



# Validation of the FRESH Austin food frequency questionnaire using multiple 24-h dietary recalls

Christine ES Jovanovic<sup>1,\*</sup> , Jacob Whitefield<sup>2</sup>, Deanna M Hoelscher<sup>3</sup>, Boajiang Chen<sup>3</sup>, Nalini Ranjit<sup>3</sup>  and Alexandra E van den Berg<sup>3</sup>

<sup>1</sup>University of Texas, School of Public Health, 1705 Guadalupe St, 2nd Floor, Austin, TX 78701, USA; <sup>2</sup>Texas Health and Human Services Commission, Austin, TX, USA; <sup>3</sup>University of Texas, School of Public Health, Austin, TX, USA

Submitted 11 August 2020: Final revision received 7 March 2021: Accepted 19 April 2021: First published online 26 May 2021

## Abstract

**Objective:** The purpose of the current study was to examine the validity of an FFQ utilised in the Food Retail: Evaluating Strategies for a Healthy Austin (FRESH Austin) study, designed to evaluate changes in the consumption of fruits and vegetables (FV) in diverse low-income communities in Austin, TX.

**Design:** The FRESH Austin FFQ was validated against three 24-h dietary recalls (24hDR). All dietary assessments were administered (in-person or by telephone) by trained investigators.

**Setting:** Recruitment was conducted at sites within the geographic areas targeted in the FRESH Austin recruitment. People at a community health clinic, a local health centre and a YMCA within the intervention area were approached by trained and certified data collectors, and invited to participate.

**Participants:** Among fifty-six participants, 83% were female, 46% were non-White, 24% had income < \$25 K/year and 30% spoke only/mostly Spanish at home.

**Results:** The FFQ and average of three 24hDR produce similar estimates of average total servings/d across FV (6.68 and 6.40 servings/d, respectively). Correlations produced measures from 0.01 for 'Potatoes' and 0.59 for 'Other Vegetables'. Mean absolute percentage errors values were small for all FV, suggesting the variance of the error estimates was also small. Bland–Altman plots indicate acceptable levels of agreement between the two methods.

**Conclusion:** These outcomes indicate that the FRESH FFQ is a valid instrument for assessing FV consumption. The validation of the FRESH Austin FFQ provides important insights for evaluating community-based efforts to increase FV consumption in diverse populations.

**Keywords**  
Validation  
Food frequency questionnaire  
24hDR

Evaluating dietary outcomes of community-based interventions is challenging, requiring both measurable change at the individual level and the instruments to detect that change. Often, because effects of community-level interventions are broad and diffuse, individual-level outcomes are difficult to capture. Historically, dietary assessment generally has focused on producing precise measures of micronutrients in an effort to discover epidemiological relations with disease outcomes. More recently, however, the focus has shifted to assessing changes in habitual dietary patterns at the food group level (e.g., fruits, vegetables, meat)<sup>(1–3)</sup>. This reflects the intention of community-based interventions to increase/decrease consumption of specific foods or food groups (e.g., increasing consumption of fruits

and vegetables (FV)), rather than targeting a change in micronutrient consumption (e.g., increasing K intake)<sup>(4)</sup>.

Among the variety of dietary assessment methods available, the FFQ may be the best option for community-based programme evaluation, as it captures usual intake in a cost-effective and minimally burdensome process<sup>(5)</sup>. Best practice recommends that FFQ be tailored to both the study aims and the study population, so that the foods queried reflect the outcomes of interest, as well as the culture and usual diet of study participants<sup>(6)</sup>. These adaptations may change the validity of the instrument, however, and ideally instruments should be validated when adapted to new studies and specific study populations<sup>(7,8)</sup>. Thus, the purpose of the current study was to examine the validity

\*Corresponding author: Email christine.e.jovanovic@uth.tmc.edu





of an FFQ utilised in the Food Retail: Evaluating Strategies for a Healthy Austin (FRESH Austin) study, designed to evaluate changes in the consumption of FV in diverse low-income communities in Austin, TX. In alignment with well-established dietary assessment protocols, repeated 24-h dietary recalls (24hDR) serve as the criterion measure<sup>(9–11)</sup>.

## Methods

### ***Food Retail: Evaluating Strategies for a Healthy Austin parent study***

FRESH Austin is a study that evaluates the fresh for less (FFL) initiative, which aims to improve access to healthy, affordable food in ethnically diverse and economically disadvantaged communities through organisational support of local mobile markets in Austin, TX<sup>(12)</sup>. By decreasing barriers to healthy food access, FFL is intended to affect purchasing behaviours and, ultimately, increase the consumption of fresh FV in the target communities. Because increased consumption of fresh FV is the primary outcome of interest, the FRESH Austin survey includes an FFQ that queried only fresh FV, rather than the entire diet. The FRESH Austin survey, which includes the FFQ, is being administered to a cohort of 400 residents over a 3-year time period. This validation study is intended to assess the validity of the FRESH Austin FFQ in comparison to three 24hDR, focusing on assessing the consumption of FV, rather than the entire diet.

### ***Participants***

Because the purpose of the current study was to validate the FRESH Austin FFQ, participants were recruited in a similar way across both studies: by location within the priority areas identified and served by the FFL. Inclusion criteria for both FRESH Austin and the validation study were the same: at least 18 years old, not pregnant or breast-feeding and able to speak English or Spanish.

