Study Protocol and Statistical Analysis Plan

Leveraging AI to transform dietary choices for cardiovascular health in young adults experiencing food insecurity

NCT Number: not yet assigned

Date approved: 06/04/2025

1) Protocol Title

a) Title: Using an AI mobile application to promote healthy eating

b) Protocol Version Date: version 1; 04/29/2025

2) Objectives

a) Describe the purpose, specific aims, or objectives:
   a. Aim 1: To assess the accuracy of PlantVillage Food against the gold standard of observed weighed food records (WFR).
   b. Aim 2: To determine the feasibility (adherence, acceptability, and usability) of the behavioral intervention of PlantVillage Food and its effects on food choices and diet quality.
   c. Aim 3. To assess whether the accuracy or feasibility of PlantVillage Food varies by food security status.

b) State the hypotheses to be tested:
   a. Hypothesis 1. PlantVillage Food will accurately estimate nutrient intake compared with WFR in U.S. college-aged students.
   b. Hypothesis 2a. The PlantVillage Food app will be considered feasible by participants. Exploratory Hypothesis 2b. Differences in food choices and diet quality will be observed by intervention vs. control groups.
   c. Hypothesis 3. There will be no differences in accuracy, adherence, acceptability, and usability measures between food secure vs food-insecure households.

c) Describe how this research will add to existing knowledge or how the outcomes of this project will be used:
   PlantVillage Food will be one of the first valid AI-assisted dietary tools in the U.S. that could transform the implementation of future dietary interventions.

3) Enrollment Numbers

a) Total number of subjects to be enrolled in this study: **180**; 66 for the validation study, 114 for the pilot RCT.

b) If this is reliance, list the number of subjects to be enrolled at each site (UC Davis: #, Relying Site 1: #, Relying site 2: #): N/A

c) Provide a rationale (e.g., statistical justification, power analysis, etc. for the number of subjects to be enrolled:
   Sixty-six participants aged 18-24 will be recruited from the UC Davis campus student population and will provide 90% power to conclude equivalence between PlantVillage Food and the WFR, with a 15% equivalence margin while allowing for 20% attrition.

4) Inclusion and Exclusion Criteria

a) Inclusion Criteria: Receiving a residential meal plan (or eat most meals from the dining commons, i.e., 5/7 day meal plan) in the dining commons, willing to use their personal smartphone with the PlantVillage Food application, and willing to record their food consumption using the app, and having their foods weighed.

b) Exclusion Criteria: Not willing to be followed up during the WFR, under 18 or over 24 years old, not receiving the 5 or 7 day meal plan, and screened for disordered eating behavior.

c) Age Range: 18-24 years

d) If applicable, describe the screening procedures to determine eligibility: The SCOFF questionnaire (https://pmc.ncbi.nlm.nih.gov/articles/PMC1070794/), consisting of five questions designed to screen for eating disorders, will be utilized to identify students who may have an eating disorder. A threshold of two or more affirmative responses to the five questions will serve as an indicator of an eating disorder. Additionally, the screening questionnaire includes a question about food insecurity experiences in the past 12 months.

5) Procedures Involved

Aim 1: Each participant will visit the dining commons on 3 days, including 1 day for recruitment, consent, and training, followed by 2 nonconsecutive days for WFR. Research assistants will obtain consent, help the participant download the PlantVillage Food app, and train participants on PlantVillage Food application use on the day of recruitment. A PlantVillage Food video tutorial will also be available on the app for reference, emphasizing a dynamic training approach. Following the recruitment visit, participants will return to collect the WFR, recording all instances of food and beverage consumption at the dining commons concurrently as participants recorded foods and beverage consumed using PlantVillage Food. Research assistants will record the weights of foods and beverages consumed in the dining commons.
Participant access to the app will be removed upon completion of the study.

Aim 2: Participants will be randomized at enrollment into either intervention or control groups, where participants in the control group will receive a version of PlantVillage Food with functionality limited to dietary assessment (here referred to as limited PlantVillage Food), and the intervention group will receive a gamified PlantVillage Food interface. The research team will separately train participants in both groups on how to record food consumption on PlantVillage Food. A 2-day troubleshooting period will allow participants to practice using PlantVillage Food to record food consumption and identify problems related to application use. On day 4, participants in both groups will be free to use PlantVillage Food for the following 6 weeks.
Participant access to the app will be removed upon completion of the study.

Aim 3: To recruit a representative sample of students from food-insecure backgrounds, we will aim to recruit first-generation, transfer, Pell Grant-eligible, and low-income students.

