Title: The reliability and validity of the PortionSize™ and MyFitnessPal apps (Study 3: Assessment in free-living conditions).

Principal Investigators: Corby K. Martin, Ph.D.; John W. Apolzan, Ph.D.

Co-Investigator: Stephanie T. Broyles, Ph.D.; Jennifer C. Rood, Ph.D

Protocol Version Date: May 4, 2023

Background

Accurately quantifying food intake is vital to promoting health and reducing chronic disease risk. Food intake encompasses energy intake, nutrient intake (macronutrients, micronutrients, vitamins, minerals), and intake of various food groups (e.g., fruits, vegetables), and thus reflects the nutritional status of individuals. Nutrition affects disease risk, including risk of developing obesity, [1] diabetes, [2] and cancer, [3] all of which negatively affect the United States (U.S). Nonetheless, accurate assessment of food and nutrient intake has remained challenging, despite an improvement in methods. Self-report methods, namely food records, are a mainstay of nutritional epidemiology research, [4] with food recall being another popular method. [5] These methods rely on the participant to accurately estimate portion size and, for food recall, remember what was consumed. The accuracy of these methods have been questioned [6-9] and the problems with human recall have been comprehensively outlined. [10] As a result, there remains a significant need for methods that are sufficiently accurate to provide researchers with good outcome data and to guide health promotion efforts.

The PortionSize™ app was designed by our laboratory to overcome the limitations outlined above, and to guide users to follow specific diets. PortionSize™ relies on users capturing images of their food selection and waste. Food intake data are immediately provided since the user relies on built in tools, including templates, to estimate portion size. Determining the accuracy and reliability of PortionSize™ is vital before users can utilize the app to obtain immediate feedback about their food intake. We accordingly aim to test the accuracy of PortionSize in free-living settings and compare this to another app that has been adopted widely (MyFitnessPal©).

Methods

Study Design

We will recruit adults to use the PortionSize app and the MyFitnessPal free-living conditions. In a within subjects and randomized, counterbalanced design, the accuracy of the PortionSize and MyFitnessPal apps at estimating energy intake in free-living conditions will be tested against the gold-standard, doubly labeled water (DLW), over two non-consecutive four-day periods that should include at least one weekend day and encompass the same days of the week during each period. Participants will be trained in our clinic to use the apps prior to their use in free-living conditions. We will integrate customizable notifications and reminders to remotely prompt participants to record data,

resolve data collection problems, etc. These notifications and reminders utilize Ecological Momentary Assessment (EMA) methodology to maximize data collection protocol adherence by reminding participants to capture and send food information at relevant times.

Participants and Recruitment

We will recruit 63 adults (approximately equal proportion of men and women) who will be randomly assigned to use PortionSize™ or MyFitnessPal first. We expect to enroll an approximately equal number of participants with body mass indexes (BMI; kg/m^2) between 18.5 to 25, 25 to 30, and 30 – 50 kg/m^2. Recruitment will be stratified based on sex and body mass index.

- Inclusion criteria include:
  - Male or female, age 18-62 years
  - Body mass index (BMI) 18.5-50 kg/m^2
  - Ownership of an iPhone model 9 or later, which the participant is willing to use for the study
  - Access to Apple ID, password, and email address and willing to use them during the study
  - Willing to use data and any accompanying charges as part of study participation
  - Willing to complete all study procedures and adhere to study visit timelines
  - Willing to be re-contacted for future research and/or follow-up
- Exclusion criteria include
  - Have been 6-month weight unstable (gain/loss >5% last 6 months) and/or active weight loss program
  - Any cardiometabolic disorder that significantly affects what or how much people eat
  - Active cancer diagnosis or treatment that significantly affects what or how much people eat, excluding some melanomas and other cancers
  - PBRC employee, as previous reviewers argued that they are not representative of the community
  - 3rd member of the same household to participate (i.e. the 1st and 2nd members of the household cannot participate simultaneously)
  - Women who are currently pregnant or breastfeeding (self-reported)
  - Diagnosed with an uncontrolled thyroid disorder (controlled ≥ 3 months of medication)
  - Have had or plan to have weight loss surgery *(gastric band removal may be allowed at PI discretion)*
  - Consume >28 alcoholic beverages per week
  - Anyone severely immunocompromised

