

Article

Minimum Dietary Diversity for Women of Reproductive Age (MDD-W) Data Collection: Validity of the List-Based and Open Recall Methods as Compared to Weighed Food Record

Giles T Hanley-Cook ¹, Ji Yen A Tung ², Isabela F Sattamini ², Pamela A Marinda ³, Kong T Thong ⁴, Dilnesaw Zerfu ⁵, Patrick Kolsteren ¹, Maria Antonia G Tuazon ² and Carl K Lachat ^{1,*}

¹ Department of Food Technology, Safety and Health, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium; giles.hanleycook@ugent.be (G.H-C.); patrick.kolsteren@ugent.be (P.K.); carl.lachat@ugent.be (C.L.)

² Nutrition and Food Systems Division (ESN), Food and Agriculture Organization of the United Nations (FAO), Rome, Italy; jiyenalexandra.tung@fao.org (JY.T.); isabela.sattamini@fao.org (I.S.); mariaantonia.tuazon@fao.org (MA.T.)

³ Department of Food Science and Nutrition, School of Agricultural Sciences, University of Zambia, Lusaka, Zambia; pamela.marinda@unza.zm (P.M.)

⁴ Department of Food Science, Faculty of Agro-Industry, Royal University of Agriculture, Phnom Penh, Cambodia; kthong@rua.edu.kh (K.T.)

⁵ Food Science and Nutrition Research Directorate, Ethiopian Public Health Institute, Addis Ababa, Ethiopia; dilnesaw2012@gmail.com (D.Z.)

* Correspondence: carl.lachat@ugent.be; Tel.: +32-9-264-9377

Received: 3 June 2020; Accepted: date; Published: date

Abstract: Minimum Dietary Diversity for Women of Reproductive Age (MDD-W) was validated as a population-level proxy of micronutrient adequacy, with indicator data collection proposed as either list-based or open recall. No study has assessed the validity of these two non-quantitative proxy methods against weighed food records (WFR). We assessed the measurement agreement of list-based and open recall methods as compared to WFR (i.e. reference method of individual quantitative dietary assessment) for achieving MDD-W and ordinal food group diversity score. Applying a non-inferiority design, data were collected from non-pregnant women of reproductive age in Cambodia (n=430), Ethiopia (n=431), and Zambia (n=476). For the pooled sample (n=1,337), proportions achieving MDD-W from both proxy methods were compared to WFR proportion by McNemar's chi-square tests, Cohen's kappa, and receiver operating characteristic (ROC) analysis. Ordinal food group diversity (0-10) was compared by Wilcoxon matched-pairs signed-rank tests, intraclass correlation coefficients (ICC), and weighted kappa. MDD-W food groups that were most frequently misreported (i.e. type I and II errors) by the proxy methods were determined. Our findings indicate statistically significant differences in proportions achieving MDD-W, ordinal food group diversity scores, and ROC curves between both proxy methods and WFR ($P < 0.001$). List-based and open recall methods overreported women achieving MDD-W by 16 and 10 percentage points, respectively as compared to WFR (proportion achieving MDD-W: 30%). ICC values between list-based or open recall and WFR were 0.50 and 0.54, respectively. Simple and weighted kappa values both indicated moderate agreement between list-based or open recall against WFR. Food groups most likely to be misreported using proxy methods were beans and peas, dark green leafy vegetables, vitamin A-rich fruit and vegetables, and other fruits. Our study provides statistical evidence for overreporting of both list-based and open recall methods for assessing prevalence of MDD-W or ordinal food group diversity score in women of reproductive age in low- and middle-income countries. Operationalizing MDD-W through qualitative recall methods should consider potential trade-offs between accuracy and simplicity.

Keywords: Cambodia; Ethiopia; List-based recall; Open recall; Minimum Dietary Diversity for Women; Weighed Food Record; Zambia

1. Introduction

Low quality, non-diverse diets are responsible for the greatest burden of morbidity and mortality worldwide, in particular for nutritionally vulnerable women and children in low- and middle-income countries (LMICs) (1). Monotonous diets, often dominated by starchy staples and a lack of fruits, vegetables, and animal-source foods, are the norm in resource-poor settings (2). Nevertheless, comprehensive data on dietary patterns, diet quality, and subsequent micronutrient adequacy from nationally representative studies are scarce.

Even though numerous methods are available for assessing individual dietary intake (3), most necessitate highly proficient enumerators and exceptionally resource-intensive data collection, processing, and analysis. Many dietary assessment methodologies also require the availability of complete food composition tables, the development of which is also resource-intensive. Therefore, there is a strong and rising demand for simple and feasible, yet accurate and precise, proxy indicators to reflect micronutrient adequacy and overall diet quality (4). The lack of indicators to allow for assessment, advocacy, and accountability has been acknowledged as a key constraint to programmatic and policy action to improve women's diet quality (5).

In response to this demand, a dichotomous (≥ 5 of 10 defined food groups) Minimum Dietary Diversity for Women of Reproductive Age (MDD-W) indicator was developed and validated as a proxy measure for assessing micronutrient adequacy from the diet in non-pregnant and non-lactating women of reproductive age (WRA) at the population-level (6). The MDD-W indicator was designed for settings where a weighed food record (WFR) or other quantitative dietary assessment methods are unfeasible. Current MDD-W guidance (7), identified two data collection approaches, namely list-based or open recall. Neither proxy method requires estimation of portion sizes (the quantity of food or drink consumed), although careful considerations are made with regard to the $\geq 15g$ threshold (7).