### ***Data collection protocol***

Recruitment was conducted at sites within the geographic areas targeted in the FFL intervention, which were the same areas in Austin where the FRESH Austin study was situated. People at a community health clinic, a local health centre and a YMCA within the FFL intervention area were approached by trained and certified data collectors, and invited to participate in the validation study. As in the FRESH Austin study, all the materials (consent, information sheet, survey) for the validation study were available in English and Spanish, and trained bilingual research staff conducted the data collection in either Spanish or English, as the participant preferred. In accordance with the approved IRB protocol, participants were given an information sheet, were invited to ask questions, and, if

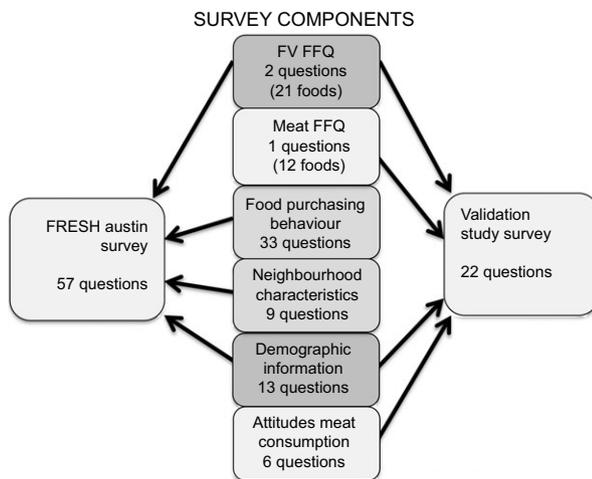
they were willing to participate, signed an IRB-approved consent. All protocols and instruments were approved by the University of Texas Internal Review Board, HSC-SPH-18-0233.

Research shows that correlations between FFQ and reference methods such as repeated 24hDR are higher for interviewer-conducted FFQ, as compared with those that are self-administered<sup>(13,14)</sup>. However, no difference in correlation has been found between FFQ conducted via telephone interview with a qualified researcher and those conducted in-person<sup>(14,15)</sup>. Therefore, the FRESH Austin validation study protocol included either in-person or telephone interviews with trained personnel. This also aligns with the protocol used in the FRESH Austin study for data collection. Using the multiple-pass method developed and validated by the United States Department of Agriculture (USDA)<sup>(16,17)</sup>, three 24hDR were administered to each participant, followed by the FRESH Austin FFQ, at four separate times. At the time of recruitment, the first 24hDR was administered in-person, and arrangements were made for the second 24hDR via telephone interview. At the time of the second 24hDR, a day and time for the third 24hDR was arranged. After three 24hDR were completed, the FRESH Austin FFQ was conducted either in-person or over the phone, depending on the availability and preference of the participant, and incentives were delivered. For both FRESH Austin and this validation study FFQ, all reported food amounts were described in terms of cups, and these amounts were compared with a hand graphic measurement guide, which featured comparisons such as a fist to a one cup portion of vegetables<sup>(18)</sup>. This allowed for a simple and consistent method of estimating portion sizes across both the FFQ and the 24hDR.

In all, sixty-nine people were recruited; five chose not to continue after the first 24hDR interview, six declined after the second and four chose not to complete the final FFQ, resulting in a final sample size of fifty-four. Participants were classified as dropped from the study after four attempts to reach the participant were made, or the participant requested to leave the study. Participants received a total of \$20 in gift cards for completing all four assessments.

### ***Food Retail: Evaluating Strategies for a Healthy Austin validation study survey***

The FFQ in the FRESH Austin survey was adapted from the Block FFQ, a semi-quantitative assessment tool which is used in the NHANES annual survey and has been widely validated<sup>(19–21)</sup>. The foods in the FRESH Austin FFQ were chosen to capture the most commonly consumed FV in the study population (diverse, low-income population in Austin, TX)<sup>(22)</sup>, taking into account local sales data<sup>(23)</sup> as well as feedback from *promotoras* who work in the study communities. The question stems were, 'Over the last month, how many times/month, week, or day did you eat the following fruit/vegetable?' and, 'When you ate the



**Fig. 1** Components of the Food Retail: Evaluating Strategies for a Healthy (FRESH) Austin survey and the FRESH Austin validation study survey

fruit/vegetable, how much did you usually eat?' The FV listed were apples, citrus, bananas, berries, grapes, melon, lettuce, dark leafy greens, broccoli or cauliflower, carrots, tomatoes, avocados, sweet potatoes, potatoes (not sweet), cabbage, peppers, maize, zucchini or other squash, and onions. In addition, respondents were given an option to mention up to four additional fruits and four additional vegetables not included in this listing.

This validation study examines only the FRESH Austin FFQ (Fig. 1), along with pertinent demographic questions. In all, the FRESH Austin validation study survey contained twenty-two questions and was administered in either Spanish or English, as preferred by the participant.

### Twenty-four hour dietary recalls

The USDA five-step multiple-pass 24hDR method has been shown to capture dietary energy and macronutrient intake within 10% of actual intake, as determined by estimated energy requirements and BMR<sup>(24)</sup>. This allows the 24hDR to be used as the 'gold standard' against which the accuracy of the FFQ can be measured<sup>(16,17)</sup>. Our study utilised three 24hDR, conducted using the five-step multiple-pass method, and was guided by scripts adapted from those provided by Nutrition Data Systems for Research (NDSR). To ensure that the data covered the same time period defined in the FRESH Austin FFQ (i.e., the 30 d), three recalls were completed in a period of 30 d, with two recalls of weekdays and one of a weekend day. A visual hand guide to portion sizes<sup>(18)</sup> was employed in all data collections (both FRESH Austin and the validation study) to assist participants in reporting portion sizes. Dietary intake data were entered and analysed using NDSR-2018 software (NDSR version 2008; Nutrition Coordinating Center, University of Minnesota)<sup>(25)</sup>. The NDSR utilises the USDA food composition database, which is maintained by the Agricultural Research Service<sup>(26)</sup>.

