9-1-2020

The Validity, Time Burden, and User Satisfaction of the FoodImage^{TM} Smartphone App for Food Waste Measurement Versus Diaries: A Randomized Crossover Trial

Brian E. Roe
The Ohio State University

Danyi Qi
Louisiana State University

Robbie A. Beyl
Pennington Biomedical Research Center

Karissa E. Neubig
Pennington Biomedical Research Center

Corby K. Martin
Pennington Biomedical Research Center

See next page for additional authors

Recommended Citation

Roe, B., Qi, D., Beyl, R., Neubig, K., Martin, C., & Apolzan, J. (2020). The Validity, Time Burden, and User Satisfaction of the FoodImage^{TM} Smartphone App for Food Waste Measurement Versus Diaries: A Randomized Crossover Trial. *Resources, Conservation and Recycling* https://doi.org/10.1016/j.resconrec.2020.104858

Brian E. Roe, Danyi Qi, Robbie A. Beyl, Karissa E. Neubig, Corby K. Martin, and John W. Apolzan

HHS Public Access

Author manuscript

Resour Conserv Recycl. Author manuscript; available in PMC 2021 September 01.

Published in final edited form as:

Resour Conserv Recycl. 2020 September ; 160: . doi:10.1016/j.resconrec.2020.104858.

The Validity, Time Burden, and User Satisfaction of the FoodImage™ Smartphone App for Food Waste Measurement Versus Diaries: A Randomized Crossover Trial

Brian E. Roe[a,*], **Danyi Qi**[b], **Robbie A. Beyl**[c], **Karissa E. Neubig**[c], **Corby K. Martin**[c,†], **John W. Apolzan**[c,†]

[a]Department of Agricultural, Environmental and Development Economics, Ohio State Univesity, 2120 Fyffe Road, Columbus, OH 43210 USA

[b]Department of Agricultural Economics and Agribusiness, Louisiana State University, Martin D. Woodin Hall, Baton Rouge, LA 70802 USA

[c]Pennington Biomedical Research Center, Louisiana State University System, 6400 Perkins Road, Baton Rouge, LA, 70808 USA

Abstract

The FoodImage™ smartphone app transmits to researchers users' photographs of food selection and food waste, and includes user-tagged information about waste reasons and destination. Twenty-four participants were trained to record food waste using FoodImage, food waste diaries requiring visual estimation of waste quantities, and diaries requiring scale weights. Participants used each method during three staged food-waste scenarios (food preparation, eating, and clean-out) in a randomized crossover trial. Two participants had extreme values for the weighed diary method; therefore, accuracy results are reported with and without these two participants' data. Error was calculated as waste estimated with the experimental method minus directly weighed waste. Mean absolute error from FoodImage was significantly smaller than or equal to the error from both diary methods in each scenario. Furthermore, the mean values from FoodImage were equivalent to directly weighed values in two out of the three tasks; while weighed diaries were equivalent in two tasks only when the two participants with extreme values were removed. Visually estimated diaries were equivalent for only one task. All 24 participants preferred FoodImage to diaries and all rated FoodImage as less time consuming. Over one week, FoodImage would require ~24 fewer minutes of users' time to record all data. Unlike food waste diaries, FoodImage also transmits data to researchers in real-time and provides detailed data on food selection and intake.

* Corresponding Author: Department of Agricultural, Environmental and Development Economics, Ohio State University, 2120 Fyffe Road, Columbus, OH 43210 USA (roe.30@osu.edu).

†Co-Senior Authorship

Supporting Information: Power analysis, detailed description of measurement methods, main analyses in Calories, figure of measurement error versus criterion values for the *Toss* task.

ClinicalTrials.gov Identifier: NCT03309306

Keywords

Food waste; measurement; validity study; household; smartphone app; food waste diaries; accuracy; time burden; user satisfaction

Introduction

The United Nations and many countries have established goals for reducing food waste.[1,2] In developed countries, households and consumer-facing institutions represent a key source of food waste.[3,4] Hence, understanding consumer response to food waste interventions is critical to meeting goals. However, valid granular consumer food waste data remains sparse.[4–6] Popular methods for collecting household food waste data include surveys, diaries (weighed and unweighed), and waste stream analysis, where each method features tradeoffs among bias, granularity, respondent burden, and cost.[7]

Surveys prompt participants to retrospectively recall the frequency, amounts, or proportions of waste for various foods used at home without the aid of diaries or other data. Surveys limit respondent time burden and cost little to administer. The collected data are as granular as is feasible given respondent memory, knowledge, and cognitive ability. The accuracy of such surveys have been questioned in several fields.[7–14] Survey data are subject to biases due to the retrospective nature of reporting, the cognitive difficulty of estimating the quantity or percent of food wasted, and social desirability bias.[7]

In waste stream composition analyses, household garbage is collected by researchers prior to landfill, sorted into key categories (non-food and different food types), and weighed. The method can be unobtrusive and limit respondent burden. Shortcomings include: the exclusion of food discarded via disposal or sink, fed to pets, composted, etc.; loss of data integrity due to water separation, compaction, co-mingling, and evaporation; an inability to assign a reason for waste; omission of waste in dine-out and other eating occasions; and the logistical burden to the researcher of capturing, sorting and weighing food waste before garbage goes to landfill.[15] Variants of this approach require participants to collect waste in separate bins, which are then collected by researchers.[16–17] This can improve data granularity but increases respondent burden and provides feedback to respondents that may alter behavior.

Diaries provide granular data (waste amounts, types, reasons, destinations), but impose a greater burden as respondents must contemporaneously record food and drink discarded at all meals and food handling situations throughout the study. In some designs, respondents visually approximate waste quantities, while in others respondents use scales to weigh waste, which may reduce approximation error but increases respondent burden and administrative cost. Systematic underreporting of waste is well established for diaries,[18] with reported magnitudes being 40%[19] to 47%[20] less than amounts reported from waste stream analyses (which, as detailed above, omit some waste). Diaries also provide respondents ongoing feedback (i.e., continual documentation of waste levels) that may alter behavior, thus confounding measurement of the effects associated with focal interventions.