We will recruit through campus events and offices on campus that interact with these student groups and have previously shown high rates of food insecurity with unlimited dining commons meal plans ensuring the reach of diverse under-represented students. Participants will then be randomly selected for participation. In addition to using the PlantVillage Food app and collecting WFR, we will also collect food insecurity status at enrollment using a demographic questionnaire.

Surveys/Questionnaires

- Background information on participants' socioeconomic status will be collected using a structured questionnaire that will include questions about age, race and ethnicity, biological sex, food insecurity, and federal nutrition assistance participation (i.e., SNAP).
- Food Security: The validated USDA 10-item Food Security Survey Module will capture food security over the prior 12 months and be administered at the recruitment visit.
- Acceptability and usability will be assessed using a structured questionnaire based on a Likert scale from 1 (strongly agree) to 5 (strongly disagree) administered at the end of the study period.

Statistical Analysis Procedures

Aim 1: Descriptive statistics will be used to characterize energy and nutrient intakes by person-day. To ensure comparability across the 2 methods, consumption periods will be matched across the 2 methods, excluding foods that were reportedly consumed outside the WFR period. Bland–Altman plots will be used to visualize differences in intakes for the different methods compared with the average intake by method. Equivalence testing with a 15% equivalence margin will be undertaken using the two 1-sided paired t-test method, examining mean differences of log-transformed nutrient intakes using mixed-effect models including a random effect at the person level to account for the repeated measures. The differences in log-transformed intakes will be equivalent to ratios where the numerator is the intake as estimated by PlantVillage Food and the denominator is the intake from the WFR. The extent of agreement between methods will be examined using the concordance correlation coefficient (CCC) with bootstrapped standard errors. Sources of error will be also examined by individual foods including *1*) instances of omissions (foods consumed but not reported) and intrusions (foods reported that were not consumed) for PlantVillage Food and WFRs and *2*) portion estimation errors, assessed by comparing the reported food amounts by PlantVillage Food to the observed amounts by WFR for the top 20 most commonly consumed foods.

Aim 2: Descriptive statistics will summarize socioeconomic characteristics for intervention and control groups. The primary objective is to estimate feasibility measures for the full sample, and as a secondary objective will be assessed separately by the intervention arm to test whether gamification reduces adherence or likability. Data will be longitudinal as outcome data will be generated throughout the 6 weeks. All models will include a random effect of participants to account for the repeated measures. Linear regression will be used for continuous outcomes and modified Poisson regression for dichotomous. Modeling will include an evaluation of group differences across the full study period as well as changes

over time (group x time interaction). The interaction between the time variable and the dummy for the PlantVillage Food version will be included because users will take time to adapt to new technologies, so the effect of PlantVillage Food could change with time.

Aim 3: Moderation by food security status will be assessed by including an interaction term between the assessment method and binary food security in the Aim 1 and 2 models. If p-for-interaction<0.10, then parameter estimates will be re-assessed within each of the two food security strata.

6) Study Timelines

a) Duration of an individual subject's participation in the study: Validation study: 3 days in the fall quarter 2025, i.e., ~40-50 minutes at enrollment visit, 5-7 hours total for 2 WFR days; Feasibility study (pilot RCT): 6 weeks in the winter-spring quarters 2026, approximately 10-20 hours' time commitment.

b) Estimated timeline to enroll all study subjects: 1 quarter (10 weeks)

c) Estimated timeline for the investigators to complete this study (complete primary analyses): The study is expected to span a period of 3 years, with data collection estimated to take about 1 year, followed by data cleaning and analysis, and dissemination of results.

7) Study Endpoints

a) Describe the primary and secondary study endpoints: **Validation study:** Completion of 2 days of weighed food records; **Outcome**: accuracy of the mobile application.
Six-week intervention period- Primary outcome: app feasibility (usability and acceptability) Secondary outcome: Diet quality. **Exit survey:** Completion of an endline survey assessing application acceptability.

b) Describe any primary or secondary safety endpoints: N/A

8) Recordings

This research involves:

☐ Audio recordings

☒ Photographs [of food]

☐ Video recordings with audio

☐ Video recordings without audio

☐ None of the above

9) Data and/or Specimen Management and Confidentiality

Before completing this section, see Privacy and Confidentiality and HIPAA Guidance.

a) List any identifiers that will be collected during this study (e.g., name, medical record number, date of birth, video recordings, etc.): Name, email address, and phone number.