### Data preparation

To minimise discrepancies in data entry, all 24hDR records were entered by two trained personnel (M.D. and J.W.), and then crosschecked in their entirety, and disagreements resolved collaboratively. In consultation with the FRESH Austin team, which includes experts in nutritional sciences, final categories of FV were chosen based on congruence between the FFQ and 24hDR. For example, the 'Deep Yellow' vegetable category from the 24hDR was mapped to sweet potatoes and carrots from the FFQ, while the 'Dark Green' category from the 24hDR was mapped to cooked dark leafy greens from the FFQ (Table 1). Final categories for fruit were citrus and non-citrus fruit, and final categories of vegetables were avocados, dark green, deep yellow, tomatoes, white potatoes, starchy and other vegetables. All quantities reported are expressed in servings/d, and all data are used, so that if a fruit or vegetable did not fall into one of the defined categories, it was included in 'Other'. In accordance with USDA convention, servings are defined as: half a cup of any cooked or raw vegetable or fruit or one cup of raw leafy greens. For a person on a 8368 kJ/d (2000 kcal/d) diet, the Recommended Dietary Guidelines for Americans 2015–2020 suggest 2.5 cups of vegetables and 2 cups of fruit/d, or 9–10 servings total<sup>(27)</sup>. Servings per day for each FV were generated by converting weekly or monthly consumption to daily, and multiplying this by servings as defined above. All variables are continuous measures of servings/d. Data were examined for outliers. Any values  $\pm$  two standard deviations were re-examined for plausibility and data collectors queried to confirm accuracy of all extreme values.

### Statistical analysis

Demographic characteristics of participants in the validation study and the FRESH Austin study were compared using  $\chi^2$  tests for sex, ethnicity, income and language at home (frequency and percentage reported) or *t*-test for age (mean and SD reported). In addition, scatterplots comparing categories of FV for the FFQ *v.* the 24hDR were generated for each of the nine food categories and examined for linearity (not provided). Crude mean and SD for categories of FV, separately and together, were computed. Shapiro–Wilks tests assessed normality of each food category variable. For non-normal variables, Spearman's  $\rho$  was reported, and Pearson's *r* was reported for normally distributed variables. Because the paired *t*-test is robust to non-normal distributions in larger ( $n > 30$ ) samples, this was used to assess differences between mean values for each food category, and *P*-values were reported<sup>(28)</sup>. Correlation estimates (Pearson's *r* or Spearman's  $\rho$ ) provide a measure of the relationship between the FFQ and the 24hDR. In order to adjust for random error from within-subjects variation across repeated 24hDR, correlations were deattenuated using the formula:

**Table 1** Validation study of food categories and their components from FRESH Austin FFQ and repeated 24hDR data

Food category	FFQ item(s)	24hDR item(s)*
Citrus	Citrus (oranges, grapefruit, etc.)	Citrus fruit (oranges, grapefruit, tangerines, lemons)
Non-citrus fruits	Apples, bananas, berries, grapes, melon other non-citrus	Fruit excluding citrus fruit
Avocados	Avocados	Avocados
Dark green	Lettuce or raw dark leafy greens, cooked dark leafy greens, broccoli or cauliflower	Spinach, collards, romaine, broccoli
Deep yellow	Carrots, sweet potatoes	Carrots, sweet potatoes, pumpkin, winter squash
Tomatoes	Tomatoes	Tomatoes
White potatoes	Potatoes (not sweet)	White potatoes, fried potatoes
Starchy	Maize, peas, jicama	Peas, maize, cassava, jicama
Other vegetables	Other	Other

24hDR, 24-h dietary recalls; NDSR, Nutrition Data System for Research.

\*Categories based on NDSR classification.

Deattenuated correlation = Crude correlation  $\sqrt{(1 + \lambda/n)}$

where  $\lambda$  is the ratio of within- to between-subject variance and  $n$  is the number of replicates of dietary data<sup>(29)</sup>. Within- and between-subject variances were obtained using one-way ANOVA of the 24hDR.

Correlations can be misleading if they are caused by a widespread sample (as when disagreement between methods is large but linear) and only provide an indication of the strength of the linear relationship between variables, rather than defining agreement. Therefore, we also present Bland–Altman plots, which plot the difference of paired variables  $v$ . their average, allowing estimates of fixed bias via mean difference. This bias is deemed significant based on its variance from zero. Limits of agreement (i.e., mean difference  $\pm 1.96$  SD of the difference) provide an estimate of variability of the agreement between the two methods and describe the range of values in which agreement between methods will fall for 95 % of the sample<sup>(30)</sup>. Bland–Altman plots were also used to explore proportional bias, which occurs when differences between methods vary across the sample.

Finally, borrowing from methods used in time series forecast analyses, mean absolute percentage errors (MAPE) provide a measure of percentage error between the paired FFQ and 24hDR observations, which can be indicative of prediction accuracy. These estimates use the 24hDR as the criterion or actual value and the FFQ as the forecast or predicted value, scaled to each category. Lower estimates suggest lower error, and better prediction, of the FFQ from the 24hDR. Significance was set at  $\alpha = 0.05$  for all tests, and all analyses were performed using STATA SE 14.2.

## Results

The participants in the validation study were similar to those in the FRESH Austin study with respect to age, ethnicity and

income, with no significant differences in these characteristics between study populations. Validation study participants were significantly more female and spoke different languages at home, especially a language other than English, Spanish, or English and Spanish equally (Table 2). Mode of data collection (in-person or via telephone) for the FFQ was not statistically different between FRESH Austin and validation study participants.