Collecting data about food waste created in day-to-day household operation is challenging. Nutritionists face similar challenges when measuring food intake in such free-living conditions. Self-report methods, such as written food intake diaries, require participants to accurately remember food types consumed and estimate portion sizes eaten. The accuracy of diaries for measuring food intake have been questioned.[10, 21–23] Importantly, about 50% of error in self-report food intake methods is due to participant inability to accurately estimate portion sizes served and left uneaten.[24] Additionally, self-report methods are associated with: larger underestimates of food intake for overweight or obese individuals;[21] respondents selectively underreporting dietary fat intake;[26] and respondents undereating during monitoring periods.[25–26] Each issue can yield data that misrepresent habitual food intake. We expect similar avenues of error with food waste measurement.

These challenges have been addressed for the measurement of food selection and food intake through the development and deployment of the Remote Food Photography Method© (RFPM) and the SmartIntake® smartphone app. The RFPM measures the energy[27–28] and nutrient intake of adults[27] with error of only 3.7% over six days in free-living conditions compared to gold standard methods.[27] Importantly, and unlike self-report methods, RFPM quantifies food selection and plate waste with food intake calculated as the difference between food intake and plate waste. The method also does not result in reactivity or alteration of energy intake during the period of observation.[27] Hence, this method has established procedures for the collection and analysis of food waste data.

Following these developments in measuring nutritional intake, an emerging alternative for measuring food waste involves respondent use of digital photography.[6–7,29–32] van Herpen and van der Lans[6] demonstrate that human raters can consistently assess the weight of wasted food from respondent photographs, consistent with previous research in the nutrition field.[27–28, 33–34] However, van Herpen et al.[7] emphasize that little is known about other sources of potential bias (e.g., underreporting) and that the cost and scalability of digital photography may limit use.

We assess the validity, time burden, and participant satisfaction associated with sel-fadministered food photography as a means for measuring household food waste as implemented via the FoodImage smartphone app, and we compare these metrics to two variants of household food waste diaries. We hypothesize that food waste will be equivalent (within 20 grams) when measured with FoodImage vs. direct weights of food waste, while food waste measured with the diary methods will not be equivalent to directly weighed food waste. We also hypothesize that the error from FoodImage will be significantly smaller than error from diary methods.

Methods and Materials

Twenty-four adults (22 women, 65.2% Caucasian, Age 18 – 59 y) from the Baton Rouge, Louisiana area participated. Sample size was determined by power analysis (see supporting information). Inclusion criteria included: age 18–65 years, body mass index 18.5 – 50 kg/m^2 (based on self-reported height and weight), conducted some of the household's food shopping and preparation, possessed an iPhone and operable Apple ID/password, and affirmed willingness to complete all study procedures. Exclusion criteria included: persons who were severely immunocompromised or pregnant. The predominance of female participants is expected given the inclusion criteria concerning household food preparation and that U.S. women spend more than twice the time in household meal preparation as men.[35]

Relevant internal institutional review boards approved the research design. Respondents provided written informed consent before enrollment. Data collection followed guidelines for ethical treatment and good clinical practice. Participants received $30 for the study.

Participants used three methods – the FoodImage app, food diary with visual estimation (diary: visual estimation), and food diary with scale (diary: with scale). Each method is fully detailed in the supporting information. Briefly, for the FoodImage app, staff verified proper installation and then instructed participants how to take photos of food and waste items within the app, which included instructions to place a standardized visual reference card in each photo (supporting information including figures S1–S4). Respondents were instructed how to verify photo quality, to add informational tags to photos concerning the source, quantity, destination, reason for waste, and any details concerning reason for waste, and to ensure data was transmitted.

For diaries, respondents were provided with a formatted color-coded data collection sheet and instructed how to record a waste item's description, source, quantity, destination, reason for waste, and any details concerning reason for waste (figures S5 and S7). A different sheet with distinct coloring was provided for each task and diary type (visual estimation and with scale). Respondent instructions were identical for both diary approaches except for the quantity estimation method. Visual estimation instructions provided standard analogies (e.g., a fist approximates one cup; figure S6) while scale instructions encouraged proper use and interpretation (e.g., reminders to record measurement units; figure S8). Participant training was matched across all methods with respect to time and intensity. For each method, participant training concluded when they showed mastery.

Respondents used each method to perform three tasks. First, participants measured food waste created during a simulated meal preparation setting (*Prep*) with foods that included edible and inedible parts. Second, they measured food waste created during simulated eating conditions with plate waste (*Eat*). The third task involved measuring food waste created during simulated cabinet and refrigerator clean-out due to the discarding of spoiled foods, *etc*. (*Toss*). In a fourth task (reported elsewhere as it was not part of the primary outcome data) participants record items and prices from a food shopping trip using a diary and via the image capture features of FoodImage.

Lab personnel directly weighed all foods to provide the criterion value (i.e., the directly weighed amount for establishing method accuracy/validity). Lab personnel also recorded the amount of time each participant took to complete each task using each method. Each task was conducted in a lab kitchen at the Pennington Biomedical Research Center (PBRC). Each respondent completed all training, tasks, and surveys during a single session, which ended with a survey that asked participants to indicate whether the app or diary approach (1) was their preferred method for recording data (*user preferred method*), (2) saved them the most time during measurement tasks (*perceived lower time burden*), and (3) was more accurate during the tasks (*perceived more accurate*). These three questions did not distinguish between the two diary types and hence represents a ranking of the app versus diary approaches in general. Twenty-four total sessions (one per respondent) were held during May – July 2018.

Each session featured three forms of randomization. First, the order of using the app and food diaries was randomized across participants, though participants always used the food diary with scale after using the food diary with visual estimation (without scale). Note the app provided no feedback to the participant concerning the amount of food wasted nor did it require participants to estimate quantities. Second, respondents were randomly assigned to one of several meals for the meal preparation and eating settings (pizza, hamburger, spaghetti, chicken, pork chop and salad-focused meals – see supplemental materials for more details) to improve robustness of results by ensuring that waste during food preparation differed in its suitability for composting and/or discard via a disposal and that plate waste differed in suitability for leftovers and/or discard via composting, garbage disposal, feeding pets, or garbage. Third, the amount of waste per meal was randomly assigned and blinded from the respondents and blinded from the research staff tasked with estimating waste amounts from images captured using the app. The amount of food waste created for each task and method was assigned randomly using an exponential distribution with the minimal being 0% remaining (i.e., no waste) and the maximum being 74%. The randomization procedure accounts for the order of measurement approaches so that the mean amount of waste for each task does not differ by measurement method.