There are, nonetheless, remaining questions with regard to the most accurate, precise, and simplest approach for collecting MDD-W data. Both the list-based and open recall methods (7) have certain operational advantages and disadvantages for measuring dietary diversity. The list-based method requires fewer enumerator capacity requirements and shorter training time; however it might be more likely to result in misclassification of foods and beverages into food groups or misreporting of some food items, particularly those consumed in trivial quantities (8,9). The non-quantitative open recall method has previously been proposed, as it might produce a more accurate and complete recall of all foods and drinks consumed; however, it requires longer training time and more proficient enumerators who have a reasonable knowledge of local foods, beverages, and recipes (3,7).

However, comparability of the list-based and open recall methods against WFRs (i.e. reference method of individual quantitative dietary assessment) has not been investigated in the context of evaluating the performance of MDD-W data collection. To our knowledge, one study has examined the relative validity of the list-based and open recall methods against each other, but not compared to WFR, to assess MDD-W in pregnant women in Bangladesh and India (8). Furthermore, Martin-Prével et al. (2010) compared simple food group diversity indicators derived from list-based recall to the same indicators derived from quantitative 24-hour recalls (9).

In this study, we address the research gap by using comparable dietary intake data from non-pregnant WRA in Cambodia, Ethiopia, and Zambia to examine the measurement agreement of list-based and open recall as compared to WFR for predicting MDD-W, ordinal food group diversity score, and individual food group consumption.

2. Materials and Methods

Our research was reported using the STROBE-nut checklist (10).

2.1. Sample size

The key aim of this multi-country study was to determine whether proxy methods are comparable to WFR at capturing the proportion of WRA achieving MDD-W. Non-inferiority tests of correlated proportions were designed for this purpose. We used PASS 16 Power Analysis and Sample Size Software (2018) to determine the sample size, using the following assumptions (11): 90% power ($1-\beta$), 5% significance level (α), 5% maximum allowable difference between proxy methods and WFR, 0% actual difference, 56% (Cambodia); 13% (Ethiopia); 57% (Zambia) standard MDD-W proportion, and 10% nuisance parameter. Our initial calculations resulted in a sample size of 374 WRA. However, accounting for a 15% dropout rate, the final sample size was 430 WRA per dietary assessment method in each country. This study pooled data from Cambodia, Ethiopia, and Zambia ($n=1,337$) as <15% of the associations between MDD-W recalled from proxy methods and WFR were explained by country-level clustering when fitting mixed effects logistic regression models (random intercept: country; supplementary Table S1).

2.2. Data sources and study population

The study countries were selected based on their diverse geographic locations, variation in dietary patterns, and the consideration that they are part of the Deutsche Gesellschaft für Internationale Zusammenarbeit's (GIZ) Special Initiative ONE World - No Hunger programme, which collected MDD-W survey data among WRA in 12 countries (12). Thus, our multi-country study was able to reference GIZ's country-specific experience and survey tools during the preparatory stage.

In Cambodia, data were collected in the rainy season between August and September 2019 in Kampong Thom Province in 34 villages (*phum*) in 5 rural communes (*khum*) from 2 districts (*srok*). In Ethiopia, data were collected in the rainy season between June and August 2019 in Amhara Region in 24 villages (*gotes*) in 13 sub-districts (*kebeles*) from 3 districts (*woredas*), characterised by rural, peri-urban, and urban residencies. In Zambia, data were collected in the hot and dry/wet seasons between October and December 2019 in Chongwe District in Lusaka Province in 6 villages from 3 peri-urban sub-districts (*wards*). Convenience samples of non-pregnant WRA (15-49 years) were selected from each study site in Cambodia ($n=430$) and Zambia ($n=476$), whereas representative samples (i.e. probability proportional to size) were selected in Ethiopia ($n=431$).

2.3. Ethical standards disclosure

The studies received ethical approvals from the National Ethics Committee for Health Research (RUA/NECHR/053/2019) in Cambodia, Ethiopian Public Health Institute Scientific and Ethical Review office (EPHI/IRB/156/2018) in Ethiopia, and the University of Zambia Biomedical Research Ethics Committee (UNZA/IRB/304/2019) in Zambia. Written informed consent were obtained from all study participants.

2.4. Preparatory phase

Prior to data collection in Cambodia, Ethiopia, and Zambia, surveys and, in particular, food lists were adapted to the local context (i.e. focus group discussions with nutritionists and health workers), based on existing country-specific MDD-W data collection tools (12). The adapted surveys were programmed into tablets, using the Open Data Kit (ODK) platform (<https://opendatakit.org>), in both English and the local language.

One-week capacity development workshops were held for MDD-W data collection teams, including classroom training and a field practice (i.e. pretest) of all three dietary assessment methods, using the tablet-based surveys. Furthermore, a pilot study was conducted on at least 50 WRA in each study country, to further improve the data collection tools, data quality checks, data management, statistical analysis procedures, and enumerator confidence and capacity (between May and August 2019).

2.5. Data collection

We used quantitative WFR and qualitative list-based and open recall, as described in FAO & FHI 360 (2016), to collect individual data on the intake of distinct food groups over a 24-hour period (7). The proxy methods were both enumerated on the day after the WFR, but by different interviewers and at different times (both at random, using MS Excel RAND function in Ethiopia and Zambia and the Random Generator application in Cambodia). Across all three study countries, no attrition was observed and no data were excluded (Figure 1).