A comparison of crude estimates of FFQ and 24hDR servings/d indicates that ‘Non-Citrus Fruits’ and ‘Other Vegetables’ have the highest values across both instruments, reflecting the inclusion of a variety of FV in each category. The FFQ and 24hDR produced similar estimates of average total servings/d across FV (6.68 and 6.40 servings/d, respectively,) as well as for individual FV categories. Further comparison of crude estimates via the paired  $t$ -test reveals that there were no significant mean differences between total fruit, total vegetables and total FV. In all, no FV categories had significant mean differences (Table 3.)

Across categories of FV, the FRESH Austin FFQ provided moderately correlated outcomes compared with the repeated 24hDR, with deattenuated correlations above 0.30 for all food groups, except potatoes (Table 4). High correlations were observed for ‘Non-Citrus Fruits’ and ‘Other Vegetables’, which were both above 0.50. All correlations were significant, except for ‘White Potatoes’ and ‘Avocados’, and 55 % of food categories had correlations above 0.40. The highest correlations were found for total fruit (0.69), total vegetables (0.79) and total FV (0.73) (Table 4).

Because our interest is in a measure of error that does not penalise larger magnitude errors more than smaller magnitude errors, the MAPE was utilised to provide further insight. The MAPE relies on average values, meaning that large deviations are not as influential, allowing for frequently observed daily variances in consumption to be captured in the 24hDR and integrated more realistically in comparison to habitual dietary consumption patterns as recorded by the FFQ. MAPE values were small for all FV, suggesting the variance of the error estimates was also small. In addition, these data

**Table 2** Comparison of selected demographic characteristics of validation study and FRESH Austin participants

	Validation study, (n 54)	%	FRESH Austin, (n 400)	%	P
Sex					
Female	45	78.33	282	70.50	0.01
Age in years					
Mean	43.56		43.89		0.87
SD	1.89		0.68		
Ethnicity					
Hispanic	12	22.22	127	31.99	
Black	3	5.56	32	8.06	
White	29	53.70	203	51.13	
Other	10	18.52	35	8.82	0.09
Income					
Less than \$25 001	12	23.53	89	23.30	
\$25 001–\$45 000	21	41.18	112	29.32	
\$45 001–\$65 000	8	15.69	70	18.32	
> \$65 000	10	19.61	111	29.06	0.30
Language at home					
Only/mostly English	24	44.44	236	59.15	
Both English and Spanish	10	18.52	51	12.78	
Only/mostly Spanish	16	29.63	109	27.32	
Mostly other	4	7.41	3	0.75	<0.01
FFQ data collection mode					
In-person	48	88.89	311	77.75	
By telephone	6	11.11	89	22.25	0.06

FRESH Austin, Food Retail: Evaluating Strategies for a Healthy Austin.

**Table 3** Crude mean servings and sD for each food category (per day) by assessment method (FFQ and 24hDR), and paired *t*-test

Food category	FFQ (serv-ings/d)		24hDR (servings/d)		Paired <i>t</i> -test <i>P</i>
	Mean	SD	Mean	SD	
<b>Fruit</b>					
Citrus	0.68	0.12	0.30	0.08	0.17
Non-citrus fruits	2.06	0.20	2.29	0.21	0.32
AVG total fruit	2.74	0.21	2.59	0.17	0.20
<b>Vegetables</b>					
Avocados	0.28	0.05	0.13	0.04	0.35
Dark green	0.41	0.09	0.61	0.10	0.28
Deep yellow	0.44	0.07	0.37	0.05	0.08
Tomatoes	0.46	0.07	0.54	0.07	0.88
White potatoes	0.25	0.06	0.30	0.05	0.12
Starchy	0.22	0.05	0.15	0.03	0.11
Other vegetables	1.88	0.21	1.71	0.14	0.49
AVG total vegetables	3.94	0.13	3.81	0.10	0.26
AVG total FV	6.68	0.14	6.40	0.13	0.51

24hDR, 24-h dietary recalls; FV, fruit and vegetables.

indicate that the errors are small in both the negative and the positive direction, which is important to the assessment of agreement between the two methods, where either a positive or a negative difference would be of interest. The larger values generated for 'Avocados' and 'White Potatoes' suggest that larger errors in assessment may make it more difficult to capture significant changes for these categories, while the small MAPE values for 'Non-Citrus Fruits' and 'Other Vegetables' suggest these categories reflect more accurate measures of FV consumption.

Finally, Bland–Altman plots (Fig. 2) provide greater detail for assessing the degree to which the two methods

agree. Each plot shows the line of equality, or the line upon which all points would appear if the FFQ and the 24hDR produced the exact same measure<sup>(30,31)</sup>. Thus, these plots provide a visual method of assessing within-subject variance. For every food category, the plots show the grouping of data points for smaller values to be closer together, while outliers are generally found at larger values. Mean differences were less than 0.50 servings/d for all categories, including the 'Non-Citrus' and 'Other Vegetables' categories, which have larger crude values due to the inclusion of a greater variety of FV. Limits of agreement for each food category describe the range of agreement among the FFQ and 24hDR for 95 % of individuals assessed<sup>(32)</sup>.