- Serious digestive disorders that significantly affect what or how much people eat Fluid imbalance and/or on diuretic
- Follows a diet mainly consisting of foods that study staff deems unquantifiable for this study (i.e., liquid supplements, gluten-free specialty items) or consumes a high quantity of specialty foods
- Any condition or circumstance that, in the judgement of the PIs, could impede study completion

Procedures

Prior to enrollment, participants will respond to recruitment, qualify for the study, and schedule a phone call with study staff. The purpose of the phone call with study staff is to ensure that the study is a good fit for the participant in regard to inclusion/ exclusion criteria, participant schedule, and other factors that may affect participant ability to complete the study and collect necessary data. This call also helps ensure that potential participants have access to their Apple ID and password for the first visit.

Screening visit

In the screening visit, written informed consent will initially be obtained. We will then obtain:

- Demographics: age, race, ethnicity, sex, marital status, education level
- Anthropometrics: height, weight and BMI
- Lifestyle Interview

Visit 1

On being deemed eligible, participants will be randomized to either the PortionSize app or MyFitnessPal app to use first. They will then receive training on the app that they will use to subsequently rate meals with the assistance of our staff members. That is, individuals randomized to the PortionSize group first will be trained on PortionSize, whereas those randomized to my MyFitnessPal group first will be trained on MyFitnessPal. Training (~30 minutes) will teach participants how to self-report food intake and estimate portion size per our standard procedures. Subsequently, they will begin using the app within free-living conditions to measure food, beverage, and supplement intake over 4-5 days, with one of these days preferably being a weekend day. Participants assigned to the PortionSize group will be instructed to use PortionSize to estimate food provision and plate waste and the app will immediately calculate food intake, with energy intake being the primary outcome variable. PortionSize will also transfer the food images to the Data Capture and Analysis Platform (DCAP) for possible expert rater analysis, and these are secondary outcome data. Participants assigned to the MyFitnessPal condition will enter estimated food intake into the app. For the DLW method for measuring the energy intake, participants will be given a mixture of 1 part 99% ^{2}H and 19 parts 10% ^{18}O. Both males and females will receive a dose of 1.5 g/kg total body weight, which provides 0.055 g/kg ^{2}H and 0.1045 g/kg of ^{18}OThe dose will be followed by ingesting 100 mL of tap water that is used to rinse the dose container. The

dose will be given early in the morning after an overnight fast and collection of two baseline urine specimens. Two urine specimens will be collected and discarded 1.5 hours and 3 hours post dose. A urine specimen will be collected 4.5 hours and 6 hours post dose. Participants will be instructed to collect the 6-hour post dose sample from home and return it to Pennington on their next visit. Participants will then record their body weight at roughly the same time each morning between this visit and Study Visit 2 at home via scales provided to them.

Visit 2

After 4-5 days, participants will record their body weight in the morning and return to the laboratory. They will then quantify satisfaction, ease of use, and adequacy of training for the app they used for 4-5 days. All participants will also complete the Computer System Usability Questionnaire, a standardized, reliable, and valid questionnaire originally designed to evaluate computer programs[11,12] that has been used to quantify the usability of mobile phone apps.

Further, participants will compete the PortionSize and MyFitnessPal-specific Dietary Health Questionnaires (DHQ III), which is modified such that participants will report the presence of vitamins and supplements during the days of food intake tracking, referred to as the DHQ III Subset. Additionally, a urine sample will be collected, and the time of collection will be documented. Body weight will be re-measured in the clinic during the morning for DLW purposes

Visit 3

Following a 1-2-day break, participants will return to laboratory. They will receive training on the second app that they didn't use in the first arm of the trial. That is, participants will receive training on PortionSize if they were randomized to use MyFitnessPal in the first condition of the trial and vice versa. Subsequently, they will begin using the second app within free-living conditions to measure food, beverage, and supplement intake over 4-5 days.