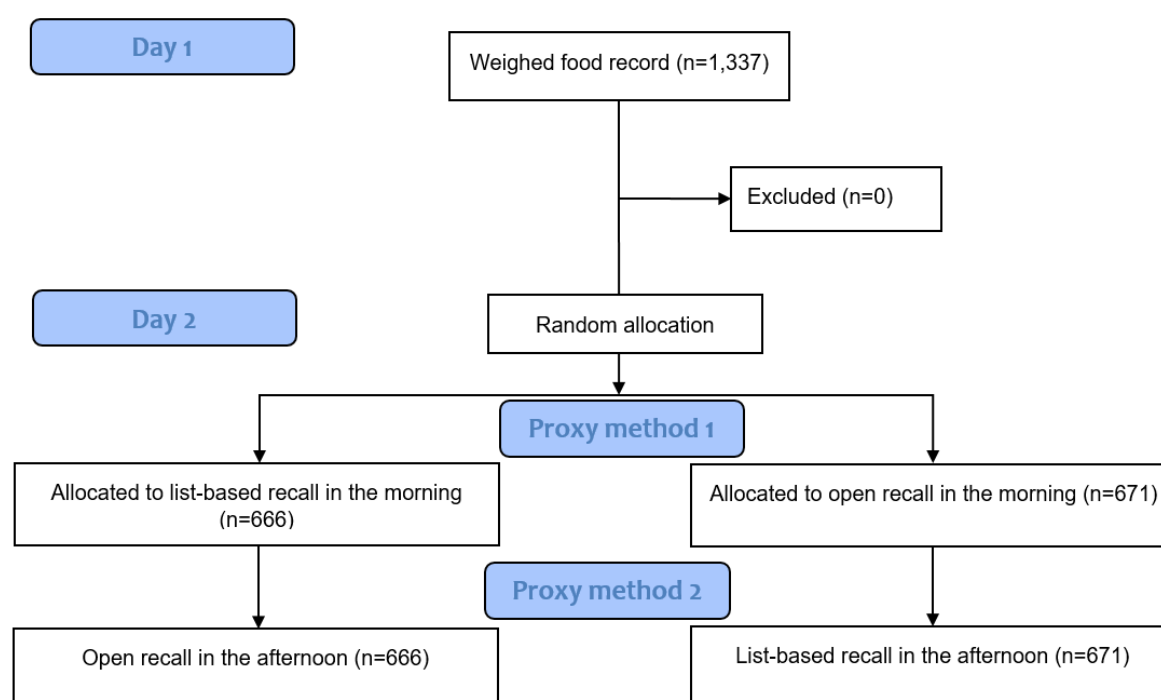


Figure 1. Study design diagram.

To conduct the WFR, each interviewer trained in the WFR method spent the entire day in a chosen household weighing and recording all foods and beverages consumed by the participant using digital scales accurate to ± 1 g and calibrated daily. Participants were instructed not to change their normal dietary pattern and to serve their own portions of food on separate utensils from other

household members on the WFR day. On the occasions when subjects attended events outside the home, the interviewer accompanied them to weigh and record any food or beverages consumed (13,14).

Detailed weighed recipe data were also collected for all the composite dishes consumed during the WFR. These data were then used to calculate the weight (g) of each ingredient, after adjustments, where necessary, for yield using US Department of Agriculture (USDA) factors (15), because there are no Cambodian, Ethiopian, or Zambian data.

During the list-based recall, the trained enumerator read a list of foods and drinks from each food group, displayed on a tablet with the ODK platform, to the respondent. A priori, local nutritionists reached a consensus on foods that would be consumed in daily quantities $\geq 15\text{g}$ threshold. The non-quantitative questionnaire was based on a list of 18 food groups in Ethiopia, and 19 food groups in Cambodia and Zambia (supplementary Table S2) that reflected the unique characteristics of food consumption in each country (e.g. including the optional insects and other small protein foods category for the latter). Various examples of local foods or composite dishes made from these local foods were provided for each group.

The open recall method evaluated food and drink consumption using an ODK tablet-based multiple pass qualitative 24-hour recall. WRAs were requested to describe all the foods and drinks they consumed during the preceding day and night, as well as the meal time. Recipes of mixed dishes were enumerated by asking the women who had prepared and consumed them to recall all the ingredients they added. Foods and ingredients usually consumed in less than a tablespoon quantity (i.e. $< 15\text{g}$) were defined during country-specific survey adaption and enumerators were trained on the principle of not counting them towards the MDD-W indicator before data collection, as described in the MDD-W measurement guide (7).

2.6. Constructing MDD-W

For the WFR, the food and drink items were re-categorised into the 10 food groups used in the MDD-W measurement guide (7): 1) starchy staple foods; 2) beans and peas; 3) nuts and seeds; 4) dairy products (milk, yoghurt, and cheese); 5) flesh foods (meat, fish, poultry, and liver/organ meats); 6) eggs; 7) dark green leafy vegetables; 8) vitamin A-rich fruits and vegetables; 9) other vegetables; and 10) other fruits. A WFR food group diversity indicator (0-10) was obtained by summing the number of food groups consumed in a quantity of at least 15g by each WRA.

For the list-based recall, the foods groups were aggregated into the 10 defined MDD-W food groups. In the qualitative open recall, all food items were assembled into the identical 10 food groups. These food group variables were used for descriptive analyses of misreporting in recalled consumption of the food group as compared to WFR. For the proxy methods, food group diversity scores (0-10) were constructed by summing the number of food groups consumed by each WRA.

2.7. Statistical analysis

First, we tested the effect of allocation of proxy methods (day 2) on MDD-W (i.e. potential “carry over” effect (16,17)), using a two-sample test of proportions for the dichotomous variable and Wilcoxon rank-sum test of the difference in food group diversity scores. We used descriptive analyses to report the proportion of WRA reaching MDD-W, median food group diversity score, and individual food group consumption.

For our main outcome of interest, we performed McNemar's chi-square tests for paired proportions to assess how well both proxy methods estimated MDD-W and individual food group consumption compared to WFR. To measure agreement amongst WRA achieving MDD-W, we used simple Cohen's kappa statistics. Kappa scores of 0.21-0.40 indicate fair agreement; 0.41-0.60, moderate agreement; 0.61-0.80, substantial agreement; 0.81-1.00, almost perfect agreement (18). We also performed a "gold standard" receiver operator characteristic (ROC) analysis to compare how well the list-based and open recall methods predicted MDD-W as compared to WFR. The area under the curve (AUC) summarises the predictive power of MDD-W for the proxy methods. An AUC significantly different from 0.5 and ≥ 0.70 was deemed satisfactory to indicate accuracy (8).