b) If any identifiers will be stored, how long will they be kept? 1-2 years after the study's completion.

c) For data that is coded with a linking key, at what point will the linking key be destroyed? 1-2 years after the study's completion.

d) For any recordings, at what point will the recordings be destroyed? N/A

e) If this research is both federally funded *and* you are using identifiable data/specimens, please explain why this research cannot be completed using de-identified data/specimens: N/A

NOTES ABOUT USE OF RECORDS

UC Davis Medical Records: UC Davis Health Electronic Health Record (EMR/EPIC) also contains the clinical data for Marshall Medical Center (MMC). MMC patient data cannot be accessed for research purposes. Researchers must take the necessary steps to ensure that MMC data is not accessed, used, or disclosed for UC Davis Health research purposes. If protected health information or personal information from the medical records will be stored on an encrypted device, investigators must follow applicable university policies (UC Davis Hospital Policy 1313, UCDHS P&P 2300-2499, and UC Business and Finance Bulletin on Information Security (IS-3)). Please contact the Biomedical Informatics Department for assistance with data security.

If identifiable protected health information is extracted from the UCDH EMR, it may not be re-disclosed/released outside the study team.

UC Davis Student Education Records: If this study involves use of UC Davis students' educational records (including records in the PI's own possession such as course exams/assignments), you must consult with the Registrar's office to see if all requirements of the Family Educational Rights and Privacy Act (FERPA) are satisfied.

10) Provisions to Monitor the Data to Ensure the Safety of Subjects

N/A

11) Withdrawal of Subjects

The sample size was calculated to account for 20% attrition.

12) Risks to Subjects

☒ This research may pose the risk of loss of confidentiality. The risk will be minimized through the data protection plan.

13) Potential Benefits to Subjects

☒ Research subjects are not likely to receive any benefit from the proposed research, but others may benefit from the knowledge obtained.

14) Provisions to Protect the Privacy Interests of Subjects

Survey questionnaires will be electronic and self-administered to ensure data privacy. At enrollment, participants will complete sociodemographic and food security surveys, and at the endline, an acceptability questionnaire using a Qualtrics survey on study-provided tablets. Research assistants will be available to answer any questions and assign participant IDs but will not have access to participants' responses.

15) Sharing of Results with Subjects

☐ Results will not be shared with subjects

☒ Results will be shared with subjects - Describe: A newsletter summarizing the general study findings will be prepared and made available to the UC Davis community.

a) If results will be shared, describe the results (study results or individual subject results such as results of investigational diagnostic tests, genetic tests, or incidental findings) to be shared with subjects or others (e.g., the subject's primary care physicians).
Note: There are restrictions on the information that can be provided to individuals when the tests performed on their specimens are considered investigational or the tests are not being performed at a CLIA-certified laboratory.

16) Data Banking

a) Will the data ever be used by you or other researchers to answer a different research aim that is not included in this study?

☒ Yes - Complete the remainder of Section 16.

☐ No - Do not complete Section 16. Go to section 17.

b) What will be banked for future use?

☒ De-identified data. Banked data cannot be linked to an individual.

☐ Identifiable data/specimens – Banked data/specimens will include identifying information.

☐ Coded data/specimens – Banked data/specimens will be stripped of identifiers and assigned a code. A key will be maintained that links the identifiers to the data/specimens.

☐ Contact information will be banked for future research opportunities.

c) Where will the data be banked? UC Davis Box folder with secure access.

d) How long will the data be banked? 5 years after the conclusion of the study.

e) Who will have access to the banked data/specimens? The PI, project manager, statistician, and graduate students involved in the study.

f) Describe the procedures to release data. Include the process to request a release, the approvals required for release, who can obtain data, and the data to be provided.
Note: Identifiable protected health information extracted from the UCDH EMR under an IRB-issued waiver of HIPAA Authorization may not be re-disclosed/released outside the study team.

Only the study's principal investigators, co-investigators, and research staff will have access to the data, which will be de-identified. Data will be made available upon reasonable request.

17) Multi-Site Research

N/A

18) Community-Based Participatory Research

N/A

19) Compensation for Research-Related Injury

N/A

20) Economic Burden to Subjects

Eligible participants will have a meal plan and access to the dining commons and as such, will not incur any additional costs.

21) Review Requirements

Some research projects require specific IRB determinations.

a) Are there any contractual obligations or other considerations that require IRB review of this research, or review at intervals other than those required by the Common Rule or FDA?

☐ Yes

☒ No

b) If yes, please describe: