Development and validation of a photographic food atlas for portion size assessment in the southern plains of Nepal

ABSTRACT

Objective: To develop and validate a photographic food atlas of common foods for dietary assessment in southern Nepal.

Design: We created a life-sized photographic atlas of 40 locally prepared foods. Between March and June 2014, data collectors weighed portion sizes that respondents consumed during one mealtime, and then a different data collector re-visited the household the next day to record respondents' estimations of their previous day's intakes using the atlas. Validity was assessed by percentage error, weighed Cohen's kappa and Bland-Altman limits of agreement.

Setting: Dhanusha and Mahottari districts in southern Nepal.

Subjects: A random sample of 95 adults in 48 rural households with a pregnant woman.

Results: Overall, respondents underestimated their intakes (mean % error = -4.5). Rice and *dal* (spiced lentil soup) intakes were underestimated (-14.1% and -34.5% respectively), but vegetable curry intakes were overestimated (+20.8%). Rice and vegetable curry portion size images were significantly reliably selected (Cohen's kappa (SE)) (rice = 0.391 (0.105)) (vegetable curry = 0.430 (0.139)), whereas *dal* images were not. Energy intake over one mealtime was underreported by an average of 136 kcal (4.5% error) using recall compared with weighing method.

Conclusions: The photo atlas is a useful tool for field estimation of dietary intake. Average errors were low, and there was 'modest' agreement between weighed and recalled portion size image selection of rice and vegetable curry food items. Error in kcal estimation was low but with wide limits of agreement, suggesting that there is scope for future work to reduce error further.

Keywords: validation, food atlas, portion size, dietary intake

INTRODUCTION

With between 27 and 41% of the South Asian population underweight and 8 to 41% overweight, accurate, affordable and appropriate methods for measuring dietary intake are needed [1, 2]. However, few dietary intake methods have been tailored to the South Asian context, where literacy rates are low [3, 4] and the burden of data collection falls on literate interviewers [5]. Interviewer-led methods may be prospective, such as weighed food records, or retrospective, such as food frequency questionnaires or 24-hour dietary recalls that rely on respondent recall to quantify their intakes [6, 7]. Prospective and retrospective methods have different sources of error, such as modified eating patterns for weighed methods or response bias for recall methods [8], but total levels of error are similar [9].

In the resource-constrained context of South Asia, recall methods are often chosen over weighed methods because they are cheaper, quicker, and more feasible for large sample sizes, and they are less intrusive so more culturally appropriate [10, 11]. For instance, recall methods take a short time whereas weighed methods take at least one full day and so recall methods are less burdensome on the (traditionally female) cook, who may have a high workload and pressure to fulfil her duties at home [12, 13]. Also, unlike weighed methods, recall methods do not require interviewers to come into contact with food. In many Hindu households, leftover food is considered ritually unclean (‘*jutho’*), and so weighing leftover food may not be permitted [14] and there might be issues with respondents perceiving that interviewers of certain castes are ‘polluting’ the kitchen [15, 16].

Despite the financial and practical benefits, a major limitation of recall methods is that they rely on respondent memory. Photographic atlases of graduated portion sizes, 3-dimensional food models, utensils and new computer-based mobile methods have all been used as aids for respondents to quantify their intakes [7, 17-19]. Evidence suggests that there is little or no benefit of using food models instead of photographs. One study found limited benefit from using food models instead of photographs, with models performing better, equally and worse than 2-dimensional images for various food types [20]. Another study found that photographs resulted in more accurate estimations than food models and measuring cups [21].

To our knowledge, no studies have tested the validity of a South Asian photo atlas on adults in this context. One study tested a photographic atlas on 80 children in Sri Lanka [22]. The authors reported that 57% of portion size estimations using life size photographs were estimated correctly (i.e. respondents selected the closest portion size image) and the ratio of estimated and actual weights was close to 1 but with wide confidence intervals. Another study from Pakistan reported that 76% to 100% of 21 respondents selected the correct portion size [23].

These studies, and numerous others not from South Asia [24-33], have reported bias (percentage error), percentage of correct photographs selected, ratios between estimated and weighed portions and / or correlations between weighed and recall methods. These are important measures for tools that are developed to aid the estimation of group-level mean nutrient intakes and risks of nutrient deficiency [5]. However, for studies aiming to assess diet at the individual level, these measures may mask large measurement errors between individuals, fail to account for image selection that would occur by chance, and show association but not agreement between weighed and recalled estimates. This means that, for studies aiming to assess diet at the individual level, additional measures of validity are needed.

Globally, few studies have validated photo atlases for individual dietary assessment, that is, using appropriate measures of agreement such as Bland-Altman limits of agreement or Cohen's kappa. To illustrate this, a non-systematic review of studies that did report agreement between weighed and recalled methods is summarised in Table 1 [34-39]. No studies were available from South Asia and only one study reported limits of agreement in terms of nutrient intakes. This step of converting portion sizes to nutrient intakes may be useful for showing the nutritional implications of bias associated with different food items.

This paper addresses these research gaps by assessing the validity of a South Asian photographic food atlas using measurements of agreement between weighed and estimated portion sizes in terms of grams and nutrient intakes. The paper also describes the cultural and practical challenges of creating and validating the atlas in the plains of Nepal.