An additional urine sample will be collected, and the time of collection will be documented. Body weight will be re-measured in the clinic during the morning for DLW purposes. Participants will be provided with a scale that automatically transmits weight to the internet and will be asked to weigh themselves at approximately the same time t every morning between this visit and Study Visit .

Visit 4

After -5days, participants will record their body weight in the morning and return to the laboratory. They will then quantify satisfaction, ease of use, and adequacy of training for the app they used for 4 days. All participants will also complete the Computer System Usability Questionnaire, a standardized, reliable, and valid questionnaire originally designed to evaluate computer programs[11,12] that has been used to quantify the usability of mobile phone apps.[13,14]. Participants will also complete User Satisfaction Surveys that are specific to either PortionSize or MyFitnessPal. Participants will

complete an app Preference Survey, enabling us to see what app they preferred. Scales will be returned to the center. Further, participants will compete the PortionSize and MyFitnessPal-specific Dietary Health Questionnaires (DHQ III), which is modified such that participants will report the presence of vitamins and supplements during the days of food intake tracking, referred to as the DHQ III Subset.

The final urine sample will be collected, and time of collection will be documented. Body weight will be re-measured in the clinic during the morning for DLW purposes. Participants will meet with IBL staff to review food intake tracking procedures and methods as needed.

Schedule of procedures

The schedule of procedures is outlined below:

	Screening Visit	Day 0	Day 1-4	Day 5*	Break (Day 6)	Day 7	Day 8-11	Day 12/13+
Informed Consent, Demographics, Anthropometrics, Eligibility Evaluation, , Lifestyle Interview	X							
Randomization and Doubly Labeled Water dosing		X						
Training on Portion Size or MyFitness Pal app		X				X		
Use app to measure food intake			X	X*			X	X+
DHQ III Subset				X				X
Urine Collection, body weight measured in clinic		X		X	X*	X		X
Record daily body weights at home			X	X	X*		X	
User Satisfaction Survey; Computer Systems Usability Questionnaire				X	X*			X
PS and MFP app preference survey								X

*Day 5 will be used as an additional food intake collection day if the data collected on Days 1-4 is deemed insufficient by study staff. In this case, participants will come to PBRC on Day 6 for clinic procedures and to complete questionnaires. Participants will return to PBRC on Day 7 for app training and proceed as originally scheduled.

+ Day 12 will be used as an additional food intake collection day if the data collected on Days 8-11 is deemed insufficient by study staff. In this case, participants will come to PBRC on Day 13 for final clinic procedures and to complete questionnaires.

The screening visit will last approximately 1hr. Visit 1 (day 0) will last 3.5-6.5 hrs., whereas Visits 2 (day 5), 3 (day 7) and 4 (day 12) will last approximately 1hr.

Participants may be contacted by study staff during food intake data collection periods to address issues with the app functionality, feedback, and data collection procedures. Day 6 may be used as a break day or a "makeup" day for measuring food and beverage intake if the participant missed one of the 4 previous days of collection or if the data collected is deemed insufficient by study staff.

Outcome measures

The **primary analyses** assessing equivalence between PortionSize (and MyFitnessPal) and the criterion measure, which is energy intake assessed with DLW. These analyses will rely on equivalence testing using the Two One-side T-test (TOST)[15,16] method. Subsequent regression-based **exploratory analyses** will determine if error from the apps varies by body mass (kg, BMI), age, or sex; and will evaluate differences in user satisfaction between the apps. Bland-Altman analysis[17] will be used to test for differences in error variance over levels of the variable being measured (e.g., food intake).