Data Preparation and Analysis

Data from the FoodImage app were prepared for analysis using methods very similar to the RFPM.[27–28] Images captured through the app were viewed by PBRC's nutrition staff who identified nutrient matches (Standard Reference 28)[36] and estimated the mass of portions of food prepared, selected, consumed, returned, and discarded at each stage.

Data from hand-written diaries were prepared for analysis via double data entry. Nutrient values and weights were assigned by PBRC's nutrition staff from the US Department of Agriculture's Standard Reference 28.[36]

One respondent, the first participant, was excluded from analyses of measurement error (though not from analyses of perceptions of methods) because appropriate procedures were not followed by study staff and, thus, data from all methods were not suitable for inclusion into analyses concerning method accuracy. The first set of method accuracy analyses includes data from all remaining participants (*All Participants*). The second set of method accuracy analyses removed data from two participants who had extreme values for one of the methods (diary with scale) as detailed in the results section. For lack of a better term, results that exclude these two participants' data are labeled as *Adherent Participants*.

Data were analyzed to assess:

1. Differences in the frequency of participant adherence to method instructions such that data were suitable for study inclusion.
2. Differences in the frequency of missed (i.e., failing to record an item) or phantom (i.e., reporting an item when none existed) items by method.
3. Equivalency to the criterion value[37] for each task and method. Equivalency was assigned to be ±20 grams (0.71 ounces) and ±50 Calories of the criterion value. These values were chosen prior to analysis and represent small but realistic bands for assessing the method's accuracy without demanding perfection.[37] Any item not recorded by a participant with a given method is coded as having a measured value of zero and is included in the equivalency analysis.
4. Differences in the mean measurement error between methods for each task, where measurement error is the value recorded by the experimental method minus the criterion value.
5. Differences in the mean absolute error (MAE) between methods for each task, where MAE is the absolute value of the measurement error.
6. Differences between the variance of the criterion values and the variance of food waste estimates for each method and task, and differences in the variance of food waste estimates between methods for each task.
7. Differences in the mean amount of time respondents spent completing measurements between methods by task.
8. Differences in the frequency of user preference across measurement methods.

The differences in measured values from criterion values and the differences between measured values for each method pair was evaluated for adherence to the normal distribution and determined to be sufficiently normal to use parametric testing for continuous variables. Associations were tested with a Fisher's Exact test for (1) and (2). Two-sided *t*-tests were used for (3), (4), (5) and (7), and an *F*-test was used for (6). A binomial test was used for (8). Statistical significance was set at 5%. Secondary analyses for each task and method include: (1) regression analyses of how measurement error and MAE is related to the criterion value, and (2) a *t*-test of whether the measurement error differs by race or age category.

Results and Discussion

A first assessment is the percent of participants that failed to adhere to instructions to such an extent that the resulting data make subsequent analyses impractical. Two of the twenty-three participants (9%) produced extreme values for some foods with the diary with scale method, while no participants were classified as such for the other two methods. While the frequency of participant adherence is not statistically different across the methods ($p>0.10$, Fisher's Exact test), the loss of data from two participants reduced the number of observations suitable for analysis from 317 to 290 (by 27 observations or 8.5%).

The next point of assessment is whether participants recorded all relevant items. Among all participants (top half of Table 1), diary methods missed several items during the eating and tossing tasks, with the difference being statistically significant for *Toss* (both diary approaches resulted in significantly more missed items than FoodImage, which had zero missed items, <DI>, p=0.01). If a week involved 14 preparation events with 4 items per event, 28 eating events with 3 items per event, and 2 clean-out events with 5 items per event, this would represent 150 total items requiring recording. Translating the per-task rates of capture to a weekly total, it suggests that a diary with visual estimation would miss 4.2% of items while a diary with scale would miss 2.8%. Similar patterns and magnitudes of missed items were observed adherent participants (bottom half of Table 1). We also note that participants using the diary methods also recorded phantom items that were not present during *Prep* (n=2) and *Toss* tasks (n=2) (not listed in table), while FoodImage resulted in no phantom items. The frequency of phantom item appearance was not significantly different across methods (p>0.10, Fisher's Exact test).

The accuracy of each approach was quantified by calculating its measurement error (criterion value for each item minus the value estimated by participants using each method, see Figure 1). We do so for all participants (top half - Table 1) and for the subset of adherent participants (bottom half - Table 1).

During the *Prep* task, the FoodImage app (but neither diary approach) were equivalent to the criterion value (i.e., the 95% confidence intervals did not contain 20 or −20, which is noted with a '*' next to the bracketed confidence interval in Table 1). The measurement error was significantly different across the three methods (i.e., none of the values in the same row share a superscript lowercase letter). Among adherent participants (bottom half - Table 1) both FoodImage and diary with scale are equivalent to the criterion value (both feature a '*' next to their bracketed confidence interval), and differences in measurement error among the three methods become non-significant (all share the same superscript letter).

For the *Eat* task, only the FoodImage and diary with visual estimation approaches were equivalent to the criterion value among all participants and the measurement error was significantly different across all the methods. When only adherent participants are considered, all three methods are equivalent to the criterion value, and measurement error no longer differed significantly between methods.

For the *Toss* task, no method was equivalent to the criterion value and each approach's measurement error was statistically different from one another with the FoodImage averaging less than the criterion value and the two diary methods averaging more than the criterion values. The same was true among adherent respondents. *Toss* proved to be the most difficult measurement task, perhaps due to product packaging as research staff struggled to identify the contents of opaque packages from participant pictures captured with the FoodImage app while the packages inflated user estimates derived from the two diary methods.

Among all respondents across the three tasks, FoodImage consistently yielded an underestimate of criterion values while the diary methods consistently yielded an overestimate. Among adherent participants, the only change was that the diary with scale yielded an underestimate for *Prep* and an overestimate for the other two tasks. When all participants are considered, diary with scale yields overestimates for all tasks and errors that were several orders of magnitude greater than the errors generated by the other methods, which arises because participants incorrectly assigned units when recording waste. The incorrect assignment of units occurred inconsistently, however, such that a researcher receiving similar diary data would be unable to simply recode all entries, e.g., from kg to g.