MATERIALS AND METHODS

Study setting and population

The study was conducted in Dhanusha and Mahottari districts in the *Terai* (southern plains) region of Nepal. This site was chosen because the photographic atlas under test was intended for subsequent use in the same districts in a cluster randomised controlled trial. The Low Birth Weight South Asia Trial (LBWSAT http://www.controlled-trials.com/ISRCTN75964374) was conducted by Mother and Infant Research Activities (MIRA) and University College London (UCL) Institute for Global Health, in partnership with the World Food Programme, Save the Children and the Institute of Fiscal Studies. It tested the effect of a pregnancy-focussed behaviour change intervention, with or without food or cash transfers, on newborn weight and infant weight-for-age. The photographic atlas was intended for a sub-study using a 24-hour dietary recall method to measure the trial effects ~~up~~on intra-household food distribution between pregnant women, their mothers-in-law and the male household heads. The main outcome of the sub-study is relative dietary energy adequacy ratio and secondary analyses refer to protein and iron.

The study districts, located on the Indian border, have a predominantly Maithili-speaking population. Poor quality roads, frequent flooding during the monsoon and high temperatures make travel difficult in the remoter parts of these districts. Being in the Gangetic plains, the land is flat and fertile and used mainly for production of rice, wheat, pulses and vegetables. Despite high food production, only 50% of households in the *Terai* are classified as food secure and there is high (30%) prevalence of underweight and anaemia in women [40], so measurement of intra-household food allocation may help to explain the causes of undernutrition in this region.

Sampling strategy

From March to June 2014, three local, Maithili-speaking, female data collectors (MK, NM, JT) conducted a cross-sectional survey in 48 households. Our sampling frame of respondents matched that of the intra-household food allocation study for which the atlas was intended. That is, we sampled the pregnant woman and, if available, the mother-in-law and male household head. Given financial and time constraints, 101 respondents were interviewed and 95 used the photographic food atlas to estimate their intakes for at least one food item. In order to reach this sample size, we randomly sampled 48 households from a list of pregnant women in their third trimester of pregnancy who were already enrolled in our trial. We sampled households sequentially until we reached 48 households and 101 respondents. To avoid being too intrusive and to capture eating on 'normal' days, we did not sample on special celebratory feasting or fasting days where households ate more or less than usual, or ate special types of food.

Development of the photographic atlas

To develop the photographic food atlas, we followed guidance from Nelson and Haraldsdóttir [41]. Working from a food list prepared for another study [42], photographs were taken of a range of commonly consumed foods, some of which were amorphous dishes (such as curry or rice), some of which were discrete items that vary in size (such as large and small mangoes) and some of which were volumes (such as spiced lentil soup, or '*dal*'). Local women and vendors from rural villages surrounding Janakpur town (Dhanusha district headquarters) prepared the dishes. We initially chose serving sizes using published data on median portion sizes from Nepal [43], to find a midpoint portion size. Local colleagues deemed some of these values implausible in this context, so, for those implausible values, we chose a different midpoint and size of increment. Different midpoints and increments were selected by asking local community members from nearby villages what a 'typical', 'small' and 'large' portion looked like, and corroborating their answers with responses from other community members. All portions were weighed accurate to 0.1 g using digital Tanita KD321 weighing scales. A study that tested which camera position was most effective (aerial or angled) showed no significant differences [25] so, in the same way as Turconi et al. [39], pictures were taken at an approximate 45^{O} angle to capture both the depth and width of the portion. The final photo atlas contained 40 different food items, with up to six different portion sizes per item; common or nutritionally important items such as rice had more options, whereas rare or small items like nuts or chutney had fewer.

Images were scaled to life size, according to findings from Thoradeniya et al. [22] that found more accuracy with life size photographs than small photographs or household utensils. The background and utensil was removed and the food image was superimposed onto an image of a plain utensil to keep the images consistent and to minimise distraction from other non-food variation. Images were processed using Microsoft Word, GNU Image Manipulation Programme (GIMP©) and Adobe Photoshop© and printed in colour. Figure 1 shows some examples of portion images, the sizes of these portions, and the cut-offs within which a selected image would correctly represent a given portion.

Validation process

Female interviewers conducted the validation study, over a two-day period per household. In all households, the cook was a woman and it was essential that the data collectors were also female because they needed to spend prolonged periods of time together. Because the data collectors were high caste we experienced no problems entering and working in the kitchen, although they were careful to respect the kitchen space and would usually work near to (but just outside) the kitchen where possible.

On the first day, for each respondent, data collectors recorded food items consumed, portion sizes of all servings, and the weights of any leftovers over one mealtime, using paper forms and weighing scales accurate to 0.1 g (Tanita KD321, Goldtech) and 0.5 g (Goldtech, ClaTronic). Weighing scales broke frequently, perhaps due to the hot and humid climate, so we gave interviewers calibration weights to check the scales before every interview and replaced scales when needed. We found that weighing a ritually unclean *jutho* plate, from which a person had already eaten, made the weighing scale *jutho* by transference, and so it was not appropriate to weigh new portions on a scale that had previously weighed leftovers. Leftovers were weighed on a separate scale, although the process remained socially uncomfortable. Respondents also reported whether leftovers were mixed with other foods and to whom any leftovers were given.

On the second day, to reduce interviewer bias, a different data collector asked the same respondents to estimate how much they each ate the previous day, using the photographic atlas. A full 24-hour recall was obtained but the corresponding recalled portions that were weighed the previous day were matched for the validation analyses. To ensure that recall data was as accurate as possible, we used a ‘five-stage multi-pass’ method that has been shown to reduce underreporting [44] in conjunction with the photographic atlas. In brief, respondents were probed to describe their food intake over the previous 24 hours using these five different ‘passes’ [44, 45]:

(1) Collect a free recall, using non-specific probes, starting from when the respondent woke up the previous morning

(2) Probe using a standard commonly forgotten foods list (such as supplements, alcoholic drinks and fruit)

(3) Ask for the time and place that each item was consumed

(4) Collect portion size information using the atlas and clarify the exact food types

(5) Use a series of final probes (referring to snacks and food eaten outside of the home) and recap all recorded foods in chronological order.

