance tools. *Journal of Public Administration and Governance*, 4(1), 45-61.
- Fiedler, B.A. (2014). Constructing Legal Authority to Facilitate Multi-Level Interagency Health Data Sharing in the United States. American Society of Public Administration Conference in Washington, DC, March 14-18. *ASPA 2014 Founders' Fellow*
- Fiedler, B.A., Goldiez, B., Sotomayer, T., Estes, J., & Washburn, D.A. (2011). Developing Data Exchange Protocol Standards for Healthcare Simulators. John Hopkins School of Medicine, MedBiquitous Annual Conference in Baltimore, MD, May 9-11

13) Please describe a 3.5-month plan to develop a working prototype during the second phase of this challenge. This should include:

- 1) Details on how you will gain access to and link data from the source(s) you've identified.

Weeks 1-2: Dedicated to collaboration with CDC and fundamental research; initialize requirements for website portal and preparedness for actual instrument in week 11. Obtain volunteer healthcare organization with digital capacity to discuss ability to carry CDC invitation to survey participants through digital scheduling system in week 11. Objective to reach consensus on direction of survey questions and span of analysis.

Weeks 3-4: Dedicated to collaboration with manufacturers of digital devices to ascertain data fields currently collected and potential to easily upload to CDC portal. Objective to determine capacity to append data to survey responses with header fields intact without disrupting survey responses.

The Healthy Behavior Data Challenge

Phase 1 Submission Template

Weeks 5-10: Dedicated to application of fundamental research gathered from the CDC and device manufacturer research coupled with survey design analysis to output the survey instrument. Objective to perform local pilot test on the survey instrument to make any minor adjustments prior to creating the online version.

Weeks 11-12: Dedicated to CDC website survey portal. Objective to pilot test survey with 'live' answers with volunteer representatives from CDC, fitness tracking device manufacturers, and the healthcare site with patients to reveal any communication flaws. Objective: validate general design and receive responses to determine validity of survey questions; verify accumulation of responses; and verify ability for system communication from CDC to providers.

Weeks 13-14: Dedicated to testing various scenarios of notification to providers in conjunction with their scheduling system; portal access; data accumulation, etc.

Upon invitation to move on to Phase II, I will establish a development plan with CDC to house the survey on a dedicated website accessible to invited participants through a URL link. Responses will be accessible to CDC personnel utilizing established security protocols and accessible to the public once annual data has been aggregated and cleansed. While respondents will not have permission to access or view incomplete raw data, they will be given the option to review their own responses, if desired, during the time of entry. Since primary survey invitations will be generated for those scheduling appointments for some form of medical treatment and diagnosis during the year, the site will be accessed only through personal invitations linked to scheduled visits or as a follow-up to emergency care, thereby providing some representation of services provided to the uninsured. Existing communication systems within the CDC will notify providers of the new method and request their cooperation in the interest of scientific and clinical discovery. Healthcare facilities serving a variety of representative population, such as Veteran's health, will be asked to extend the information regarding survey participation when requesting pre-registration or actual registration information. Further, notice of the option to participate can be embedded in URLs in conjunction with other federal agencies, <https://supertracker.usda.gov/>, or organizations that offer fitness tracking portals such as www.myfitnesspal.com who may be willing to collaborate by drawing attention to the value of their services. Inviting organizations such as Fitbit to review the value of their existing data fields will allow the CDC to positively influence manufacturers of the added value to data collection if data fields converge. Until then, utilization of self-reporting and appending external sources of data with their existing data fields will provide an immediate method to obtain information for assessment, level of current contribution to research, and offer potential changes to data structure leading to reporting standards. Finally, this approach endeavors to collect data using terminology that will be comfortable to a greater number of participants. Therefore, questions will ask information using easily recognizable vocabulary.

Healthcare organizations that specialize in scheduling, such as AppointmentQuest LLC in Colorado, can be utilized in the pilot study as a survey invitation mechanism to supplement tracking data through local provider scheduling systems. The addition of these organizations will offset local providers who do not have their own system in place.

The Healthy Behavior Data Challenge

Phase 1 Submission Template

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

The survey host website will be secured using general HIPPA protocols and the general premise of secure data utilized in scientific investigation. Surveys will not collect personal information such as name, patient identifiers, address, or other items such as social security numbers that could be tracked back to the study participant.

- 3) Your approach to comparing results from your prototype to that generated from existing public health surveillance programs

The new approach will still collect frequency data to establish a link to previously gathered information displayed in database format such as Microsoft Excel or Access, comma delimited, and similar. But advanced features, such as transforming data to various data types, will be incorporated into the portal for authorized members to retrieve data, create pertinent pivot tables by selecting relevant combinations of data field analysis, and later for the public, researchers, or other agencies to access complete data.

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

The data can be linked to an option to perform pivot tables for agencies and other end users to extract data in a relevant combination of data field analysis. I propose providing direct statistical interface to end users through a limited number of statistical analysis methods that can enrich frequency data and introduce statistical relevance to reported data. For example, creating videos to demonstrate how to generate valuable information with analysis tools such as Microsoft Excel, SPSS, or SAS.

The design of questions will permit advanced statistical analysis generating complex results that link predictive behaviors to existing health outcomes to inform the decision-making process regarding viable programs, assessment methods, and adjustments to physical and mental health care. This is based on the survey collection process that will overcome the limitations of the current collection procedure, collect information from a larger demographic, and thus provide a more accurate representative of behavior patterns associated with actual health status from which to derive and recommend treatment, care programs, and other services.

- 5) Costs you expect to incur during this prototyping phase

Costs:

- 1) \$5,000 Travel, hotel accommodations, and presentation materials for CDC direct.

The Healthy Behavior Data Challenge

Phase 1 Submission Template

- 2) \$5,000 Establishing point of contact and working with various fitness tracking manufacturers to supply data field constraints; suggest incorporating other information such as differentiating alcohol consumption in caloric intake; and graphic methods that increase use of tracking devices and translation of data.
- 3) \$5,000 Establishing point of contact and working with various statistical software vendors to develop vimeos to discuss a limited number of automated tools to interface and enrich frequency data, encourage advanced analysis and reporting.
- 4) Balance: cost to offset research dedicated to theory, constructs, and analysis methods towards the further development of the survey instrument.

(1500-word limit)

14) 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.

Extending Public Health Surveillance Reporting Through Digital Data Collection of Behavior Patterns and Existing Conditions

This proposal is based on the collection of data through a digital method using existing pre-registration clinical service portals to invite study participants to discuss their behavioral patterns in relation to their diagnosed clinical status. The self-reporting collection method engages participants using a method of introduction to study, invitation to take the survey, and reminder notification through digital contact such as text and email, and also social media, and existing online communities that track relevant information relating to physical behavior and caloric intake. The method is an opportunity to include a greater representative population using personally reported behaviors and an accurate clinical assessment of existing diagnosis and health conditions embedded in the global diagnosis standard of the International Classification of Diseases. Then, proposes to optimize the historical collection of data using frequency counts by introducing embedded analysis methods that add value and statistical significance to reported data. Cumulatively, these modifications provide the foundation for an evidence-based approach to public health data analysis, the path to overlay survey data with data from other agencies (i.e., pollution particular matter concentration by zip code), and empower decision makers with analysis results using advanced statistical analysis methods.