Furthermore, as the 10-point ordinal score might be preferred in research and certain programmatic contexts (6), we also performed Wilcoxon matched-pairs signed-rank test, intraclass correlation coefficients (ICC), and weighted Cohen's kappa to assess how well both proxy methods estimated food group diversity score compared to WFR. The ICC was used to estimate the consistency between dietary assessment methods with a higher ICC representing a higher degree of consistency (19,20). We calculated the correlation of proxy method against WFR using Spearman's rank correlation coefficient, which is fitting, since we considered our food group diversity variables to be ordinal.

We quantified the frequency of misreporting (i.e. type I and II errors) for MDD-W and each individual food group using confusion matrices and identified the food groups that were most often misreported by the proxy methods.

Data management and statistical analysis were conducted in Stata version 15.1 (21). A two-sided significance level of $P < 0.05$ was applied for all analyses.

3. Results

The allocation sequence of proxy methods was non-significant for the dichotomous MDD-W indicator ($P = 0.64$) and ordinal food group diversity score ($P = 0.49$). Therefore, our subsequent analyses included all study data from list-based and open recall methods (both $n = 1337$).

3.1. Sample characteristics

Most non-pregnant WRA (>60%) were married with a mean $\pm$ SD age ranging between 36 ± 8 years in Cambodia ($n = 430$) and 29 ± 9 years in Zambia ($n = 476$). In Cambodia almost all WRA were Buddhist (>90%), whereas in Ethiopia and Zambia large proportions of WRA were Christians (>60%). In all three study counties >50% of WRA were the **wife of a male household head** and >30% were homemakers by occupation. Our WFR findings showed that, in Cambodia, 44% of WRA achieved MDD-W, 8% in Ethiopia ($n = 431$), and 38% in Zambia. The median [interquartile range (IQR)] food group diversity score in Ethiopia (3 [3, 3]) was 1 food group lower than in Cambodia (4 [4, 5]) and Zambia (4 [4, 5]), based on WFR.

3.2. Comparison of the performance of proxy methods and WFR to predict dichotomous MDD-W

The proportion [95% confidence interval (CI)] of WRA achieving MDD-W (consumed ≥ 5 food groups; $n = 1,337$) was 30% (95% CI: 28-33) as stated by the WFR, whereas according to the list-based recall it was 46% (95% CI: 43-48) and 40% (95% CI: 37-42) using the open recall method (Table 1; both $P < 0.001$). The list-based and open recall methods correctly classified (i.e. true positive and negative values) 77% and 80% of WRA as (not) achieving MDD-W, respectively (Table 2). Simple kappa values for the dichotomous indicator indicated moderate agreement for both proxy methods against WFR

($k=0.51-0.57$). Furthermore, list-based recall had a sensitivity (i.e. truly achieving MDD-W) of 87% and a specificity (i.e. truly not achieving MDD-W) of 72%, whereas open recall had a sensitivity and specificity of 83% and 79%, respectively (Figure 2). The pooled performance of proxy methods to predict dichotomous MDD-W, compared to WFR (AUC: 1), was comparable for the list-based and open recall methods (AUCs ranged from 0.68 to 0.70 in Cambodia, 0.89 to 0.91 in Ethiopia, and 0.76 to 0.80 in Zambia). Nevertheless, the pooled AUCs for the list-based and open recall methods were significantly different from WFR (both $P<0.001$).

Table 1. Pooled proportions of non-pregnant women (15-49 years) having consumed food groups and achieved Minimum Dietary Diversity for Women of Reproductive Age, based on weighed food record, list-based, and open recall methods.

	Weighed food record (n=1,337)	List-based recall (n=1,337)	Open recall (n=1,337)
All starchy staple foods	100	99.6	100
Beans and peas	41.6	46.8***	46.8***
Nuts and seeds	4.2	8.8***	7.3***
Dairy	2.9	6.6***	6.1***
Flesh foods	48.5	53.9***	52.3***
Egg	22.1	23.4*	21.5
Dark green leafy vegetables	40.4	53.7***	53***
Vitamin A-rich fruits and vegetables	26.4	28.3	23.6**
Other vegetables	92.1	92.2	92.6
Other fruits	14.6	31.3***	22.1***
MDD-W	30.1	45.5***	39.5***

