Title

Workgroup 3 - **Study Design** *Recommendations from the Digital Sensing Workshop held at UCLA Feb 28-March 2, 2023*

Author

Digital Sensing Workshop Participants (https://ucla.box.com/v/dig-sensing-wkshp-pubroster)

Publication Date

2023-05-22

STUDY DESIGN

WORKGROUP #3 – COMPILED RECOMMENDATIONS

These recommendations were compiled as an output of the Advancing the Utility of Digital Sensing Tools for Mental Health Research workshop ("Digital Sensing Workshop") sponsored by the UCLA Depression Grand Challenge, Wellcome Trust and NIMH. Workshop participant roster may be found at: https://ucla.box.com/v/dig-sensing-wkshp-pubroster.

Note: Workgroup 3 produced two reports—one on study design and one on reporting.

Please submit feedback at https://bit.ly/dig-sensing-report-feedback by August 31, 2023.

Design

The following a list of elements to consider in the the design of digital mental health studies. Not all elements may apply to all study designs.

- **Study Purpose.** When specifying study purpose/value to inform study design, consider:
    - The type of study you plan to run (exploratory, hypothesis testing, RCT, multiple goals etc.?) and how that should influence your design. Is it possible/appropriate to address multiple goals in the same study?
        - For an RCT have you consulted the CONSORT checklist to ensure the study design and procedures conform to standards and produce data required for publication?
        - For any digital mental health cohort study, have you consulted the EVIDENCE-MH checklist?
        - Have you considered broadly the impact of your study on populations that you are studying? (e.g., might it increase visibility, stress, stigma, etc.)
        - Have you considered broadly the impact of your study on populations that you are not studying? (e.g., might it be seen as exclusionary?).
        - Identifying the broader impacts of your study (and possible unintended outcomes)

- **Study Team.** Does your team represent all areas of expertise and experience necessary?
    - Including representative individuals with lived experience in your study team.
    - Including representative individuals from necessary disciplines (e.g., psychology, human factors/HCI, ubiquitous computing, software engineers, statisticians, computer scientists, data scientists, etc)
    - Have you designed the study and protocol with team science in mind?
        - Does your study plan represent the needs and professional interests of all investigators.
        - Are there processes in place to respectfully manage differences in scientific methods across the different disciplines?
        - Do you have a strategy to embed people with lived experience in the team including development, implementation, interpretation and dissemination of results?

- **Feasibility and Usability.** When establishing feasibility and usability (lived experience input), consider:
    - Feasibility, usability and performance of the:
        - Devices
            - Affordability (e.g., are these types of devices affordable for people in a broad range of contexts, communities, countries etc.)
            - Availabilty (e.g., are these types of devices and any services (e.g. data plans) available to people in a broad range of contexts, communities, countries etc.)
            - Stigma (e.g., will the devices be stigmatizing)
            - Burden and unintended effects (e.g., will the devices present an unreasonable burden to participants and/or influence their behaviors in unintended ways).
        - Algorithms
            - (see below) Are algorithms understandable and unbiased for the population you are studying? Will you be at risk of finding spurious correlations with their measurements?
        - Data management and flows
            - Where will the data be transferred to and stored? Is this process understandable and acceptable to the participants? Is it ethical and legal?
            - Will there be data deletion policies that need to be adhered to? How will these impact analysis?
        - Duration
            - Will fatigue influence adherence to the study over time? How will that be mitigated?
            - Have you considered the impact of environmental factors, such as the impact of seasonality on mental health.

- **Study Participants.** Digital mental health sensing studies often have to balance the effects of heterogeneity on algorithms and generalizability to a larger population. Differences in age, lifestyle, income, culture, geographic location, weather, and many other factors can impact both how people use the devices from which the sensor data are coming as well as the outcomes measured. Most studies are not large enough to account for all of the possible heterogeneity. Narrowly defined study samples may produce stronger effect sizes or greature accuracy, but may also be less generalizable. In identifying study population and tailoring your recruitment strategy:

    - Have you considered how your inclusion and exclusion criteria may affect effect sizes, accuracy, and generalizability?
    - What plans do you have to ensure your sample is representative with in your intended population in terms of race, ethnicity, income, age, gender, geography. If you are considering a narrower sample, within that sample, are there groups that you need to ensure are adequately represented?