- Energy intake (kcal)
- User satisfaction and usability of the app
- PortionSize vs. MyFitnessPal Preference

<u>Power calculation</u>

Estimated energy (kcal) values from the PortionSize and MyFitnessPal apps will be compared to directly DLW, which will serve as the criterion measure. Equivalence tests that follow the Two One-sided T-test (TOST)[15,16] method will be used (error reflects the difference between DLW energy intake and estimates from either PortionSize or MyFitnessPal). Based on data from the literature, we originally estimated that we had sufficient power in Study 3 to detect equivalence within ± 12% of intake measured by DLW with 53 participants in the PortionSize condition and 43 participants in the MyFitnessPal condition. This initial power analysis assumed the following for the equivalence tests of PortionSize: energy (kcal) intake as the outcome variable, alpha = 0.05, power >0.80 as acceptable, and mean intake from DLW = 2,360 kcal with an SD (difference) = 694 kcal. The equivalence tests of MyFitnessPal assumed: energy (kcal) intake as the outcome variable, alpha = 0.05, power > 0.80 as acceptable, and mean intake from DLW = 2,564 kcal (lunch) with an SD (difference) = 671 kcal.

We have since completed a pilot study with PortionSize that compared the app to weighed intake in a laboratory setting. We found that our initial power analysis dramatically overestimated power, particularly for the PortionSize equivalence tests. We reran the power analysis and found that, with 58 participants completing the PortionSize

condition, we estimated that we have power to detect equivalence within ± 18.5% of the directly weighed value. This power analysis assumed: energy (kcal) intake as the outcome variable, alpha = 0.05, power > 0.80 as acceptable, and mean intake from DLW = 2,360 kcal with an SD (difference) = 1,120 kcal. These new power analyses caused us to modify the Statistical Design and Power section as follows. We will now test for equivalence with a properly powered test that relies on equivalence bounds of ± 18.5% of the directly weighed value. We will also test for equivalence at the original bound of ± 12%, as well as a new bound of ± 15%, though we recognize that these tests will likely be underpowered.

The reliability of each apps' food intake estimates will be quantified with intraclass correlation coefficients (ICCs), and reliability will be supported if the ICC is significant and ≥ 0.50. A power analysis indicates that, with 2 observations per subject, an ICC of 0.54 can be detected with 23 subjects at 80% power. The power analysis was conducted in R software using the 'ICC.Sample.Size' package [18] based on Zou. [19] Regression-based exploratory analyses are powered to detect an R^2 of 0.106 with 68 participants and thus we have ample power.

Overall, 63 adults will be enrolled. We accounted for up to a 7% to 8% drop-out rate, resulting in 58 completers, though all available data from participants will be analyzed regardless of drop-out status.

Provisions to Monitor the Data to Ensure the Safety of Subjects

Adverse events will be monitored at each intervention visit. The mPIs and their co-investigators will review all data continuously to ensure the safety of each subject.

Withdrawal of Subjects

Subjects may be withdrawn from the study if he/she misses study visits and will be notified of their withdrawal via telephone or mail. If a subject voluntarily withdraws from the study, no additional data will be collected, and they will be considered dropouts in the study.

Risks to Subjects

This study involves no greater than minimal risk. The main risk is breach of confidentiality, and the PBRC team will work to minimize this during data collection, handling, and analysis. With questionnaires, participants do not have to answer any questions they do not wish to answer. Measuring energy expenditure by DLW carries no risk. The two natural atoms given in the water are not dangerous at all and are often given to small children and pregnant women. The water drunk during this test is prepared under sterile techniques.

Potential Benefits to Subjects

Participants may benefit by increased awareness of their portion sizes with various food groups.

Sharing of Results with Subjects

The participants may receive information such as body weight and anthropometrics. Those results will be provided to the participant at the end of the study if requested.

Setting

The Injestive Behavior, Weight Management, and Health Promotion Laboratory at Pennington Biomedical Research Center.

Compensation

Participants will receive $400 for the successful completion of the study.

Provisions to Protect the Privacy Interests of Subjects

All attempts will be made to maintain a subject's privacy. Safeguards such as password-protected computers and networks have been put in place to limit access to subject data. Subjects will be given ample time to read over the consent, ask questions, and agree to participate in the research study. Subjects may decline to answer questions with which they are not comfortable. Each procedure will be explained to the subject before it is performed. We will always ensure the privacy of the subjects.

Compensation for Research-Related Injury

No compensation will be provided for research-related injury.

Economic Burden to Subjects

All study-related tests and procedures will be at no cost to the subject. The subject will incur transportation costs in getting to Pennington Biomedical Research Center.

Consent Process

The PI or one of the designated staff will obtain informed consent during the Screening Visit. Ample opportunity will be given for the subject to review the consent and ask any questions prior to signing the consent form. If subjects wish, they can take the form home and return at a different visit. We are aware that consent is an ongoing process.

References

1. Swinburn BA, Caterson I, Seidell JC, James WP. Diet, nutrition and the prevention of excess weight gain and obesity. *Public Health Nutr.* 2004;7(1A):123-146.
2. Ley SH, Hamdy O, Mohan V, Hu FB. Prevention and management of type 2 diabetes: dietary components and nutritional strategies. *Lancet.* 2014;383(9933):1999-2007.
3. Huang T, Yang B, Zheng J, Li G, Wahlqvist ML, Li D. Cardiovascular disease mortality and cancer incidence in vegetarians: a meta-analysis and systematic review. *Ann Nutr Metab.* 2012;60(4):233-240.
4. Willett WC. *Nutritional epidemiology.* New York: Oxford University Press; 1998.
5. Lytle LA, Nicastro HL, Roberts SB, et al. Accumulating Data to Optimally Predict Obesity Treatment (ADOPT) Core Measures: Behavioral Domain. *Obesity (Silver Spring).* 2018;26 Suppl 2:S16-S24.
6. Beaton GH, Burema J, Ritenbaugh C. Errors in the interpretation of dietary assessments. *Am J Clin Nutr.* 1997;65(4 Suppl):1100S-1107S.
7. Tran KM, Johnson RK, Soultanakis RP, Matthews DE. In-person vs telephone-administered multiple-pass 24-hour recalls in women: validation with doubly labeled water. *J Am Diet Assoc.* 2000;100(7):777-783.
8. Schoeller DA, Bandini LG, Dietz WH. Inaccuracies in self-reported intake identified by comparison with the doubly labelled water method. *Can J Physiol Pharmacol.* 1990;68(7):941-949.
9. Bandini LG, Schoeller DA, Cyr HN, Dietz WH. Validity of reported energy intake in obese and nonobese adolescents. *Am J Clin Nutr.* 1990;52(3):421-425.
10. Archer E, Pavela G, Lavie CJ. The Inadmissibility of What We Eat in America and NHANES Dietary Data in Nutrition and Obesity Research and the Scientific Formulation of National Dietary Guidelines. *Mayo Clin Proc.* 2015.
11. Lewis JR. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International*

Journal of Human-Computer Interaction. 1995;7(1):57-78.
12. Lewis JR. Usability: Lessons Learned ... and Yet to Be Learned. *International Journal of Human-Computer Interaction.* 2014;30(9):663-684.
13. Al Ayubi SU, Parmanto B, Branch R, Ding D. A Persuasive and Social mHealth Application for Physical Activity: A Usability and Feasibility Study. *JMIR mHealth and uHealth.* 2014;2(2):e25.
14. English LL, Dunsmuir D, Kumbakumba E, et al. The PAediatric Risk Assessment (PARA) Mobile App to Reduce Postdischarge Child Mortality: Design, Usability, and Feasibility for Health Care Workers in Uganda. *JMIR mHealth and uHealth.* 2016;4(1):e16.
15. Limentani GB, Ringo MC, Ye F, Bergquist ML, McSorely EO. Beyong the t-test: statistical equivalence testing *Analyitical Chemistry.* 2005:221-226.
16. Walker E, Nowaki AS. Understanding Equivalence and Noninferiority Testing. *J Gen Intern Med.* 2010;26(2):192-196.
17. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-310.
18. *Calculation of Sample Size and Power for ICC* [computer program]. 2015.

19. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat Med.* 2012;31(29):3972-3981.