## Discussion

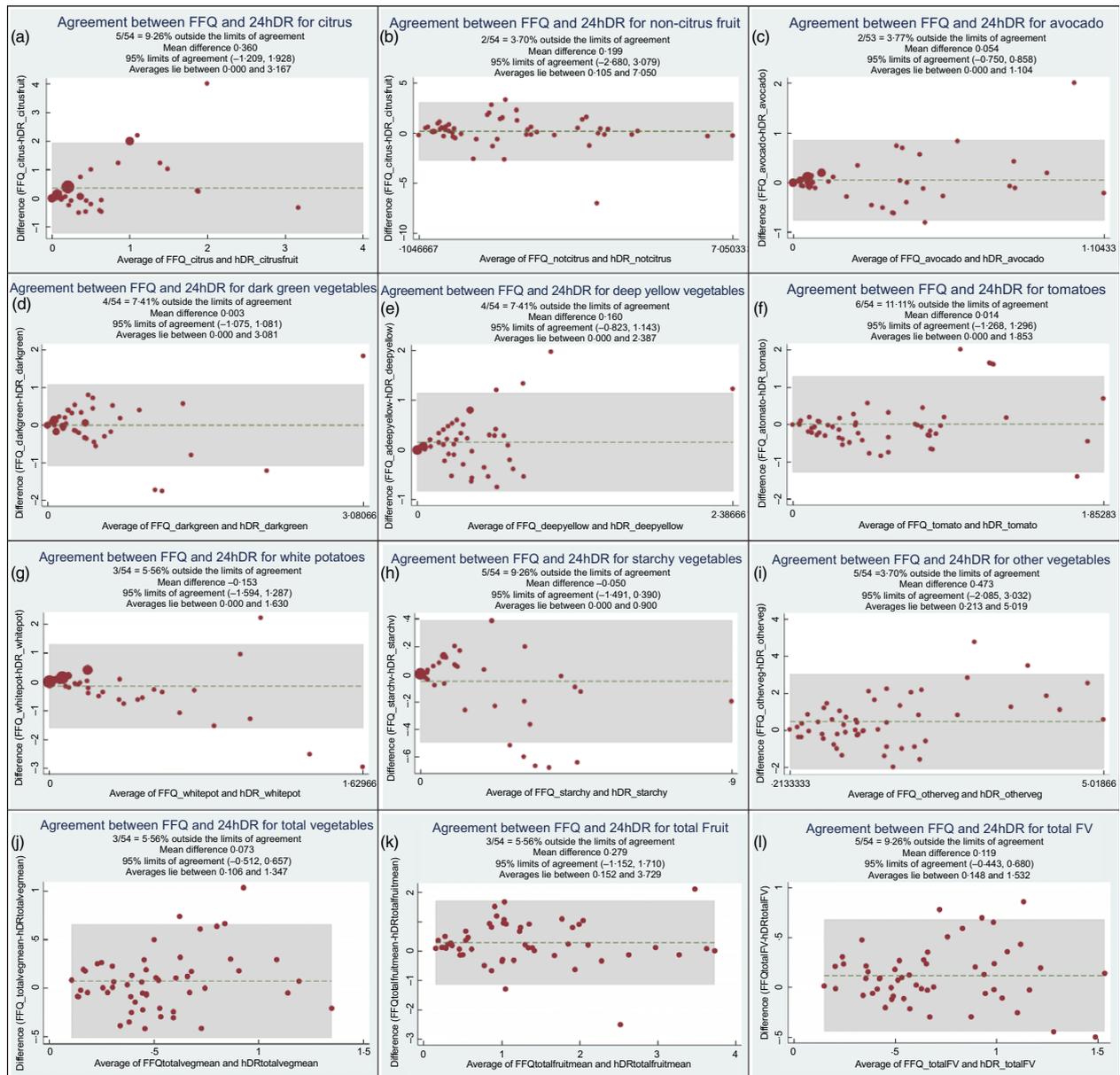
The central question of the current study is whether the FRESH Austin FFQ provides a valid measure of FV consumption, compared with the 24hDR<sup>(33)</sup>. This is motivated by the desire to use the least burdensome dietary assessment method that is sufficiently accurate to detect differences in FV consumption. The 24hDR is often used as the reference method compared with the FFQ for several reasons: it has less reliance on long-term memory than the FFQ (requiring recall of only the previous day), utilises a trained interviewer to enhance details and accuracy and elicits a detailed record of consumption, including methods of preparation and details of brands and sources<sup>(7,19,34)</sup>. However, the strengths of the 24hDR also make it burdensome, requiring repetition and trained personnel to achieve validity in line with established protocols. A consistent



Table 4 Spearman's  $\rho$  or Pearson's  $r$  mean absolute percentage error (MAPE), based on servings/d,  $n$  56

Food category	Spearman $\rho$ /Pearson's $r$	$P$	Deattenuated correlation	$P$	MAPE (%)	SD
Citrus	0.48	<0.01	0.51	0.01	3.0	0.06
Non-citrus Fruits	0.57	<0.001	0.60	0.004	0.9	0.10
Total fruit	0.64	<0.001	0.69	<0.001	2.2	0.05
Avocados	0.38	0.38	0.40	0.43	3.5	0.05
Dark green	0.33	0.02	0.34	0.02	0.7	0.03
Deep yellow	0.51	<0.01	0.56	0.02	1.1	0.06
Tomatoes	0.47*	<0.01	0.47	0.03	0.9	0.05
White potatoes	0.01	0.93	0.12	0.61	2.6	0.10
Starchy	0.30	0.03	0.32	0.02	1.5	0.03
Other vegetables	0.59*	<0.001	0.59	<0.001	0.4	0.10
Total vegetables	0.74	<0.001	0.79	<0.001	1.7	0.06
Total FV	0.69	0.02	0.73	0.03	0.7	0.06

\*Pearson's  $r$ , variable normally distributed after log-transformation.



FFQ, food frequency questionnaire; 24hDR, 24-hour dietary recall; FV, fruit and vegetable

Fig. 2 (colour online) Bland–Altman plots comparing the average of daily serving differences between FFQ and 24-h dietary recall (24hDR) estimates for all FV categories



agreement of the FRESH Austin FFQ and repeated 24hDR allows the use of the less burdensome option, in this case the FFQ, to be deemed reliable and effective at detecting important changes in consumption in the population of interest<sup>(35,36)</sup>.