On both days, data collectors recorded the food items by entering a 4-digit food code (rather than food name) on a paper form. Because of the large number of food items, food names and their corresponding codes were listed on an Android application (OpenDataKit, ODK Collect 1.4.3; an open-source, cloud-based platform)[(46)] that the data collectors used to look up food items and find the correct code.

Data collectors were trained to put the respondents at ease and to be non-judgemental about food intake, and they were provided with a training manual with guidelines on how to minimise social desirability bias and examples of non-leading probes that they could use. Because anthropometric status is thought to be associated with response bias [(47, 48)] mid-upper arm circumference (MUAC) of all respondents, and weight and height of non-pregnant respondents, were taken using Seca circumference tapes, Tanita solar scales 302, and Shorr Board stadiometers respectively.

Kilocalorie (kcal) intakes were calculated using a food composition table that HHF compiled from multiple sources, including the US Department for Agriculture [(49)], McCance and Widdowson The Composition of Foods Integrated Dataset 2015 [(50)], Bangladesh Food Composition Table [(51)], Nepal Food Composition Table, [(52)] and other peer-reviewed published sources for rare items. For a few items, such as supplements and some locally packaged foods, nutritional data on the packets were used.

For mixed dishes made with multiple ingredients, data collectors collected 174 local recipes during the creation of the atlas, piloting, and the validation study. The number of recipes for each mixed dish depended on how common the dish was, ranging from between 1 and 32 recipes per dish. All raw ingredients and the final weight of the mixed dish were weighed and the nutritional content was calculated by summing the nutritional contents from all raw ingredients and calculating the summed nutrients as a proportion of the final dish weight. This was then reported as nutrients per 100 g of the mixed dish. For food items with more than one recipe, the average nutritional composition was calculated. For items with no recipe (for example rare meat curries or out-of-season vegetable curries), the most similar recipe was used and the main ingredient was substituted. For example, to create a duck meat curry recipe, duck meat was replaced with goat meat and the rest of the curry ingredients were kept the same. A total of 127 dish recipes were analysed from 174 locally collected recipes, 45 imputed recipes (based upon substitutions using locally collected recipes), 3 published recipes [(51, 53)] and 6 recipes from various online sources that were referenced in full in the food composition table.

The validation study method was iteratively modified during a series of pilot studies in 16 households. Data collectors received eight days of training in the office, and then practised the validation seven times each in nearby villages. None of the pilot or practice data were included in the results because the method changed substantially during the piloting process, and the practice data were expected to have high levels of error.

Supervisors (HHF and PP) monitored 10% of the interviews and completed an observation checklist to ensure adherence to protocol. The checklist items included: obtained consent, had all equipment in clean and working order, kept weighing scale on a flat surface, used the tare function on the scale correctly, reported leftover food, all sections of the form completed, non-judgemental interviewing technique. Supervisors also checked data and resolved any illogical or missing data by discussion with the data collectors. Data were then entered into an MS Excel database and checked for errors.

Analysis

Total weighed portion for a particular food was calculated as the sum of all servings, minus any leftover food. The total portion included any shared foods that were originally served to someone else. Weights of shared and leftover foods that were mixed with other foods (such as rice and spiced lentil soup mixed together) were estimated by assuming equal proportions of food items in the first serving as in leftovers or shared foods.

Bias was calculated as percentage error: 100 × (recalled portion - weighed portion) / weighed portion. Cohen's weighted kappa (κ_w) statistic was calculated to assess the agreement between the selection of portion size images and the portion size image that should have been selected according to the weighed portion [55]. To do this, the weighed portion size was converted into an ordinal variable to represent the image number that the respondent should have chosen. The cut-off points were the midpoints (shown in Figure 1) between each portion size in the atlas (Figure 1 and Table 3). Respondents were allowed to choose in-between two portion sizes but, because few respondents used this option (and therefore the atlas was used without this option in a later study), these observations were excluded from the analyses. Analyses with these 'in-between' values produced similar results.

The kappa statistic adjusts for agreement in selection of portion sizes that might occur due to chance, and quadratic weights allowed for partial agreement, giving proportionally larger penalties for greater distances between observed and selected images. For example, if a respondent ate a portion size of 10 g and had an option of three images depicting 10, 50 and 90 g, then image 1 would be the best option with perfect agreement (weighted 1), image 3 would show no agreement (weighted 0) and image 2 showing 50 g would be worse than image 1 but better than image 3 (weighted 0.75).

Bland-Altman plots for intakes of kcal, protein and iron were used to show the agreement between weights and recall estimates [54]. These show the difference in nutrient intake between recalled and weighed portions plotted against the mean intakes calculated by the two methods. Limits of agreement (LOA) at 5% significant level were calculated as mean kcal difference +/- 1.96 SD. Confidence limits for the mean error were calculated as the mean +/- 1.96 SE of the mean, and for the LOA as the limit +/- 1.96 SE of the limits. Standard error of the limits was approximated as $\sqrt{3\, s^2/n}$ because of the smaller number of scores at the limits [56]. Bland-Altman plots and limits of agreement for portion size weights were not calculated for all 245 recalled portions because each respondent reported multiple portion sizes, and so the assumption of independence does not hold.