- **Recruitment.** When designing a recruitment strategy, consider:
    - Have you tailored your recruitment strategy to reduce the potential for bias due to exogenous effects? Some considerations include:
        - Are the recruitment methods likely to attract a sample that differs in any ways from the population to which you intend to generalize. As this is often the case, are there ways that you can mitigate or understand how this expected sample bias impacts your ability to generalize?
        - Seasonality (e.g, changes in symptoms related to season, such as Seasonal Affective Disorder). Will your recruitment timing and follow-up period be affected by any seasonal effects?
        - Economic/social events
    - How will you continually adjust the sampling strategy to achieve a representative sample
    - Collect data to ensure that participant flow can be monitored.

- **Sample Size and Duration.** When determining a sample size and study duration, consider:
    - Is your study suitably powered to answer your questions?
    - What does a representative sample look like?
    - Key considerations:
        - Number and distribution of features. Measurement error of sensors for each feature.
        - Number/frequency of mental health assessments/events of interest.
        - Extent of missing data, particularly in multivariable analyses where multiple datastreams are required for analysis.
        - If conducting a longitudinal study, consider accuracy versus ease of collection for sensors and psychometric instruments.
        - Rate of change of measures of interest
        - Size of population and distribution of labels

- **Model.** Specify a conceptual model
    - Consider before beginning the study how the sensed data will map onto the psychological and behavioral constructs you intend to detect. Is there a conceivable model for how the sensor data and features might predict your outcomes?
    - Consider factors that might impact or bias those models. For example
        - the impact of weather, climate, or age on GPS or accelerometry data
        - income or connectivity on the acquisition of data
        - patterns of using or turning off the device
        - or any other potential factors?

- **Devices.** When using sensors/devices in a study, consider:
    - Whether you need to provide devices or can rely on participants' own devices (e.g., their own smartphone). If the latter, how will you control for differences across devices/software (e.g., OS). Consider how relying on participants' own devices might influence participation of groups frequently under-represented in research (e.g. ethnic minorities, lower SES).

- **Self-Report.** When designing or using self-report instruments, consider:
    - Behavioral/environmental context
    - Advocating for intentional design of self-report content AND delivery/context
        - Which questionnaires are most appropriate?
        - Where possible, use standardized measures (e.g., measures that have appropriate psychometrics) (e.g., PHQ-8/9, GAD-7, etc.) that are recommended as common data elements by the NIMH, Wellcome Trust, or similar agencies or that are commonly used and accepted.
        - How will these be delivered and how will that impact responsiveness, response time and responses:
            - On what surface?
            - At what time/in what place?
            - Will the participant have some control (e.g., can they postpone, choose when to complete, etc.) and if so, how will that impact interpretation of the data?
        - If using non-standardized measures, consider using measures that have been used by others to support broader interpretation. If developing questions, ensure that they are written at an appropriate educational level, that they are understood by the target population as intended by the research team.

- **Risk Mitigation.** When designing risk mitigation strategies with respect to mental health, consider:
    - What unforeseen/unanticipated events might be detectable through assessments or by sensors. Develop a plan for managing these risks (or be clear with participants during the consent process that risks will not be managed).
    - For risks inherent in the population (e.g. suicidal ideation, worsening of symptoms, etc.) develop a management plan.

- **Data Quality and Management.** When design processes for ongoing data quality management/validation, consider:
    - How you will identify and manage data anomalies
    - How ambiguities in sensor data will be investigated.
    - The steps taken to minimize missing data and whether you will intervene if missing data is detected.

- **Data Analysis.** When designing a data analysis strategy, consider:
    - Algorithms and Models. When applying machine learning models or ``tuned'' algorithms, consider:
        - Do the assumptions of each model hold for the data in your study?
        - Are there any reasons that the performance of those models/algorithms would differ in the population you are studying from other populations, or differ across populations you are studying?
        - Are there any reasons that the performance of those models/algorithms would differ in the context you are deploying them from other contexts, or differ across populations you are studying?
        - Do you have a strategy to collect a validation set to quantify these effects?
        - Can the study population understand the function and limitations of these models/algorithms sufficiently well to provide adequate informed consent?
    - How will missing data be reported and managed?
    - How will participant engagement with the study protocol be reported?
    - How different stakeholders and representative individuals with lived experience will have have access to the data and in what form. Are visualizations, data aggregations or dashboards needed?

- **Other considerations:**
    - Does your observational study constitute an intervention? In observational digital studies there may be considerable contact with participants including steps taken to reduce missing data or minimise clinical risks. Consider whether such steps may have an impact on the outcomes under study (including mental health).