In previous studies, the FFQ generally overestimates usual intake<sup>(6–8,19,35,37)</sup> and yields correlations between 0.40 and 0.70 across food groups and nutrients, compared with the 24hDR<sup>(38)</sup>. In our study, we found correlations between 0.30 for 'Starchy' vegetables and 0.79 for 'Total Vegetables', indicating that the FRESH Austin FFQ provides moderately valid measures for FV surveyed, except for 'Potatoes', which had a very low correlation of 0.12. We suspect this low correlation is due to NDSR software separating fried from other potatoes and combining fried starchy vegetables with fried potatoes, while the FFQ did not differentiate.

Our study found correlations in line with similar research, such as Hebden *et al.*'s<sup>(39)</sup> comparison of a tailored FFQ to repeated food records. In that study, fruit servings were correlated ( $r=0.58$ ) in line with our results ('Citrus'  $r=0.51$ , 'Non-Citrus'  $r=0.60$ ), as were vegetable servings ( $r=0.57$ ) in comparison to our 'Other Vegetables' ( $r=0.59$ )<sup>(39)</sup>. As found in other studies, the FFQ slightly overreported the consumption of FV compared with the criterion measure<sup>(5,6,9,39,40)</sup>. This may be attributed to social desirability bias, as participants may report habitual intake to resemble their own intended consumption or their perceptions of the interviewer's expectations, rather than actual intake. In contrast, the 24hDR recall may reduce this bias by asking more immediate questions of recent intake, providing less opportunity to edit consumption to align with intentions. Further, as noted by Boucher *et al.*, higher numbers of recall days are associated with greater correlations with FFQ values, suggesting that for some food categories, more than two or three recalls are required to account for daily variance in consumption<sup>(37)</sup>. While the brevity of the FRESH Austin FFQ reduced survey burden, it also eliminated the ability to calculate total energy, since the entire diet was not evaluated. This limitation would be important for any investigations into associations with disease, since consumption cannot be scaled by total energy. However, the truncated FFQ may be appropriate for studies intended to capture changes in FV consumption, rather than deattenuated values<sup>(41,42)</sup>.

Agreement between the FFQ and 24hDR was reported via the paired *t*-test, which found that none of the mean differences in any FV category were significantly different from zero (i.e., all comparisons were nonsignificant). In addition, the small error estimates obtained via the MAPE suggest the two methods substantially agree.

The pattern observed via the Bland–Altman plots, showing closer agreement at smaller quantities and greater discrepancies at higher quantities, argues for the careful examination and possible exclusion of outliers in pre/post-assessments. Similar results were reported in other

studies<sup>(43–45)</sup> where greater variation was found at higher levels of intake. These values may be 'true', in the sense that large quantities of a specific food were eaten, but 'untrue' as an indicator of habitual consumption. Because measures of agreement in the Bland–Altman plots are distributed both above and below the line of equality, no systematic bias is detected. However, the limits of agreement were wide (Fig. 1, panels (a)–(1)), which is similar to research reported in two studies by Conway *et al.*<sup>(16,17)</sup>, Bautista *et al.*<sup>(46)</sup> and a comparable validation study among Lebanese children<sup>(43)</sup>, suggesting that the FFQ may lead to important under- or over-estimation of actual intake. This feature, as well as the truncated food list of the FFQ (only FV, rather than the whole diet), limits the use of this instrument in exploring associations with chronic disease, where precise and calibrated nutrient estimates are critical. However, because mean estimates of FV consumption exhibit no indication of systematic bias, this FFQ would be appropriate for valid comparisons of cohort designs<sup>(46)</sup>. This FFQ would be useful to provide estimates of changes in usual consumption over time, where the outcome of interest is not a measure of true intake, but rather an assessment of changes in consumption. The lower survey burden and moderate precision of the FFQ may be well-suited to community-level interventions that are intended to change consumption patterns.

Limitations of the current study include a lack of energy adjustment from total kcal, which was not possible because the entire diet was not assessed via the FFQ. This limits the ability to make conclusions regarding associations with disease, since the results cannot be scaled by total consumption. In addition, the FFQ food list and NDSR software were not aligned for 'Potatoes', with NDSR separating fried from other potatoes, whereas the FFQ did not. We suspect this is the cause of the low correlation in this category and may not indicate a true problem with the FFQ. More worrisome, however, is the low correlation for avocados, which can be an important part of the diet for Hispanic populations. Future research may need to explore better ways of capturing this category of consumption more accurately, perhaps including descriptions of how avocados may be consumed, such as in dips or sauces. Our study protocol may have introduced bias by setting up appointments for subsequent data collections, increasing the opportunity for social desirability bias to affect dietary behaviours in advance of our interviews. The current study's strengths include using an FFQ food list that was carefully adapted to the population of interest, fully bilingual materials and research staff, and multiple methods of analyses (*t*-tests, correlation, MAPE and Bland–Altman graphs) to explore the validity of the FFQ.

## Conclusion

Every FV group assessed by the FRESH Austin FFQ showed acceptable levels of association between FFQ and 24hDR, with the exception of potatoes and avocados, suggesting



that this tailored FFQ is able to capture usual consumption with sufficient accuracy to enable valid assessment of changes in FV intake. The FFQ minimises respondent burden, which is an especially important condition for retention in cohort studies and helps ensure sufficient sample sizes and power to detect changes in outcomes. In addition, the FRESH Austin FFQ's focus on whole foods is aligned with evaluation of community-based interventions, such as Austin's FFL programme, aimed at improving access in high-need communities<sup>(47–49)</sup>. As in other community-level programming, the outcomes of interest are changes in patterns of consumption, specifically increases in FV intake. This FFQ aligns evaluation with implementation, providing a measure of change that is important for programme evaluation, as well as for assessment of an important determinant of desired health outcomes.

### Acknowledgements

**Acknowledgements:** The authors would like to thank the Southeast Health & Wellness Center, the St. John's Community Center and the Northwest YMCA in Austin, TX for their participation in this research. **Financial support:** Funding was provided by the Foundation for Food and Agriculture Research (06-67320005-57015-0013764-15). This work was undertaken as part of the first author's PhD dissertation at the University of Texas, School of Public Health. **Conflict of interest:** There are no conflicts of interest. **Authorship:** C.E.S.J. designed the study; supervised data collection, cleaning and analysis; and drafted the manuscript. J.W. assisted with creating study instruments and protocols. J.W. and M.D. assisted with data collection, data entry and quality control. D.M.H. assisted with study design and editing. B.C. advised on statistical analysis and results. N.R. assisted with study design, coding and analysis. A.E.V. secured funding, supervised all aspects of the research and co-wrote the final manuscript. **Ethics of human subject participation:** The current study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving research study participants were approved by the Internal Review Board of the University of Texas. Written informed consent was obtained from all subjects/patients.

### References

- Cummins S, Petticrew M, Higgins C *et al.* (2005) Large scale food retailing as an intervention for diet and health: quasi-experimental evaluation of a natural experiment. *J Epidemiol Community Health* **59**, 1035–1040.
- McCormack LA, Laska MN, Larson NI *et al.* (2010) Review of the nutritional implications of farmers' markets and community gardens: a call for evaluation and research efforts. *J Am Diet Assoc* **110**, 399–408.
- Sharpe P, Wilcox S, Bell B *et al.* (2019) Evaluation of a food hub initiative's effect on food shoppers' perceptions, shopping behavior, diet, and weight in a community of low income and low access to healthy food (OR16–06–19). *Curr Dev Nutr* **3**, nzz051.
- Kim DJ & Holowaty EJ (2003) Brief, validated survey instruments for the measurement of fruit and vegetable intakes in adults: a review. *Prev Med* **36**, 440–447.
- Rodrigo CP, Aranceta J, Salvador G *et al.* (2015) Food frequency questionnaires. *Nutr Hosp* **31**, 49–56.
- Willett W & Lenart E (2013) Reproducibility and validity of food-frequency questionnaires. *Nutr Epidemiol* **3**.
- Subar AF, Thompson FE, Kipnis V *et al.* (2001) Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study. *Am J Epidemiol* **154**, 1089–1099.
- Cade J, Thompson R, Burley V *et al.* (2002) Development, validation and utilisation of food-frequency questionnaires—a review. *Public Health Nutr* **5**, 567–587.
- Jackson MD, Walker SP, Younger NM *et al.* (2011) Use of a food frequency questionnaire to assess diets of Jamaican adults: validation and correlation with biomarkers. *Nutr J* **10**, 28.
- Fraser GE, Jaceldo-Siegl K, Henning SM *et al.* (2016) Biomarkers of dietary intake are correlated with corresponding measures from repeated dietary recalls and food-frequency questionnaires in the Adventist Health Study-2. *J Nutr* **146**, 586–594.
- Boeing H, Bohlscheid-Thomas S, Voss S *et al.* (1997) The relative validity of vitamin intakes derived from a food frequency questionnaire compared to 24-h recalls and biological measurements: results from the EPIC pilot study in Germany. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* **26**, S82.
- Michael & Susan Dell Center for Healthy Living (2021) Food Retail: evaluating Strategies for a Healthy Austin (FRESH Austin). <https://sph.uth.edu/research/centers/dell/project.htm?project=9707359b-4d0d-426e-b915-760809f89b4a> (accessed January 2021).
- Thompson FE, Dixit-Joshi S, Potischman N *et al.* (2015) Comparison of interviewer-administered and automated self-administered 24-h dietary recalls in 3 diverse integrated health systems. *Am J Epidemiol* **181**, 970–978.
- Hughes AR, Summer SS, Ollberding NJ *et al.* (2017) Comparison of an interviewer-administered with an automated self-administered 24 h (ASA24) dietary recall in adolescents. *Public Health Nutr* **20**, 3060–3067.
- Harnack LJ & Pereira MA (2017) Strategies for improving the validity of the 24-h dietary recall and food record methods. In *Advances in the Assessment of Dietary Intake*, pp. 67–84 [DA Schoeller & MS Westerp-Plantenga, editors]. Boca Raton, FL: CRC Press.
- Conway JM, Ingwersen LA, Vinyard BT *et al.* (2003) Effectiveness of the US Department of Agriculture 5-step multiple-pass method in assessing food intake in obese and nonobese women. *Am J Clin Nutr* **77**, 1171–1178.
- Conway JM, Ingwersen LA & Moshfegh AJ (2004) Accuracy of dietary recall using the USDA five-step multiple-pass method in men: an observational validation study. *J Am Diet Assoc* **104**, 595–603.
- Rocheleau Z (2018) Visual Hand Guide to Portion Sizes. [https://medium.com/@Zach\\_Rocheleau/visual-hand-guide-to-portion-sizes-c499142d2734](https://medium.com/@Zach_Rocheleau/visual-hand-guide-to-portion-sizes-c499142d2734) (accessed April 2021).
- Block G, Thompson F, Hartman A *et al.* (1992) Comparison of two dietary questionnaires validated against multiple dietary records collected during a 1-year period. *J Am Diet Assoc* **92**, 686–693.
- Lutz LJ, Nakayama AT, Karl JP *et al.* (2019) Serum and erythrocyte biomarkers of nutrient status correlate with short-term A-Carotene,  $\beta$ -Carotene, Folate, and vegetable intakes