Intraclass correlation coefficients (ICC) were used to assess the strength of possible within-household clustering expressed as a random effect. Limits of agreement for individual food items were not reported because the estimated portion sizes were ordinal, rather than continuous, and so Cohen's kappa statistic was deemed more appropriate. Non-parametric methods were used to measure associations between respondents and percentage error in kcal estimation because percentage error was negatively skewed. Statistical significance was defined at 5% level. All analyses were performed using Stata SE 14 (*StataCorp 2015. Stata Statistical Software: Release 14. College Station. TX: StataCorp LP).*

RESULTS

Response rate

Figure 2 shows the response rate at the individual and household levels. We visited 58 households to obtain our target of 48 households (83% response rate). Seven households were empty and three refused. Within these 48 households we aimed to sample three household members - the pregnant woman, household head, and mother-in-law. This gave a maximum of 144 potential respondents. However, in some cases, the pregnant woman or mother-in-law was also the head of the household (Figure 2) and so only one or two household members could be sampled. For instance, if the mother-in-law was also the household head, she was sampled along with the pregnant woman, giving only two respondents. If the pregnant woman was the head of a nuclear household, that is, not living with her in-laws, then only the pregnant household head was sampled. Some household members were temporarily unavailable or not living in the home, and a few did not use the photographic atlas to estimate their portion sizes because they consumed discrete food items (such as bananas) for which no atlas images had been created. In total, we obtained dietary recalls from 95 individuals (58% total response rate), and 245 validated portion sizes estimations.

The total kcal intake over one main meal from this sampled number of individuals (n = 95) ranged from 90 to 2246 kcal, with mean 823 kcal and standard error of 34 kcal. The mean and SD of the differences in kcal intake estimated from using the photographic atlas and the intakes calculated from weighed food portions were 138 and 351 kcal. Following Bland and Altman [54], these summary statistics and sample size ensure a confidence limit of agreement with length 62 kcal.

Study population and diet characteristics

Respondent characteristics are provided in Table 2. The average respondent age was 36 years and 76% respondents were women (all household heads were male). These variables have been described as possible determinants of recall estimates [41]. MUAC was used as a comparative anthropometric measure for all respondents because Body Mass Index is difficult to interpret during the third trimester of pregnancy. In most households the pregnant woman was the main cook (83%).

The 245 portions, estimated by recall using the food atlas, mainly came from the six most frequently consumed items, plus 25 other portions for other food items. The mean bias associated with the six most frequently consumed items, and the overall mean bias from all 245 portions, is shown in Table 3. This overall mean bias shows that respondents tended to underestimate portion sizes by 4.5% (SE=3.9). Rice and *bhujiya* (spiced fried potato) had the smallest bias (-11% and -13% respectively) whereas *sag* (green leafy vegetables, cooked with salt and oil) had the largest (+40%).

Selection error and weighted kappa (κ_w) for rice, *dal* (spiced lentil soup) and vegetable curry portion sizes are shown in Table 4. The selection error shows how close respondents were to choosing the correct portion size image. The portion sizes for rice, *dal* and vegetable curry depicted in the atlas and the cut-off points for the selection of each image were shown in Figure 1. Over three quarters of the respondents chose the correct portions to within one image larger or smaller. For rice and vegetable curry, selection of portion sizes was significantly better than chance (κ_w 0.39 and 0.43 respectively) whereas, for dal, there was no significant agreement in choice of portion size.

Correlation coefficient between kcal intakes calculated from weighed and recalled portion sizes of individual dishes was 0.446 (p<0.0001). The Bland-Altman plots showed agreement between weighed and recalled measures of kcal, protein and iron intakes (Figure 3, 4 and 5 respectively).

Since Bland-Altman plots rely on independence of scores, we measured clustering of household members' mealtime kcal intakes within households. One-way random effects regression models found very low intraclass correlation coefficients between pregnant women and their mothers-in-law (ICC=0.003; *n*=29), pregnant women and the household head (ICC=0.160; *n*=18) and mothers-in-law and household heads (ICC=-0.016; *n*=14).

Recalled measures of kcal intakes per respondent (over the one mealtime that was validated) were underreported by an average 138 kcal (95% CI 67, 208). 95% limits of agreement between weighed and recalled methods were -551 and 826 kcal. Protein intakes were underreported by 3.7 g (95% CI 1.7, 5.6) and 95% limits were -15.3, 22.7 g. For iron, intakes were underreported by 0.5 mg (95% CI 0.1, 0.9) and 95% limits were -3.8, 4.8. Unlike kcal and protein plots, the iron plot shows heteroscedasticity, with agreement decreasing as iron intakes increased.

We checked for plausibility of outliers, and differences in respondent characteristics between outliers and non-outliers. The outlier in Figure 4, where the respondent had much higher intakes of protein, was mainly due to consumption of a large portion of meat curry. Outliers were defined as % error in kcal estimation of >75 or <-75; *n*=8. They were not significantly associated with gender (Fisher's exact test, *p*=0.675); age ((odds ratio (95%)) (1.01 (0.96, 1.20), *p*=0.735), years of education ((0.97 (0.96, 1.06)), *p*=0.809) or mid-upper arm circumference ((1.35 (0.99, 1.86)), *p*=0.055). Outliers were also evenly distributed between the three female interviewers.

Univariable analyses found no association between percentage error in kcal estimation and gender (Wilcoxon rank-sum test (*z*=0.113, *p*=0.910), education category (any or no years of schooling) (*z*=-0.175, *p*=0.861), age (Spearman's correlation (ρ=0.062, *p*=0.551), or MUAC (ρ=-0.069, *p*=0.548). Multivariable quantile regression to adjust for possible confounding gave similarly non-significant results.

DISCUSSION

The photo atlas was a useful aid because it enabled the estimation of dietary intakes in populations with low literacy levels, using affordable, practical and culturally appropriate methods. The overall underestimation error of 4.5 percent was small compared with the typical range of between 5% and 100% error reported by Nelson and Haraldsdóttir [(57)].

The different directions and variance in error associated with different food items illustrated the importance of measuring agreement instead of only mean differences. Around 85% of respondents chose the correct portion to within one option bigger or smaller. Although agreement in image selection calculated from weighted Kappa results (0.43 and 0.39 for vegetable curry and rice) could be categorised as 'modest' [(58)], it is significantly better than random selection. The small bias for rice is important because it is the staple food and so the main source of energy. Only one other study measured agreement using weighted kappa and it found better agreement than our study (κ_w 0.60 compared with κ_w 0.24) [(34)].

One possible reason for the higher percentage error in our population is that respondents may be less able to conceptualise portion sizes and less practised in estimating measures. Also, there may have been more coding error from matching recalls with their corresponding weighed portions because recalls were collected over a full 24-hour period whereas weights were only collected for a single meal. People often ate sequentially rather than together in one sitting, and the person eating would eat in private because it was considered rude to eat in front of others who were not eating. This meant that it was sometimes difficult for the data collector to see if all the food was eaten, or if the cook had quickly served another portion on the respondent's plate. The data collectors paid close attention to record any additions or leftovers as far as possible. Alternatively, the difference in agreements may be attributed to the comparative heterogeneity in our sample (we included pregnant women, older women and men rather than only women of reproductive age). Although no significant effects of respondent characteristics (such as age or gender) were found in our study, this may be due to insufficient statistical power rather than absence of a trend.

There is also an intrinsic, random error that exists from using any photographic atlas because it converts continuous portion sizes into ordinal portions. As actual portion sizes decrease, this error increases; for instance, a difference of 100 g in a large actual portion size of 900 g is 11.1%, but in a small actual portion of 100 g the error is 100%. This error approaches infinity as actual portion sizes approach zero. Since intervals between portion size images are approximately equal, this intrinsic error will be larger (despite still selecting the closest portion image) if actual intake distributions are closer to the lower end of the atlas scale and depending on the intervals between portion sizes. Therefore, differences in percentage error between studies may exist if the respondents were equally discriminant and absolute differences in portion size estimation were equal, but respondents' actual portion sizes were different. This variance in random error is complicated by the trend for agreement to decrease as portion sizes increase, as shown in the heteroscedasticity in agreement of iron estimations (Figure 5) and in agreement shown elsewhere [39].

Limits of agreement between estimated and weighed measurements were wide, although part of this will be explained by the intrinsic error of the ordinal portions in the atlas. 95% limits were wider than those in the one other study that reported limits (for example -551 to 826 kcal, compared with 49 to 162 kcal) [36]. This was to be expected, however, since the other study tested a novel method using photographs taken by the respondents to assist respondents with their the portion size image selection [36].

Strengths and limitations

The lack of agreement associated with *dal* (spiced lentil soup) may be because it is often spooned directly over rice, and so images of ladles may have been more appropriate than bowls. Also, the recipes showed that the thickness of the *dal* ~~was~~ varied, and so the densities of *dal* in households may have been different from the density of *dal* depicted in the image. Therefore, respondents may have chosen the image that best represented the volume, but not the gram weight, of their portions.

Nutrient retention factors, used to correct for the change in nutritional value of foods that occurs when cooking, were not applied when calculating the nutrient composition of dishes. This was because it would not have affected the validity of the atlas, and because the atlas was intended for comparisons of dietary intakes in relation to other household members or between trial arms, rather than for exact calculations of nutritional adequacy. However, if the recipes were to be used for other purposes, the recipes may need to be reanalysed, to account for these factors.

Data collectors could only weigh intakes over one mealtime, due to the severe cultural challenges that they faced when they initially attempted a full 24-hour weighed food record. It was not safe for the women to travel home in the dark after the respondents had eaten their evening meal, and they faced complaints and criticisms from their own communities and the respondents for spending nights and long periods away from home. This meant that 24-hour weighed food records were not possible and evening meal validation was limited to 25% of the sample. For this 25%, the three data collectors sampled households that were near to each other and stayed overnight together, or were collected by a guardian or MIRA staff member.

Although we measured internal validity (i.e. the ability of the tool to measure what it should measure), we were unable to assess the external validity; further assessment is needed to know if the atlas is valid in other South Asian contexts.

The validation method bears international relevance for individual-level dietary assessment, because it is one of few studies that have used measures of agreement to test a photographic atlas for this purpose, under 'real' field conditions that the atlas would be used in (e.g. in respondents' homes, containing similar coding errors, estimating self-served portions of own-made food, and collecting recalls 24 hours after consumption). However, this approach gave the disadvantage that not all respondents ate the same foods. This meant that the number of observations for each food item was small, and only three items in the atlas could be well tested. We must therefore rely on the assumption that people's ability to recall common and rare items is similar.

The characterisation of agreement using Bland-Altman plots and Cohen's weighed kappa shows the full extent of the error associated with the atlas, rather than masking errors in both directions by simply reporting the mean bias. Nutrient analyses also add to a scarce body of literature describing the nutritional implications of these errors.

Future research

Future work to re-assess the validity could test if edited images result in improved accuracy. In many studies, food atlases were tested during or immediately after serving [57], and so further research could test how bias changes with time delay between food consumption and recall.

It is hoped that this is the beginning of an effort to make the measurement of dietary intake more feasible, ~~and~~ sources of bias better understood, and that other researchers will use the atlas. The lack of recent evidence linking cultural factors (such as food taboos and gender discrimination) with inadequate diets and nutritional status indicates the need for culturally appropriate dietary assessment methods at the individual level [14]. The findings and context-appropriate images in the atlas will enable better understanding of nutritional adequacy and inequity on a large scale, particularly in Nepali and South Asian populations.

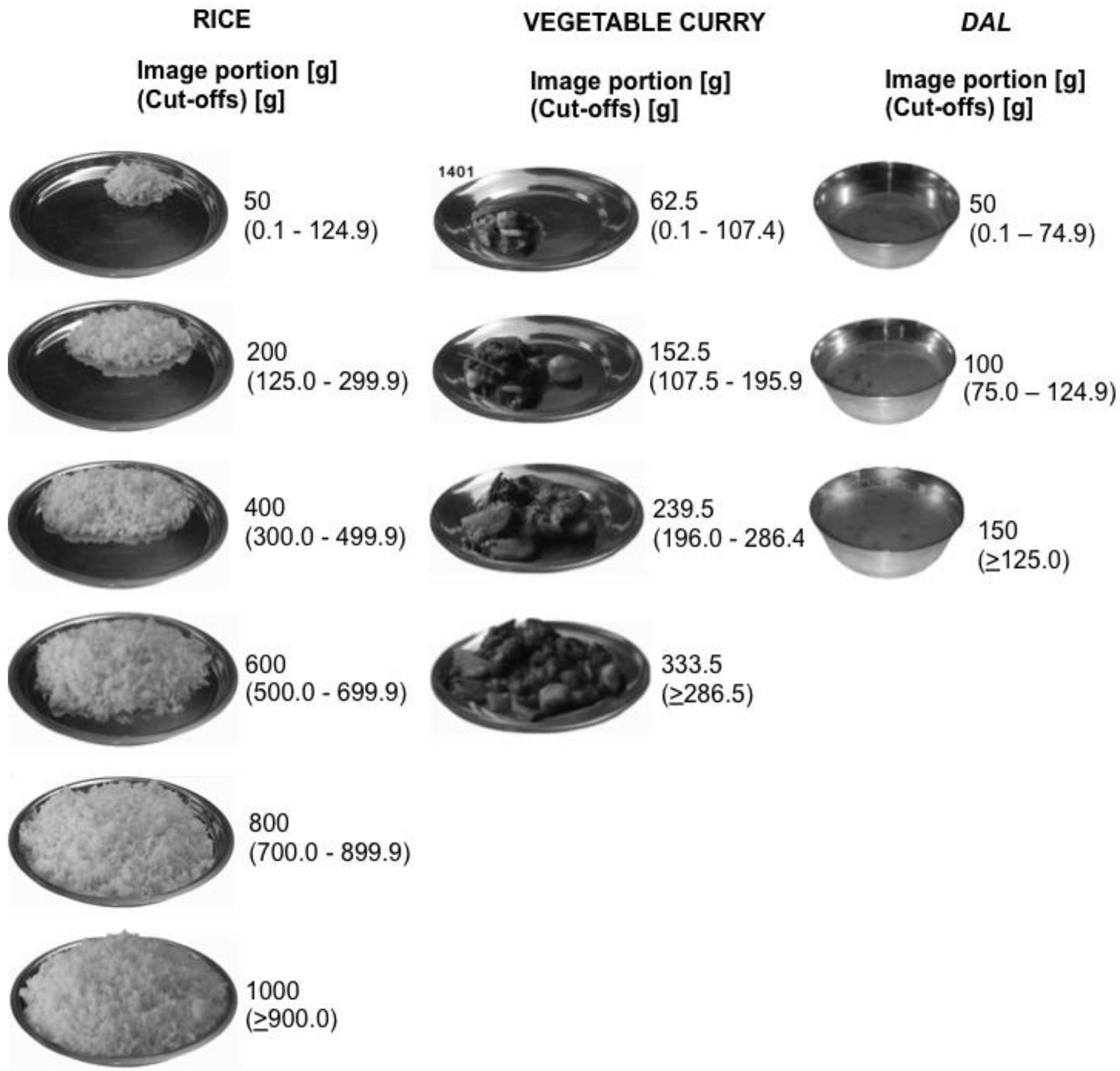

Figure 1: Examples of portion size images (not to scale), their sizes [g] and the cut-offs [g] within which a selected image would correctly represent a given portion

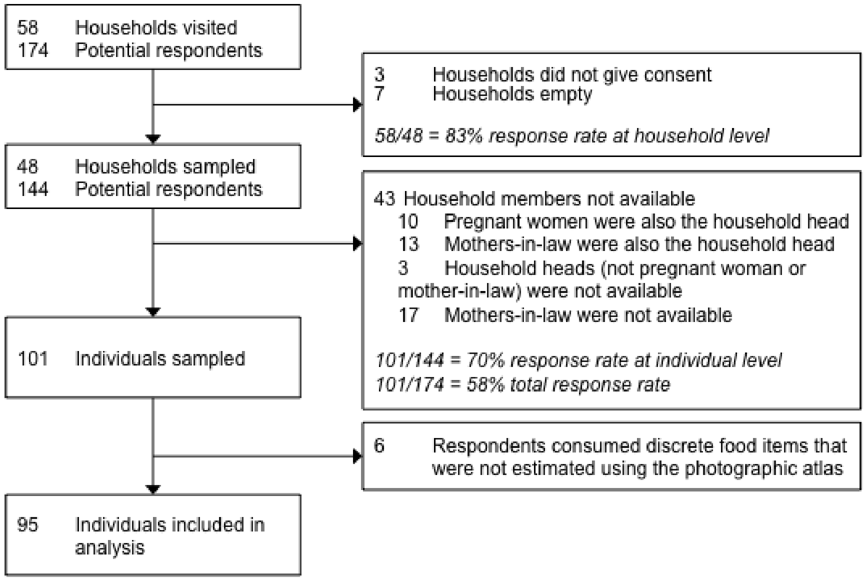

Figure 2: Response rate of households that were randomly sampled in their homes

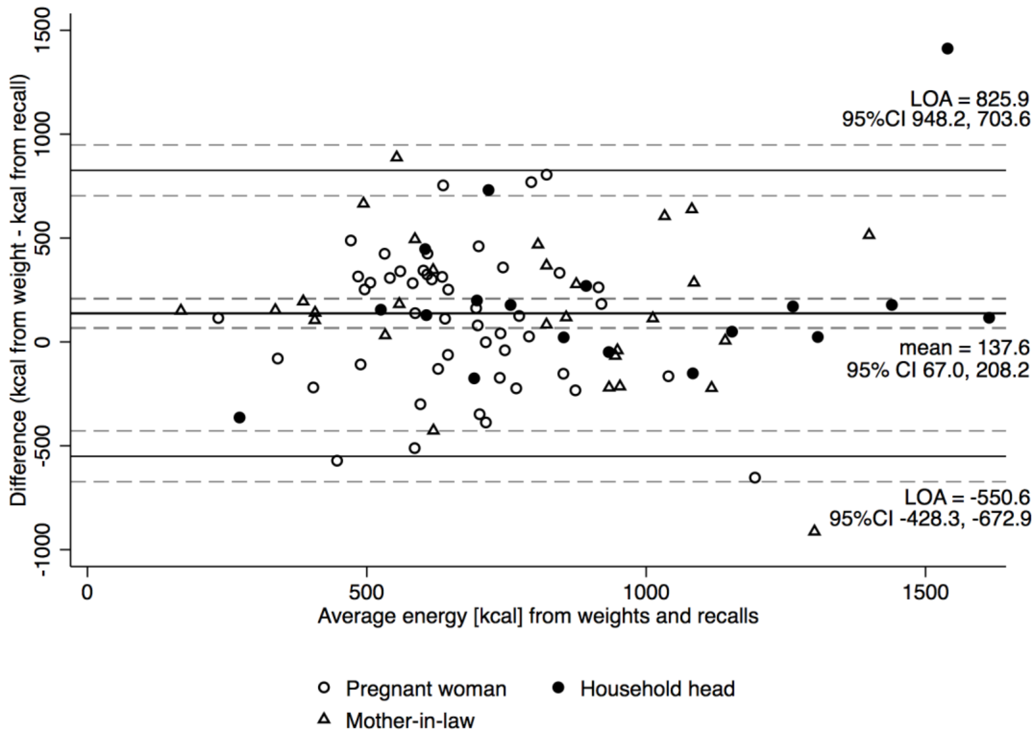

Figure 3: Bland-Altman plot of kcal intakes calculated from weighed and recalled methods

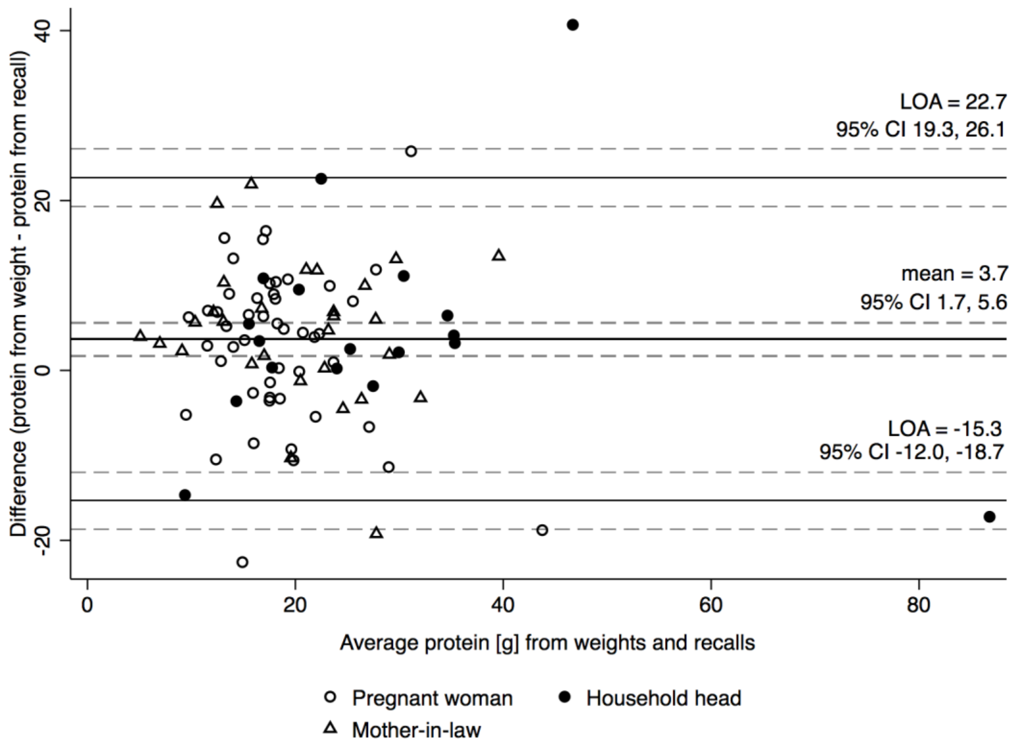

Figure 4: Bland-Altman plot of protein (g) intakes calculated from weighed and recalled methods