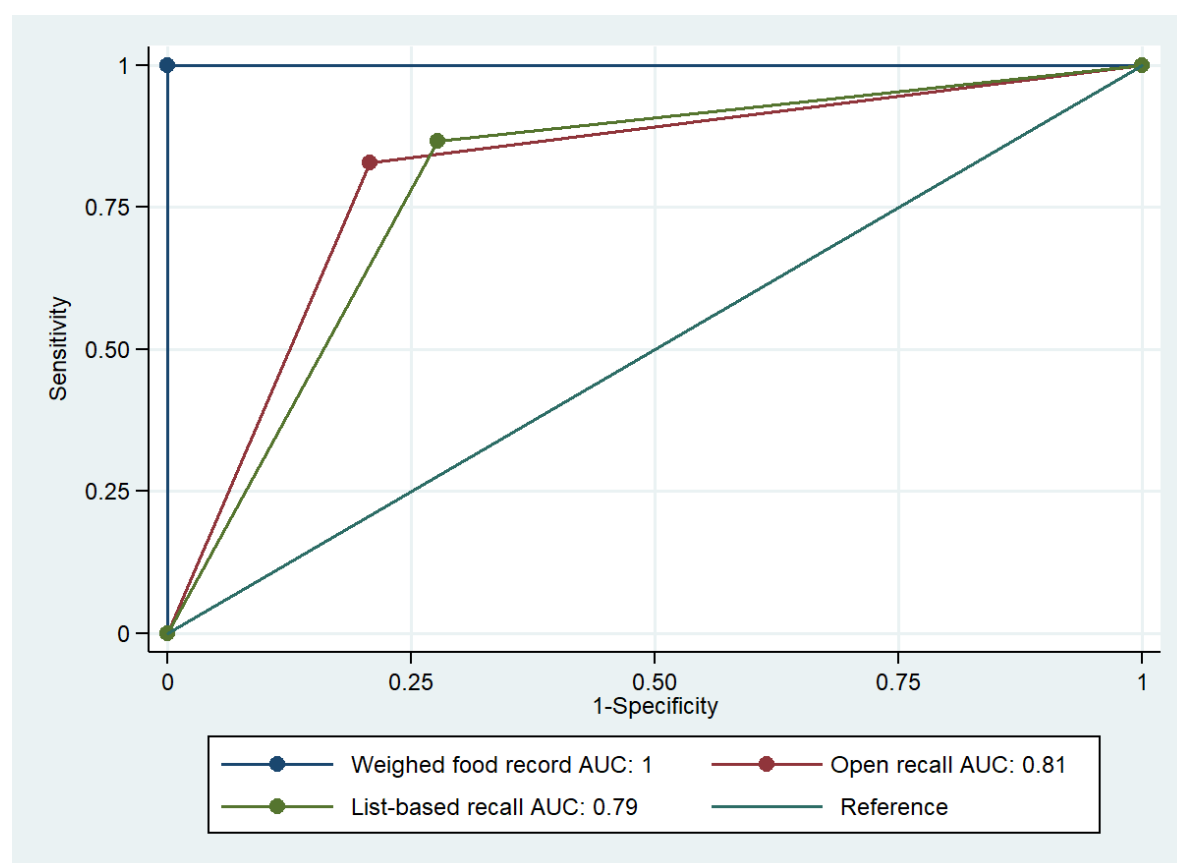
Statistically different proportion between weighed food record and proxy methods are indicated by * $P<0.05$,

** $P<0.01$, and *** $P<0.001$. MDD-W, minimum dietary diversity for women of reproductive age.

Table 2. Agreement between dichotomous Minimum Dietary Diversity for Women of Reproductive Age measured by list-based or open recall as compared to weighed food record.

	Weighed food record		Agreement statistics	
	< 5 food groups n (%)	≥ 5 food groups n (%)	% Agreement	Cohen's kappa
List-based recall				
< 5 food groups	675 (50.5)	54 (4)‡	76.6	0.51***
≥ 5 food groups	259 (19.4)†	349 (26.1)		
Open recall				
< 5 food groups	740 (55.3)	69 (5.2)‡	80.3	0.57***
≥ 5 food groups	194 (14.5)†	334 (25)		

[†]False positive finding (type I error). [‡]False negative finding (type II error). * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

**Figure 2.** Receiver operating characteristic analysis for dichotomous Minimum Dietary Diversity for Women of Reproductive Age measured by list-based and open recall methods as compared to weighed food record. AUC, area under the curve.

3.3. Measurement agreement between list-based or open recall and WFR for food group diversity score

The distributions of food group diversity scores, by dietary assessment method, are presented in supplementary Figure S1. The pooled median (IQR) number of WFR food groups was 4 (3, 5). Ordinal food group diversity scores for the list-based and open-recall were statistically higher than

WFR (both $P < 0.001$; Table 3). The Spearman correlations (ρ) between WFR and proxy methods were 0.67 and 0.70 for list-based and open recall, respectively (both $P < 0.001$).

ICC values ranged from 0.27 to 0.28 in Cambodia, 0.55 to 0.60 in Ethiopia, and 0.43 to 0.52 in Zambia for list-based and open recall, respectively. The agreement of food group diversity score was 47% for list-based and 52% for open recall. Weighted kappa values for the ordinal food group diversity score ($k = 0.47$ – 0.52) were very similar to the simple kappa for the dichotomous MDD-W in all three countries.

Table 3. Agreement between ordinal food group diversity score measured by list-based or open recall as compared to weighed food record.

	Median (IQR range) [†]	ICC (95% CI)	% Agreement	Weighted kappa
Weighed food record (reference)	4 (3, 5)	-	-	-
List-based recall	4*** (3, 5)	0.50 (0.15, 0.85)	47.3	0.47***
Open recall	4*** (3, 5)	0.55 (0.18, 0.92)	52.2	0.52***

[†]Wilcoxon matched-pairs signed-rank test. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$. CI, confidence interval; ICC, intraclass coefficients; IQR, interquartile range.

3.4. Misreporting of food groups by list-based and open recall as compared to the reference method

For the list-based recall, 83% of the measurement errors were type I (false positive values); as compared to 74% for open recall (Table 2). Our individual MDD-W food group confusion matrices indicated that for the list-based recall 69% of type I errors arise from overreporting of beans and peas (11%), dark green leafy vegetables (20%), vitamin A-rich fruits and vegetables (12%), and other fruits (25%) food groups, whereas 31% of type II errors (false negative values) were from underreporting vitamin A-rich fruits and vegetables (Table 4). In parallel for open recall, 68% of type I errors come from overreporting of beans and peas (14%), dark green leafy vegetables (25%), vitamin A-rich fruits and vegetables (11%), and other fruits (19%) food groups, whereas 32% of type II error are from underreporting vitamin A-rich fruits and vegetables.

Table 4. Agreement between 10 Minimum Dietary Diversity for Women of Reproductive Age food groups measured by list-based or open recall as compared to weighed food record.

	Weighed food record		Agreement statistics	
	< 15g n (%)	≥ 15g n (%)	% Agreement	Cohen's kappa
All starchy staple foods				
List-based recall				
< 15g	0 (0)	5 (0.4) [‡]	99.6	1
≥ 15g	0 (0) [†]	1332 (99.6)		
Open recall				
< 15g	0 (0)	0 (0) [‡]	100	-
≥ 15g	0 (0) [†]	1337 (100)		

Beans and peas				
List-based recall				
< 15g	666 (49.8)	46 (3.4) [†]	88	0.76***
≥ 15g	115 (8.6) [†]	510 (38.1)		
Open recall				
< 15g	668 (50)	43 (3.2) [†]	88.3	0.76***
≥ 15g	113 (8.5) [†]	513 (38.4)		
Nuts and seeds				
List-based recall				
< 15g	1199 (89.7)	20 (1.5) [†]	92.4	0.38***
≥ 15g	82 (6.1) [†]	36 (2.7)		
Open recall				
< 15g	1222 (91.4)	17 (1.3) [†]	94.3	0.48***
≥ 15g	59 (4.4) [†]	39 (2.9)		
Dairy				
List-based recall				
< 15g	1247 (93.3)	2 (0.1) [†]	96	0.57***
≥ 15g	51 (3.8) [†]	37 (2.8)		
Open recall				
< 15g	1252 (93.6)	4 (0.3) [†]	96.3	0.57***
≥ 15g	46 (3.4) [†]	35 (2.6)		
Flesh foods				
List-based recall				
< 15g	610 (45.6)	7 (0.5) [†]	93.6	0.87***
≥ 15g	78 (5.8) [†]	642 (48)		
Open recall				
< 15g	624 (46.7)	14 (10.5) [†]	94.2	0.88***
≥ 15g	64 (4.8) [†]	635 (47.5)		
Egg				
List-based recall				
< 15g	993 (74.3)	31 (2.3) [†]	94	0.83***
≥ 15g	49 (3.7) [†]	264 (19.7)		
Open recall				
< 15g	1007 (75.3)	43 (3.2) [†]	94.2	0.83***
≥ 15g	35 (2.6) [†]	252 (18.8)		
Dark green leafy vegetables				
List-based recall				
< 15g	587 (43.9)	32 (2.4) [†]	81.9	0.64***
≥ 15g	210 (15.7) [†]	508 (38)		
Open recall				
< 15g	593 (44.4)	36 (2.7) [†]	81.7	0.65***
≥ 15g	204 (15.3) [†]	504 (37.7)		
Vitamin-A rich fruits and vegetables				

279	List-based recall				
280	< 15g	857 (64.1)	102 (7.6) [‡]	82.9	0.57***
	≥ 15g	127 (9.4) [†]	251 (18.8)		
281	Open recall				
	< 15g	894 (66.9)	128 (9.6) [‡]	83.7	0.57***
282	≥ 15g	90 (6.7) [†]	225 (16.8)		
	Other vegetables				
283	List-based recall				
284	< 15g	47 (3.5)	57 (4.3) [‡]	91.3	0.40***
	≥ 15g	59 (4.4) [†]	1174 (87.8)		
285	Open recall				
	< 15g	42 (3.1)	57 (4.3) [‡]	90.9	0.36***
286	≥ 15g	64 (4.8) [†]	1174 (87.8)		
	Other fruits				
287	List-based recall				
288	< 15g	888 (66.4)	31 (2.3) [‡]	78.7	0.42***
	≥ 15g	254 (19) [†]	164 (12.3)		
289	Open recall				
	< 15g	985 (73.7)	57 (4.3) [‡]	84	0.47***
290	≥ 15g	157 (11.7) [†]	138 (10.3)		

[†]False positive finding (type I error). [‡]False negative finding (type II error). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Several food groups that made up the dichotomous MDD-W indicator and food group diversity score varied significantly ($P < 0.05$) when enumerated by the list-based and open recall methods (Table 1). In Cambodia, comparing proxy methods with WFR, our findings indicate overreporting of beans and peas (~ 6 percentage points (pp)); nuts and seeds (5-9 pp); dark green leafy vegetables (~ 27 pp); and other fruits (23 pp for list-based recall). In Ethiopia, similarly, respondents overreported beans and peas (by ~ 10 pp); dark green leafy vegetables (2-5 pp); and other fruits (2-7 pp) using the list-based and open recall. In Zambia, the picture was slightly different: when comparing proxy methods to WFR, our findings indicate overreporting of flesh foods (8-13 pp), dark green leafy vegetables (9 pp); and other fruits (17-21 pp; supplementary Table S3). Overall, the list-based method reported higher proportions of MDD-W food group consumption than the open recall.

4. Discussion

Our findings indicate poor dietary diversity in surveyed non-pregnant WRA from Cambodia, Ethiopia, and Zambia with only 44% (95% CI: 39-49), 8% (95% CI: 5-11), and 38% (95% CI: 33-42) achieving MDD-W, respectively (based on WFR). Both the list-based and open recall methods did not meet our non-inferiority criterion in predicting dichotomous MDD-W (i.e. > 5 % difference as compared to WFR). Nevertheless, AUCs were 0.79 for list-based and 0.81 for open recall, which were both considered satisfactory for predictive performance (although statistically different from WFR).

Food group diversity scores measured by the proxy methods had moderate agreement, but were also statistically different when compared to WFR (although median (IQR) were identical). In all countries, both the list-based, in particular, and open recall were inclined to misreport consumption of specific MDD-W food groups. To illustrate, overreporting for both proxy methods as compared to

WFR, exceeded 10% for beans and peas, nuts and seeds, dairy, dark green leafy vegetables, and other fruits.

We attempted to identify the specific foods and beverages (and/or potential recall biases) that were accountable for the misreporting of food group consumption by looking in more detail at the list-based and open recall, as compared to the WFR data. The overreporting of food groups with list-based and open recall are, in part, due to the reporting of food items not consumed in sufficient quantities (≥ 15 g; e.g. nuts and seeds often used as “condiments and seasonings”). This is consistent with findings from Bangladesh and India (8) and Burkina Faso (9), where misreporting was common for ingredients used in composite dishes or in small quantities for sauces. The overreporting of certain food groups, not consumed according to the reference method, could be a result of memory lapses, social desirability and/or social approval bias e.g. flesh foods, dairy, eggs, certain sorts of vegetables (3).

The poorer agreement and lower correlation among proxy methods and WFR in Cambodia as compared to Ethiopia and Zambia need further examination. In all three countries, we conducted theoretical and practical enumerator capacity development workshops on dietary assessment methodologies, with a standardised adaption of questionnaires. The nature of the diet in Cambodia is the most complex, with numerous mixed dishes containing a large number of ingredients. Nevertheless, special focus was placed during the enumerator training to make sure that composite dishes and foods items with multiple ingredients were allocated in the correct food groups. As previously described, ingredients with trivial quantities mainly used to add flavour were to be classified in the “condiments and seasonings” food group.

The *Minimum Dietary Diversity for Women: A Guide to Measurement* describes the use of two proxy qualitative methods (i.e. list-based or open recall) to measure food group diversity (7), however no evidence existed on the validity of these two non-quantitative methods at the time of publication. Our multi-country study used WFR, which is accepted as the reference method for dietary assessment (3). At present, however, there is no true gold standard for dietary intake, since all methodologies entail error. Therefore, our comparisons of the dichotomous MDD-W indicator and ordinal food group diversity score calculated from list-based and open recall against WFR might still be considered as a type of relative validity. Both the list-based and open recall methods in our study covered the same period of recall (i.e. the day and night prior to the survey, on which WFR was conducted), thereby preventing differences related to varying time frames.

To the authors’ knowledge, this is the first multi-country study that uses dietary intake data from countries in sub-Saharan Africa and Southeast Asia, with a wide range of food cultures, to assess the validity of two proxy methods as compared to WFR for non-pregnant WRA. Strengths of the study include that we compare the two qualitative methodologies described for MDD-W against WFR (i.e. reference method of individual quantitative dietary assessment). Furthermore, our study collected data, with high-quality capacity development and methods, resulting in high-quality dietary intake data, with sufficiently large sample sizes per country. The present findings are timely, given that there is a widespread uptake of MDD-W (e.g. Demographic and Health Surveys-8 (22)) and strong support to scale-up the indicator at country-level from the European Union, United Nations agencies, and international NGOs (e.g. national information systems and nutrition-sensitive agricultural programmes).

Our simple Cohen's kappa and AUC results showed similar reliability and accuracy between the list-based and open recall, demonstrating that list-based data collection performed almost as well as open recall in estimating proportions of WRA achieving MDD-W. These findings have important implications, because the list-based recall has several operational advantages, such as lower enumerator capacity requirements and shorter training time, which in turn might reduce the cost of MDD-W data collection. Nevertheless, our findings confirm statistical inferiority (i.e. overreporting) of both list-based and open recall, as simple tools to measure dietary diversity in WRA. Furthermore, for assessing the prevalence specific MDD-W food group consumption, estimates from list-based and open recall differed quite substantially for distinct food groups in each country, so additional methodological work from a wider range of (urban) contexts and seasons is required to understand the potential of using simple proxy methodologies to assess the consumption of individual food groups.

This research was part of a Food and Agriculture Organization of the United Nations led project that will result in updated operational guidance for measuring MDD-W. The guidance will also be informed by operational data, which will factor into a recommendation on the choice of MDD-W data collection method.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: Distribution of ordinal food group diversity score, by dietary assessment method, Table S1: Associations between proportions reaching achieving Minimum Dietary Diversity for Women of Reproductive Age from proxy methods and weighed food record proportion, Table S2: Food categories used during list-based recall and of Minimum Dietary Diversity for Women of Reproductive Age in Cambodia, Ethiopia, and Zambia, Table S3: Proportions of non-pregnant women (15-49 years) having consumed food groups and achieved Minimum Dietary Diversity for Women of Reproductive Age in Cambodia, Ethiopia, and Zambia, based on weighed food record, list-based, and open recall methods.

Author Contributions: Conceptualization, Giles T Hanley-Cook, Ji Yen A Tung, Patrick W Kolsteren, Maria Antonia G Tuazon and Carl K Lachat; Data curation, Giles T Hanley-Cook, Pamela A Marinda, Kong T Thong and Dilnesaw Zerfu; Formal analysis, Giles T Hanley-Cook; Funding acquisition, Maria Antonia G Tuazon; Investigation, Giles T Hanley-Cook, Ji Yen A Tung, Isabela F Sattamini, Pamela A Marinda, Kong T Thong and Dilnesaw Zerfu; Methodology, Giles T Hanley-Cook, Maria Antonia G Tuazon and Carl K Lachat; Project administration, Ji Yen A Tung, Isabela F Sattamini, Maria Antonia G Tuazon and Carl K Lachat; Resources, Maria Antonia G Tuazon and Carl K Lachat; Software, Giles T Hanley-Cook and Carl K Lachat; Supervision, Patrick W Kolsteren, Maria Antonia G Tuazon and Carl K Lachat; Validation, Pamela A Marinda, Kong T Thong and Dilnesaw Zerfu; Visualization, Giles T Hanley-Cook; Writing – original draft, Giles T Hanley-Cook; Writing – review & editing, Giles T Hanley-Cook, Ji Yen A Tung, Isabela F Sattamini, Pamela A Marinda, Kong T Thong, Dilnesaw Zerfu, Patrick W Kolsteren, Maria Antonia G Tuazon and Carl K Lachat. All authors have read and agreed to the published version of the manuscript.

Funding: Financial support from the German Federal Ministry for Food and Agriculture (BMEL) and technical leadership from the Food and Agriculture Organization of the United Nations (FAO).

Acknowledgments: We gratefully acknowledge the data collection by the Faculty of Agro-Industry, Royal University of Agriculture Cambodia, Food Science and Nutrition Research Directorate of the Ethiopian Public Health Institute, and the Department of Food Science and Nutrition, University of Zambia, with support from Boran Altincicek. We acknowledge the invaluable comments from the Mary Arimond, Terri Ballard, Estefania Custodio, Gina Kennedy, Markolf Maczek, Yves-Martin Prével, and Pattanee Winichagoon. We heartily thank all the women for participating in this multi-country study.

Conflicts of Interest: Ji Yen A Tung, Isabela F Sattamini, and Maria Antonia G Tuazon are FAO staff members. The opinions and statements in this manuscript are those of the authors and may not reflect official FAO policies. The authors declare that they have no conflicts of interest. The funders of this multi-country study had no role in study design, data collection, analysis, interpretation, or writing of the manuscript. The corresponding author has full access to all of the data in the study and had the final responsibility for deciding to submit the manuscript for publication.

References

1. Afshin A, Sur PJ, Fay KA, Cornaby L, Ferrara G, Salama JS, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2019;393(10184):1958–72.
2. Arimond M, Wiesmann D, Becquey E, Carriquiry A, Daniels MC, Deitchler M, et al. Simple Food Group Diversity Indicators Predict Micronutrient Adequacy of Women's Diets in 5 Diverse, Resource-Poor Settings. *J Nutr*. 2010;140(11):2059S–2069S.
3. Gibson RS, Charrondiere UR, Bell W. Measurement Errors in Dietary Assessment Using Self-Reported 24-Hour Recalls in Low-Income Countries and Strategies for Their Prevention. *Adv Nutr An Int Rev J*. 2017;8(6):980–991.
4. Development Initiatives. Global Nutrition Report 2018: Shining a light to spur action on nutrition [Internet]. Claydon J, editor. Bristol: Development Initiatives Poverty Research Ltd; 2018. 3–165 p. Available from: <http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/128484>
5. Ruel MT, Deitchler M, Arimond M. Developing Simple Measures of Women's Diet Quality in Developing Countries: Overview. *J Nutr*. 2010;140(11):2048S–2050S.
6. Martin-Prevel Y, Arimond M, Allemand P, Wiesmann D, Ballard TJ, Deitchler M, et al. Development of a Dichotomous Indicator for Population-Level Assessment of Dietary Diversity in Women of Reproductive Age. *Curr Dev Nutr*. 2017;1(12):e001701.
7. FAO, FHI 360. Minimum Dietary Diversity for Women- A Guide to Measurement. Minimum Dietary Diversity for Women: A Guide for Measurement. 2016.
8. Nguyen PH, Martin-Prevel Y, Moursi M, Tran LM, Menon P, Ruel MT, et al. Assessing Dietary Diversity in Pregnant Women: Relative Validity of the List-Based and Open Recall Methods. *Curr Dev Nutr*. 2019;4(1):nzz134.
9. Martin-Prevel Y, Becquey E, Arimond M. Food Group Diversity Indicators Derived from Qualitative List-Based Questionnaire Misreported Some Foods Compared to Same Indicators Derived from Quantitative 24-Hour Recall in Urban Burkina Faso. *J Nutr*. 2010;140(11):2086S – 93.
10. Lachat C, Hawwash D, Ocké MC, Berg C, Forsum E, Hörnell A, et al. Strengthening the Reporting of Observational Studies in Epidemiology – Nutritional Epidemiology (STROBE-nut): An extension of the STROBE statement. *PLoS Med*. 2016;13(6):e1002036.
11. NCSS LCC. PASS 16 Power Analysis and Sample Size Software [Internet]. Kaysville, Utah, USA; 2018. Available from: ncss.com/software/pass.
12. GIZ. Appropriate Diet for All: Cross-country evaluation of nutrition outcomes [Internet]. Bonn, Germany; 2019. Available from: [https://www.giz.de/en/downloads/GIZ_GV ESRS_Cross country_evaluation_2019.pdf](https://www.giz.de/en/downloads/GIZ_GV_ESRS_Cross_country_evaluation_2019.pdf)
13. FAO. Dietary Assessment: A Resource Guide to Method Selection and Application in Low Resource Settings. Rome, Italy; 2018. 152 p.
14. Gibson RS. Principles of Nutritional Assessment. 2nd ed. New York: Oxford University Press; 2005. 908 p.
15. US Department of Agriculture. Food Yields Summarized by Different Stages of Preparation. Agriculture Handbook. Washington, DC: US Department of Agriculture; 1975. 102 p.
16. Sibbald B, Roberts C. Understanding controlled trials Crossover trials. *BMJ*. 1998;316(7146):1719–20.
17. Mills EJ, Chan AW, Wu P, Vail A, Guyatt GH, Altman DG. Design, analysis, and presentation of crossover trials. *Trials*. 2009;10(27).
18. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977;33(1):159–74.
19. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–8.
20. McGraw KO, Wong SP. Forming Inferences about Some Intraclass Correlation Coefficients. *Psychol Methods*. 1996;1(1):30–46.
21. StataCorp LLC. Stata Statistical Software: Release 15. College Station, Texas, USA; 2017.
22. FAO. Minimum Dietary Diversity for Women (MDD-W) Indicator [Internet]. Rome, Italy; 2019. Available from: fao.org/nutrition/assessment/tools/minimum-dietary-diversity-women