- estimated by food frequency questionnaire in military recruits. *J Am Coll Nutr* **38**, 171–178.
21. Tanner KJ & Watowicz RP (2018) Development of a tool to measure the number of foods and beverages consumed by children using National Health and Nutrition Examination Survey (NHANES) FFQ data. *Public Health Nutr* **21**, 1486–1494.
  22. Sanjeevi N, Freeland-Graves J & George GC (2017) Relative validity and reliability of a 1-week, semiquantitative food frequency questionnaire for women participating in the supplemental nutrition assistance program. *J Acad Nutr Diet* **117**, 1972–1982.
  23. Austintexas.gov (2021) Fresh for Less. <https://www.austintexas.gov/department/fresh-less> (accessed April 2021).
  24. Burkholder-Cooley NM, Rajaram SS, Haddad EH *et al.* (2017) Validating polyphenol intake estimates from a food-frequency questionnaire by using repeated 24-h dietary recalls and a unique method-of-triads approach with 2 biomarkers. *Am J Clin Nutr* **105**, 685–694.
  25. Nutrition Coordinating Center (1992) *Nutrientdatasystem, Release 2.4*. Minneapolis, MN: University of Minnesota.
  26. Ahuja J, Montville JB, Omolewa-Tomobi G *et al.* (2012) *USDA Food and Nutrient Database for Dietary Studies, 5.0—Documentation and User Guide*. Beltsville, MD, USA: US Department of Agriculture, Agricultural Research Service, Food Surveys Research Group.
  27. Millen BE, Abrams S, Adams-Campbell L *et al.* (2016) The 2015 dietary guidelines advisory committee scientific report: development and major conclusions. *Adv Nutr* **7**, 438–444.
  28. Rasch D & Guiard V (2004) The robustness of parametric statistical methods. *Psychol Sci* **46**, 175–208.
  29. Willett W (1998) Correction for the effects of measurement error. *Nutr Epidemiol*, 302–320.
  30. Bland JM & Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* **8**, 135–160.
  31. Bland JM & Altman D (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **327**, 307–310.
  32. Bunce C (2009) Correlation, agreement, and Bland–Altman analysis: statistical analysis of method comparison studies. *Am J Ophthalmol* **148**, 4–6.
  33. Bingham SA, Gill C, Welch A *et al.* (1994) Comparison of dietary assessment methods in nutritional epidemiology: weighed records *v.* 24 h recalls, food-frequency questionnaires and estimated-diet records. *Br J Nutr* **72**, 619–643.
  34. Shim J-S, Oh K & Kim HC (2014) Dietary assessment methods in epidemiologic studies. *Epidemiol Health* **36**, e2014009.
  35. Shatenstein B & Payette H (2015) Evaluation of the relative validity of the short diet questionnaire for assessing usual consumption frequencies of selected nutrients and foods. *Nutrition* **7**, 6362–6374.
  36. Thompson FE & Subar AF (2017) Dietary assessment methodology. In *Nutrition in the Prevention and Treatment of Disease*, 4th ed., pp. 5–48 [AM Coulston, CJ Boushey, MG Ferruzzi *et al.*, editors]. Cambridge, MA: Academic Press.
  37. Boucher B, Cotterchio M, Kreiger N *et al.* (2006) Validity and reliability of the Block98 food-frequency questionnaire in a sample of Canadian women. *Public Health Nutr* **9**, 84–93.
  38. Willett WC (1994) Future directions in the development of food-frequency questionnaires. *Am J Clin Nutr* **59**, 171S–174S.
  39. Hebden L, Kostan E, O’Leary F *et al.* (2013) Validity and reproducibility of a food frequency questionnaire as a measure of recent dietary intake in young adults. *PLoS One* **8**, e75156.
  40. Dehghan M, Martinez S, Zhang X *et al.* (2013) Relative validity of an FFQ to estimate daily food and nutrient intakes for Chilean adults. *Public Health Nutr* **16**, 1782–1788.
  41. Do M, Kattelman K, Boeckner L *et al.* (2008) Low-income young adults report increased variety in fruit and vegetable intake after a stage-tailored intervention. *Nutr Res* **28**, 517–522.
  42. Pearson N, Atkin AJ, Biddle SJ *et al.* (2010) A family-based intervention to increase fruit and vegetable consumption in adolescents: a pilot study. *Public Health Nutr* **13**, 876–885.
  43. Moghames P, Hammami N, Hwalla N *et al.* (2015) Validity and reliability of a food frequency questionnaire to estimate dietary intake among Lebanese children. *Nutr J* **15**, 4.
  44. Streppel MT, de Vries JH, Meijboom S *et al.* (2013) Relative validity of the food frequency questionnaire used to assess dietary intake in the Leiden Longevity Study. *Nutr J* **12**, 75.
  45. Siebelink E, Geelen A & de Vries JH (2011) Self-reported energy intake by FFQ compared with actual energy intake to maintain body weight in 516 adults. *Br J Nutr* **106**, 274–281.
  46. Bautista LE, Herran OF & Pryer JA (2005) Development and simulated validation of a food-frequency questionnaire for the Colombian population. *Public Health Nutr* **8**, 181–188.
  47. Resnicow K, Jackson A, Wang T *et al.* (2001) A motivational interviewing intervention to increase fruit and vegetable intake through Black churches: results of the Eat for Life trial. *Am J Public Health* **91**, 1686–1693.
  48. Haire-Joshu D, Elliott MB, Caito NM *et al.* (2008) High 5 for Kids: the impact of a home visiting program on fruit and vegetable intake of parents and their preschool children. *Prev Med* **47**, 77–82.
  49. Yoder ABB, Liebhart JL, McCarty DJ *et al.* (2014) Farm to elementary school programming increases access to fruits and vegetables and increases their consumption among those with low intake. *J Nutr Educ Behav* **46**, 341–349.