The mean absolute error (Figure 2) reveals a similar (though not identical) pattern of measurement error as the equivalency results depicted in Figure 1 and Table 1. When all participants are considered, FoodImage yields the smallest MAE for each task with the difference being significant against diary with visual estimation for all tasks and significantly different for the *Toss* task against diary with scale (see Table S2 for all test statistics supporting lower case letters in Figure 2). When only adherent participants are considered, FoodImage and diary with scale are never significantly different from one another, though both of these methods have significantly smaller MAE than diary with visual estimation for the *Prep* and *Toss* tasks.

Hence, among adherent participants, the comparison of error across methods becomes much closer, with the diary with scale approach, which theoretically should be the most accurate, matching FoodImage in terms of being equivalent to the criterion values for two of the three tasks (*Prep* and *Eat*) and matching FoodImage on all tasks in terms of MAE. However, this requires dropping 4 *Prep* items (7%), 7 *Eat* items (10%), and 16 *Toss* items (8%), respectively, when dropping the 2 non-adherent participants. Furthermore, even once these two participants were removed, the adherent participants using the diary methods still failed to record more than 3% of individual items presented to them in the lab setting. Combining non-adherence and missed items among adherent participants, the diary with scale yielded 37 (11.7%) fewer usable data points than FoodImage. The diary with visual estimation yielded 14 (4.4%) fewer usable data points than FoodImage, which was attributable only to missed individual items. As noted earlier, additional error is introduced by both diary methods due to the inclusion of phantom food items that were reported but not actually present.

When comparing the variance of values encoded by each measurement method to the variance of the accompanying criterion values for each task (Table 1), we find that FoodImage yielded variance that was statistically similar to the variance of the criterion values for all tasks whether all participants or only adherent participants were considered. Among all participants, the diary approaches yield variances that were significantly greater than the variances from the criterion values in all but one case (diary with visual estimation for the *Prep* task). Among adherent participants, the diary approaches yield variances that were statistically similar to the variances from the criterion values in about half of the tasks. In particular, and in line with results from the error measures discussed above, the diary with scale approach yield variances that are several orders of magnitude larger than the other two measures when all participants are considered, but remain consistently larger even when only the adherent subset is used (though only one difference is statistically significant).

Secondary analyses (Table 2) reveal that, for three of nine conditions, measurement error becomes more downward biased as food waste increases. This is observed from the significant negative regression coefficient for both the FoodImage app and diary with scale among adherent participants for the *Toss* task (top panel, Table 2) and for the significant negative regression coefficient for the *Prep* task for the diary with visual estimation. This increase in downward bias with item mass can be seen in Figure S9, which depicts all measurement error among adherent participants for the *Toss* task.

The MAE consistently increases with criterion weights, which can be observed as eight of the nine regression coefficients are positive and significant in the second panel of Table 2. Training for the FoodImage app and diary methods should focus on larger items as the mean absolute value of the deviation between the criterion and measured value tends to increase with the criterion value (i.e., for larger items).[1]

While measurement accuracy and consistency are essential, each measurement method requires participant time and engagement. Reducing respondent time burden and providing respondents with a satisfying user experience promotes its use and, potentially, supports consistent data collection over longer periods of time. A key manifestation of respondent burden is the time spent conducting measurement (Table 3, top panel). Measurement with the FoodImage app took participants significantly less time than the diary with visual estimation or the diary with scale during *Prep* (25 and 32 s, respectively), *Eat* (19 and 26 s, respectively), and *Toss* tasks (106 and 137 s, respectively), where each of these differences is statistically significant (see Table S3 for test statistics). If a week involved 14 preparation events, 28 eating events, and 2 clean-out events, this would represent more than 18 and 24 minutes less time (33% and 43% less time) using FoodImage rather than the diary with visual estimation and scale methods, respectively.

When directly asked to compare the two overarching measurement approaches (FoodImage vs. pen and paper diaries) all respondents chose FoodImage when asked “Which method did you prefer for recording the foods you eat and throwaway?” and when asked which method saved the most time in measuring waste. The majority (69.6%) perceived that the app was more accurate for measurement purposes.

Conclusions

Participant measures of food waste using the FoodImage app are equivalent to the gold standard measures obtained by direct researcher measurement for two of the three tasks. The diary with scale method also featured equivalency on two of the three tasks, but only after two participants (8.5% of sample) with extreme values on the diary with scale method were eliminated from the data. The diary with visual estimation method yielded equivalency on only one task. Similar patterns emerge from examination of mean absolute error by method. Further, the measurement bias for the one non-equivalent task is predictable for the FoodImage app, yielding promising avenues for enhanced training and implementation. The

[1]Secondary analysis also suggests that accuracy features little association with respondent characteristics (age or race, bottom of Table 2) regardless of task or method. Furthermore, accuracy is largely unchanged when the analysis is conducted in Calories rather than grams (Table S1).

FoodImage app is also perceived more favorably by participants as it is preferred in direct comparison, perceived to incur a lower time burden, and perceived to provide greater accuracy.

This study begins to address van Herpen et al.'s[7] question about whether photo-based data collection methods will suffer from underreporting in the realm of food waste. FoodImage captured all items presented to respondents, and no respondents had to be removed from data analyses due to extreme values. In comparison, even adherent diary users omitted a significant percent of items during the *Toss* task even within the limited confines of a laboratory setting immediately following training and in the presence of laboratory staff. We hypothesize that the number of items omitted in diary-based food waste data collections will increase in free-living (non-lab) settings and may help explain previous reports of that diaries provide lower estimates of food waste than waste stream analysis.[19–20]

This positions the FoodImage app as a promising new tool for measuring food waste that is preferred by participants over more traditional self-report methodologies. However we note this study features several limitations. While the sample provided enough power to conduct our planned analyses, we look forward to future validations featuring larger samples that include more geographic and gender diversity. Instructions provided to participants on app usage continue to be refined, which could reduce the error associated with measurement, particularly for *Toss* incidents. Further work is also needed to streamline the process by which research staff interpret and encode photos to reduce per-participant costs, which are currently higher for FoodImage than for diaries.

Our analysis also provides insight for practitioners choosing between the two diary variants. Providing a scale yields improved accuracy over visual estimation without a significant change in the number of missing items, though this comes at the cost of removing a fraction of users who fail to accurately record scale weights. Importantly, identifying extreme or invalid weights in 'real-world' settings is particularly challenging since large amounts of food waste can be generated when preparing food; cleaning up after family meals; and cleaning out the refrigerator, freezer, or cabinets. Indeed, in our study, the participants that produced extreme values appeared to use the diary with scale method correctly but simply recorded units incorrectly (e.g., g, kg) but did so inconsistently, negating the ability to simply recode the entries in a real-world situation. We had the luxury in this study of being able to identify extreme values as being gross errors since the criterion values were known, but in real world conditions some extreme values will reflect real and accurate data (e.g., discarding ten pounds of rotting potatoes).

In addition, user time burden is estimated to increase by about 6 minutes (7.7%) a week when using a scale versus visual estimation while data collection costs will increase because scales must be purchased. Moreover, the ability of participants to carry and use a scale when outside the home significantly limits the feasibility and use of this approach. Hence, researchers must consider the tradeoff between the enhanced accuracy provided by diaries with scales against the higher rates of usable data and lesser time burden and cost associated with diaries reliant on visual estimation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

The authors would like to thank H. Raymond Allen for his assistance. Also we appreciate the participants' time and effort. Funding: This work is supported in part by the following grants: 2017–6702326268 from the USDA National Institute of Food and Agriculture, U54 GM104940 from the National Institute of General Medical Sciences of the National Institutes of Health, which funds the Louisiana Clinical and Translational Science Center, and NORC Center Grant # P30DK072476 entitled "Nutrition and Metabolic Health Through the Lifespan" sponsored by NIDDK. The content is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. Department of Agriculture or the National Institutes of Health.

References

1. EU Platform on Food Losses and Food Waste: Term of Reference (ToR). European Commission 2016 https://ec.europa.eu/food/sites/food/files/safety/docs/fw_eu-actions_flwplatform_tor.pdf (accessed: 27 October 2019).
2. USDA and EPA Join with Private Sector, Charitable Organizations to Set Nation's First Food Waste Reduction Goals. U.S. Department of Agriculture, Office of Communications 2015 9 16 https://www.usda.gov/media/press-releases/2015/09/16/usda-and-epa-join-private-sector-charitable-organizations-set (accessed 1 December 2019).
3. Griffin M; Sobal J; Lyson TA An analysis of a community food waste stream. Agriculture and Hum. Values, 2009, 26(1–2), 67–81.
4. Xue L; Liu G; Parfitt J; Liu X; Van Herpen E; Stenmarck Å; O'Connor C; Östergren K; Cheng S Missing food, missing data? A critical review of global food losses and food waste data. Environ. Sci. Technol, 2017, 51(12), 6618–6633. DOI 10.1021/acs.est.7b00401 [PubMed: 28492315]
5. Porpino G Household food waste behavior: avenues for future research. J. of the Association for Consumer Res, 2016 1(1), 41–51.
6. van Herpen E; van der Lans I A picture says it all? The validity of photograph coding to assess household food waste. Food Quality and Preference, 2019, 75, 71–77. DOI 10.1016/j.foodqual.2019.02.006
7. van Herpen E; van der Lans IA; Holthuysen N; Nijenhuis-de Vries M; Quested TE Comparing wasted apples and oranges: An assessment of methods to measure household food waste. Waste Management, 2019, 88, 71–84. DOI 10.1016/j.wasman.2019.03.013 [PubMed: 31079652]
8. Elimelech E, Ert E and Ayalon O Exploring the drivers behind self-reported and measured food wastage. Sustainability, 2019, 11(20), 5677 DOI 10.3390/su11205677
9. Dhurandhar NV, Schoeller D, Brown AW, Heymsfield SB, Thomas D, Sørensen TI, Speakman JR, Jeansonne M, Allison DB and Energy Balance Measurement Working Group. Energy balance measurement: When something is not better than nothing. Int. J. of Obesity, 2015, 39(7), 1109–1113.
10. Schoeller DA, Thomas D, Archer E, Heymsfield SB, Blair SN, Goran MI, Hill JO, Atkinson RL, Corkey BE, Foreyt J and Dhurandhar NV Self-report–based estimates of energy intake offer an inadequate basis for scientific conclusions. The Am. J. of Clin. Nutr, 2013, 97(6), 1413–1415. [PubMed: 23689494]
11. Van de Mortel TF Faking it: Social desirability response bias in self-report research. Aust. J. of Adv. Nurs, 2008, 25(4), 40–48.
12. van Herpen E, van Geffen L, Nijenhuis-de Vries M, Holthuysen N, van der Lans I and Quested T A validated survey to measure household food waste. MethodsX, 2019, DOI 10.1016/j.mex.2019.10.029
13. Delley M and Brunner TA Household food waste quantification: Comparison of two methods. British Food J, 2018, 120(7), 1504–1515. DOI 10.1108/BFJ-09-2017-0486

14. Giordano C, Alboni F and Falasconi L Quantities, determinants, and awareness of households' food waste in Italy: A comparison between diary and questionnaires quantities. Sustainability, 2019, 11(12), 3381 DOI 10.3390/su11123381
15. Lebersorger S and Schneider F Discussion on the methodology for determining food waste in household waste composition studies. Waste Management, 2011, 31(9–10), 1924–1933. DOI 10.1016/j.wasman.2011.05.023 [PubMed: 21705207]
16. Ramukhwatho F, duPlessis R, Oelofse S Preliminary drivers associated with household food waste generation in South Africa. Appl. Environ. Educ. Commun. 2017, 17, 254–265. DOI 10.1080/1533015X.2017.1398690
17. Elimelech E, Ayalon O and Ert E What gets measured gets managed: A new method of measuring household food waste. Waste Management, 2018, 76, 68–81. DOI 10.1016/j.wasman.2018.03.031 [PubMed: 29576512]
18. World Resources Institute. 2016 Food Loss and Waste Accounting and Reporting Standard: Version 1.0. Washington, DC Available online: http://www.wri.org/sites/default/files/REP_FLW_Standard.pdf (access August 9, 2016)
19. Quested TE, Parry AD, Easteal S, & Swannell R Food and drink waste from households in the UK. Nutr. Bull, 2011, 36(4):460–467.
20. Hoover D, 2017 Estimating quantities and types of food waste at the city level. Natural Resources Defense Council https://www.nrdc.org/sites/default/files/food-waste-city-levelreport.pdf (accessed 1 December 2019).
21. Bandini LG, Schoeller DA, Cyr HN, Dietz WH. Validity of reported energy intake in obese and nonobese adolescents. Amer. J. of Clin. Nutr, 1990, 52(3), 421–425. DOI 10.1093/ajcn/52.3.421 [PubMed: 2393004]
22. Beaton GH, Burema J, Ritenbaugh C. Errors in the interpretation of dietary assessments. Amer. J. of Clin. Nutr, 1997 65(4), 1100S–7S DOI: 10.1093/ajcn/65.4.1100S
23. 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. of the Amer. Dietetic Assoc. 2000, 100(7), 777–83. DOI: 10.1016/S0002-8223(00)00227-3
24. Beasley J, Riley WT and Jean-Mary J Accuracy of a PDA-based dietary assessment program. Nutr., 2005, 21(6), 672–677. DOI: 10.1016/j.nut.2004.11.006
25. Goris AH, Westerterp-Plantenga MS and Westerterp KR Undereating and underrecording of habitual food intake in obese men: Selective underreporting of fat intake. The Amer. J. of Clin. Nutr, 2000, 71(1), 130–134. [PubMed: 10617957]
26. Goris AH and Westerterp KR Underreporting of habitual food intake is explained by undereating in highly motivated lean women. The J. of Nutr, 1999, 129(4), 878–882. [PubMed: 10203564]
27. Martin CK, Correa JB, Han H, Allen HR, Rood JC, Champagne CM, Gunturk BK, Bray GA Validity of the Remote Food Photography Method (RFPM) for estimating energy and nutrient intake in near real -time. Obesity. 2012, 20(4), 891–9. DOI: 10.1038/oby.2011.344. [PubMed: 22134199]
28. Martin CK, Han H, Coulon SM, Allen HR, Champagne CM, Anton SD 2009. A novel method to remotely measure food intake of free-living individuals in real time: The remote food photography method. British J. of Nutr 2009, 101(3), 446–56. DOI: 10.1017/S0007114508027438
29. Roe BE, Apolzan JW, Qi D, Allen HR and Martin CK Plate waste of adults in the United States measured in free-living conditions. PloS One, 2018, 13(2), p.e0191813.
30. Farr-Wharton G, Foth M, Choi JHJ, 2012 Colour coding the fridge to reduce food waste. In: Proceedings of the 24th Australian Computer-Human Interaction Conference, pp. 119–122.
31. Farr-Wharton G, Foth M, Choi JH. Identifying factors that promote consumer behaviours causing expired domestic food waste. J. of Cons. Behav. 2014, 13(6), 393–402. DOI: 10.1002/cb.1488
32. Porpino G, Parente J, Wansink B Food waste paradox: Antecedents of food disposal in low income households. Int. J. Consum. Stud, 2015, 39, 619–629. DOI 10.1111/ijcs.12207
33. Williamson DA, Allen HR, Martin PD, Alfonso AJ, Gerald B, Hunt A Comparison of digital photography to weighed and visual estimation of portion sizes. J. Am. Diet. Assoc. 2003, 103(9), 1139–1145. [PubMed: 12963941]

34. Williamson DA, Allen HR, Martin PD, Alfonso A, Gerald B, Hunt A Digital photography: A new method for estimating food intake in cafeteria settings. Eat. Weight. Disord. 2004, 9(1), 24–28. [PubMed: 15185830]

35. Hamrick KS, McClelland K 2016 Americans' Eating Patterns and Time Spent on Food: The 2014 Eating & Health Module Data, EIB-158, U.S. Department of Agriculture, Economic Research Service, July.

36. SR28 US. National Nutrient Database for Standard Reference, Release 28. US Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory [cited on: 26 November 2019] Available from: http://www.ars.usda.gov/ba/bhnrc/ndl

37. Meyners M Equivalence tests–A review. Food Quality and Preference, 2012, 26(2), 231245.

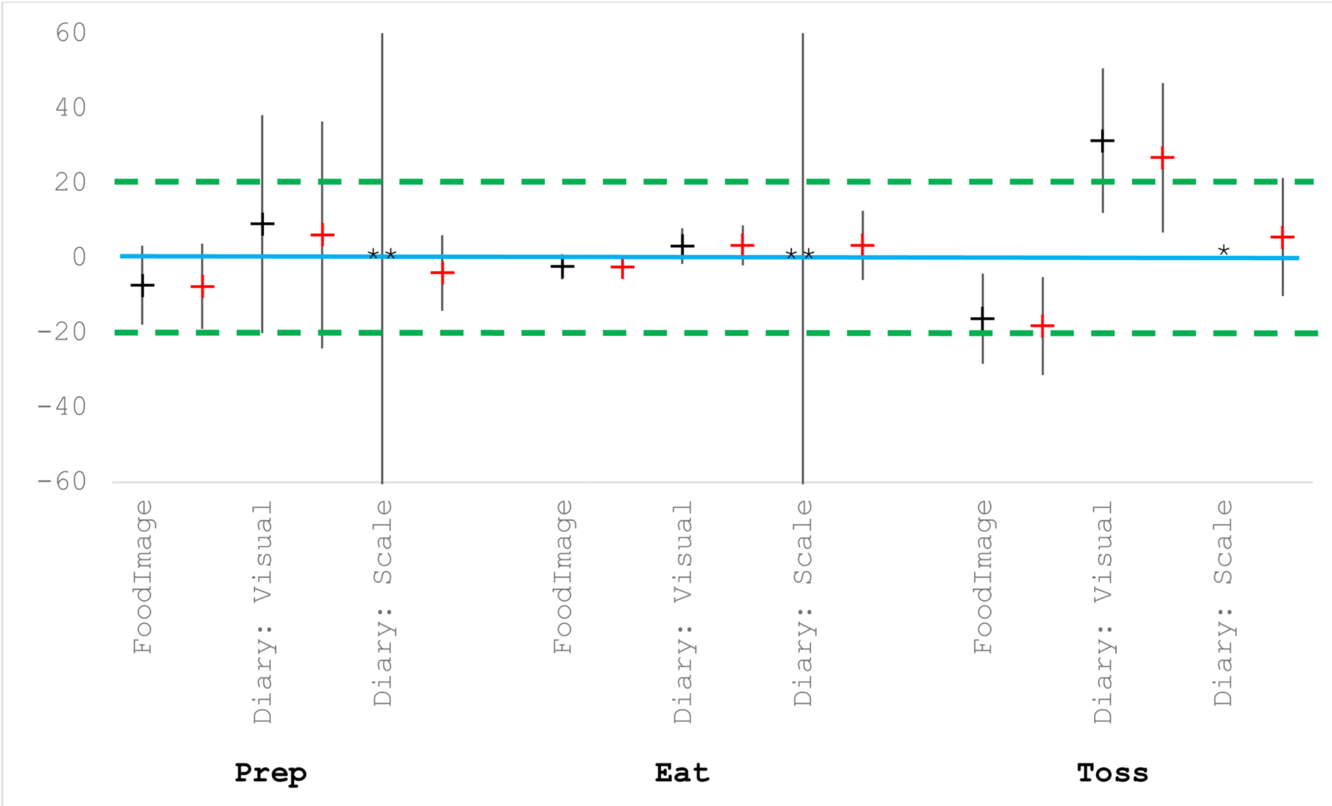

Figure 1.

Mean error with 95% confidence interval bars by measurement method and task for all participants (black cross) and adherent participants (red cross). *Notes*: 20 g equivalency band depicted with green dashed lines. **The 95% confidence interval for diary with scale confidence interval extends in both directions beyond the region depicted on the graph. *The mean and entire 95% confidence interval for diary with scale lies above the region depicted on the graph.

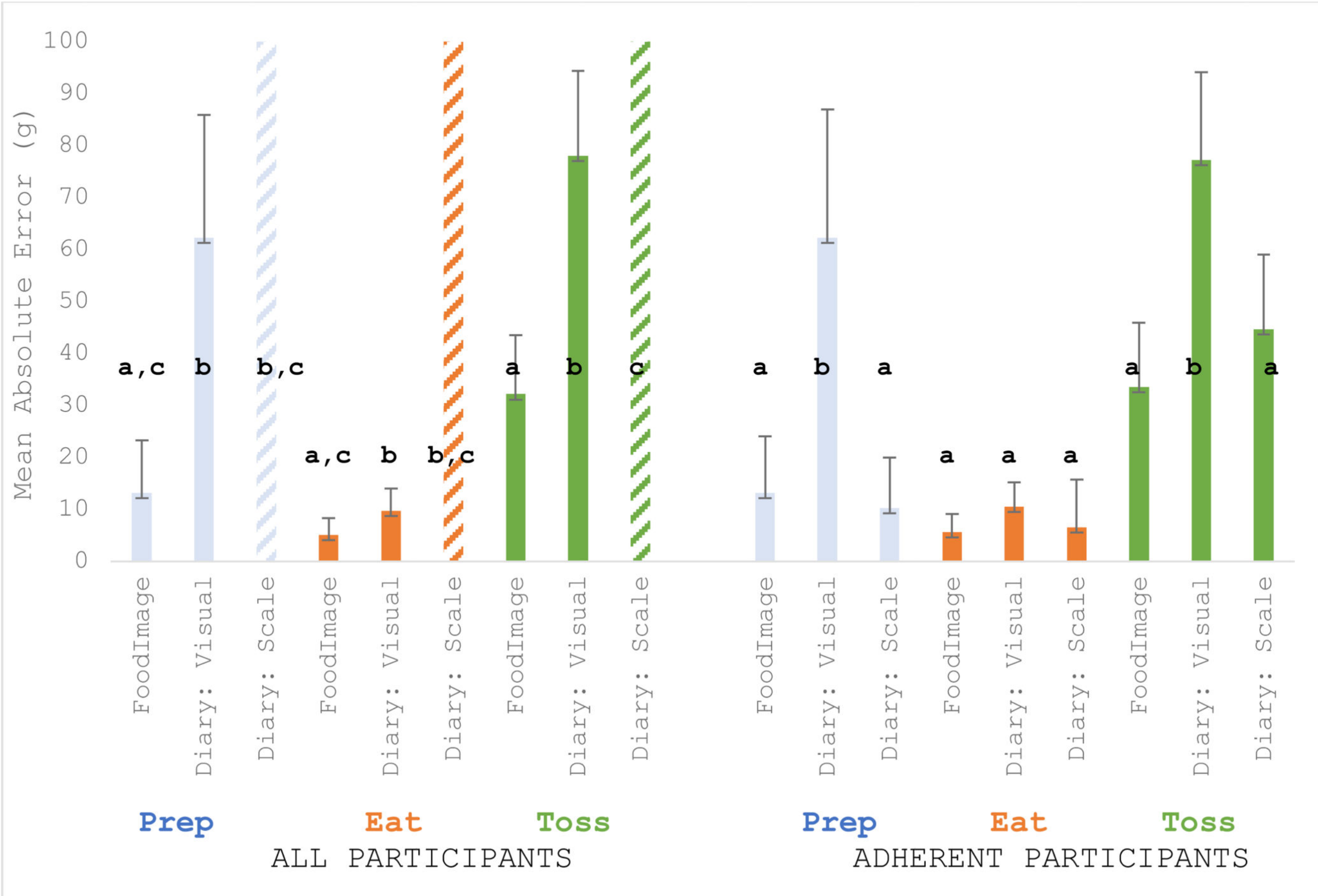

Figure 2.

Mean Absolute Measurement Error by Method and Task. Whiskers are 95% confidence intervals. Bars within the same task cluster that share a letter are not significantly different at the 5% level. Cross-hatched bars extend outside the graph's range.

Table 1.

Item Coverage and Measurement Accuracy by Method

	FoodImage	Diary: Visual Estimation	Diary: Scale
All Participants			
Items Missed	% Items Not Measured		
Prep (N=54)	0[a]	0[a]	0[a]
Eat (N=73)	0[a]	6.8[a]	5.5[a]
Toss (N=190)	0[a]	4.7[b]	3.2[b]
Accuracy (g)	Measurement error: deviation from criterion value (g) (standard deviation) [95% confidence interval]		
Prep (N=54)	−7.33[a] (38.67) [−17.88, 3.23]*	9.07[b] (106.60) [−20.01, 38.16]	4232.0[c] (21,718.2) [−1695.9, 10,159.9]
Eat (N=73)	−2.30[a] (14.50) [−5.68, 1.08]**	3.18[b] (20.72) [−1.65, 8.02]*	536.7[c] (2730.6) [−100.4, 1173.8]
Toss (N=190)	−16.25[a] (84.72) [−28.38, −4.13]	31.34[b] (134.80) [12.04, 50.63]	20,553.6[c] (111,151) [4647.1, 36,460.1]
Variance	Variance of measured values (variance of criterion values)		
Prep (N=54)	9227.0[a] (9499.5)	9787.7[a] (9499.5)	472,062,529[c] (9499.5)**
Eat (N=73)	842.1[a] (746.6)	1363.3[b] (746.6)**	7,455,630[c] (746.6)**
Toss (N=190)	34,596[a] (35,419)	48,312[b] (35,419)**	12,370,778,176[c] (35,419)**
Adherent Participants			
Items Missed	% Items Not Measured		
Prep (N=50)	0[a]	0[a]	0[a]
Eat (N=66)	0[a]	6.1[a]	6.1[a]
Toss (N=174)	0[a]	5.2[b]	3.4[b]
Accuracy (g)	Measurement error: deviation from criterion value (g) (standard deviation) [95% confidence interval]		
Prep (N=50)	−7.63[a] (40.00) [−19.00, 3.74]*	6.20[a] (107.00) [−24.21, 36.61]	−3.99[a] (35.60) [−14.11, 6.12]*
Eat (N=66)	−2.46[a]	3.36[a]	3.35[a]
	FoodImage	Diary: Visual Estimation	Diary: Scale
	(15.24) [−5.59, 0.67]*	(21.77) [−1.99, 8.71]*	(37.79) [−5.94, 12.65]*
Toss (N=174)	−18.19[a] (87.76) [−31.33, −5.06]	26.81[b] (134.00) [6.76, 46.85]	5.52[c] (105.50) [−10.25, 21.30]
Variance	Variance of measured values (variance of criterion values)		
Prep (N=50)	9803.40[a] (10,080.16)	8828.74[a] (10,080.16)	10,920.25[a] (10,080.16)

	FoodImage	Diary: Visual Estimation	Diary: Scale
Eat (N=66)	916.63[a] (808.42)	1480.68a,b (808.42) **	2144.89[b] (808.42) **
Toss (N=174)	28,056.25[a] (30,450.25)	41,861.16[b] (30,450.25) **	35,268.84[a,b] (30,450.25)

Notes: shared superscript letters in the same row denote figures (means or variances, depending on the row) that are not significantly different from one another.

* denotes that the measurement method is equivalent to the criterion value, which was defined prior to analysis as ± 20 g of the weight obtained by research staff in the test kitchen.

** denotes the variance of the experimental measurement values differ significantly from the variance of the criterion values.

Table 2.

Secondary Analyses of Accuracy by Method among Adherent Participants

	FoodImage	Diary: Visual Estimation	Diary: Scale
Measurement Error vs. Criterion Weight	Regression Slope (standard error) [*t*-value]		
Prep (N=50)	−0.58 (0.35) [−1.66]	−0.59 (0.11) [−5.39]*	−0.17 (0.41) [−0.42]
Eat (N=66)	−0.23 (0.22) [−1.08]	0.26 (0.16) [1.65]	−0.02 (0.09) [−0.25]
Toss (N=174)	−0.62 (0.14) [−4.41]*	−0.15 (0.10) [−1.54]	−0.25 (0.12) [−2.10]*
Absolute Value of Measurement Error vs. Criterion Weight	Regression Slope (standard error) [*t*-value]		
Prep (N=50)	1.14 (0.34) [3.36]*	0.95 (0.09) [10.01]*	1.38 (0.37) [3.71]*
Eat (N=66)	1.34 (0.16) [8.38]*	0.95 (0.13) [7.08]*	−0.01 (0.09) [−0.07]
Toss (N=174)	0.57 (0.15) [3.79]*	0.23 (0.11) [2.11]*	0.49 (0.13) [3.63]*
Variance of Measurement Error			
Prep (N=50)	1600.01[a]	11,449.00[b]	1267.10[a]
Eat (N=66)	232.39[a]	474.04[b]	1428.26[c]
Toss (N=174)	7702.50[a]	17,956.00[b]	11,130.30[c]
Measurement Error by Respondent Age	Difference: (25 or younger) - (older than 25) (Standard Error)		
Prep	−29.07 (24.18)	−63.61 (38.00)	−20.02 (20.31)
Eat	−2.44 (2.94)	1.12 (5.95)	−9.59 (9.77)
Toss	−2.37 (17.66)	−29.88 (27.12)	0.74 (19.85)
Measurement Error by Respondent Race	Difference: (not Caucasian) - (Caucasian) (Standard Error)		
Prep	4.03 (27.24)	−77.57 (37.96)	8.79 (22.45)
Eat	0.79 (3.27)	−9.88 (6.26)	−7.84 (11.08)
Toss	10.18 (19.03)	−43.18 (28.69)	−33.99 (20.21)

Notes:

* denotes the top value in the cell (regression slope or between-group difference) is significantly different from zero at the 5% level. Figures in the same row that share the same superscript letter are not significantly different at the 5% level.

Table 3.

Time Burden and User Perceptions Among All Participants

	FoodImage	**Diary: Visual Estimation**	**Diary: Scale**
Time Burden (s)	Mean (standard deviation) N		
Prep	69.1[a] (37.3) 45	94.1[b] (45.3) 45	101.0[b] (43.5) 45
Eat	74.7[a] (39.4) 44	94.1[b] (42.6) 44	100.5[b] (45.7) 43
Toss	138.4[a] (66.1) 44	244.3[b] (66.1) 44	277.4[c] (69.4) 43
User Perceptions ***			
User preferred method	100.0[a]	0.0[b]	
Perceived lower time burden	100.0[a]	0.0[b]	
Perceived more accurate	69.6[a]	30.4[b]	

Notes. Figures in the same row that share a superscript letter are not significantly different from one another.

*** Comparative user perceptions were elicited for FoodImage and for diaries in general (without distinguishing between diaries with scale versus visual estimation). All values represent the percent of the 24 participants who chose the specified approach in response to the question.