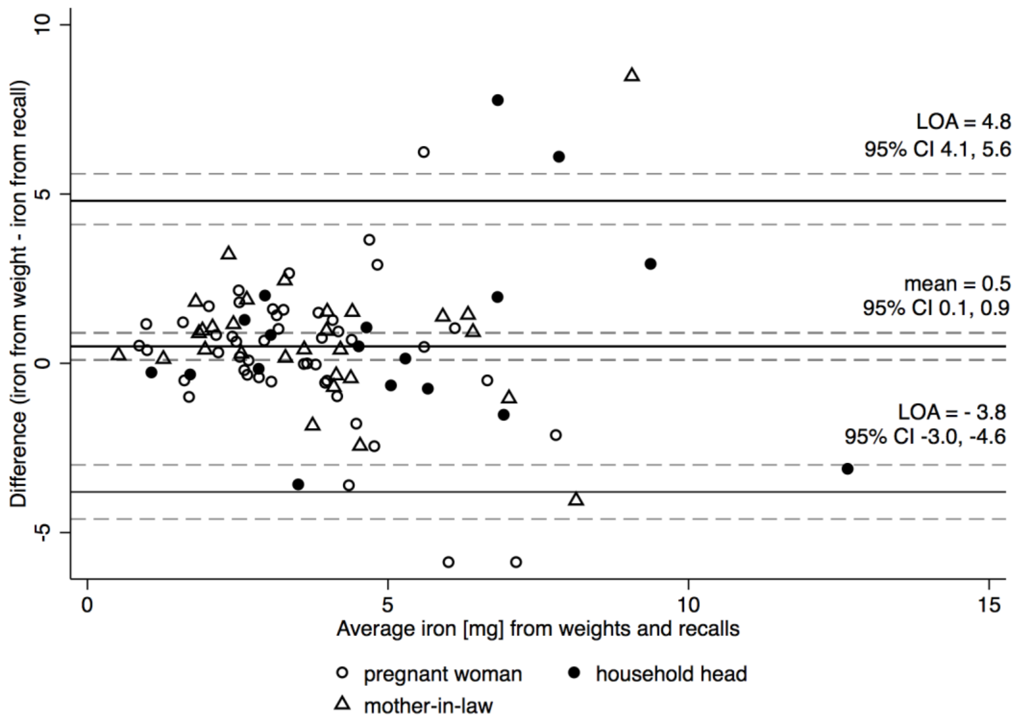

Figure 5: Bland-Altman plot of iron (mg) intakes calculated from weighed and recalled methods

Table 1 Review of studies that have reported agreement between weighed and recalled portion estimations using a photographic atlas

	Country	Sample (*n*)	Food types (*n*)	Time between consumption and recall	Bland-Altman		Cohen's kappa
					Mean difference (weight - recall estimate)	95% limits of agreement	
[34]	Burkina Faso	137 women for all food types	8	24 h	Not applicable	Not applicable	Range 0.52, 0.92
[35]	Mozambique	99 girls (aged 13 - 18 years) 25 - 52 recalls per food type	5	30 m	Range -2 to -61 g between food types	Approx. 120, -225 g for rice and stiff maize porridge. Exact values not reported; only shown on plots.	Not applicable
[36]	Bolivia	34 women 15 - 198 recalls per food type	10	24 h	Range -13 to 4 g (median) between food types 56.7 kcal (mean) for all types.	Proportional and log limits reported * 49.0, 162.6	Not reported
[37]	South Africa	92 adolescents *n* per food type not reported	11	30 s †	Range -66.0 to 29.6 g between food types. 0.6 kcal for all types	Not reported Only reported 2.7% and 0.2% fell above or below the limits.	Not reported
[38]	Lebanon	50 adults 11 - 67 recalls per food type	9 (212 items)	24 h	Range -36.8 to 17.1 g between food types	Reported that they showed 'good agreement for all dishes' *	Not reported
[39]	Italy	448 adults and children 45 - 3513 per food type	6 (434 items)	5 - 10 m	Range -1.3 to 23.2 g for different food types Overall 13.6 g	Widest limits -153.9, 107.5 * Overall limits -114.9, 87.8 *	Not reported

* Multiple measurements per individual in these analyses mean that data assumption of non-independence of scores for Bland-Altman limits of agreement does not hold.

† Respondents were only shown the portion and did not consume it.

Table 2: Respondent characteristics								
	All respondents		**Pregnant woman**		**Household head**		**Mother-in-law**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Number of respondents	95	-	48	-	18	-	29	-
Age	35.6	15.5	22.9	5.1	48.0	14.9	48.9	9.8
Gender (% female)	75.8	-	100	-	0	-	100	-
Years of schooling	2.3	3.6	2.8	3.7	2.8	4.2	1.1	2.9
Mid-upper arm circumference *	24.6	3.0	23.9	2.2	26.5	3.3	25.2	3.8

* Response rate = 83.2%; all other variables had 100% response rate.

Table 3: Bias (percentage error) between paired weighed and recalled portion sizes

Food item	***n***	**Weighed portion size**		**Recalled portion size using food atlas**		**% error from within-pair differences**		**Portion sizes shown in atlas [g]**
		Mean [g]	**SD**	**Mean [g]**	**SD**	**Mean**	**SE**	
Rice	84	498.2	198.1	408.3	216.1	-11.4	5.6	50, 200, 400, 600, 800, 1000
Vegetable curry	55	145.6	72.6	158.4	86.8	20.8	9.6	62.5, 152.5, 239.5, 333.5
Dal (spiced lentil soup)	53	218.6	75.4	135.1	50.6	-34.5	3.6	50, 100, 150
Sag (green leafy vegetables, cooked with salt and oil)	11	49.5	24.3	50.9	30.2	40.1	35.0	20, 60, 100
Bhujiya, spiced fried potato	10	94.7	96.3	60.3	11.0	-13.0	9.2	25, 60, 115
Flat breads, all types including *roti*, *paratha* and *puri*	7	222.8	88.2	160.9	96.4	-22.5	14.3	31.5, 40.5, 127.5
Other foods	25	117.6	112.4	106.3	125.6	16.0	14.8	
Total	245	275.1	213.1	223.4	194.8	-4.5	3.9	

~~* Using widest interval of portion sizes but assuming no outliers (very high or low portion sizes)~~

Table 4: Difference between the selected photograph number and the most closely matching image number, and agreement in selection of portion size images

Food type	% subjects with selection error of *n* images					Weighted kappa			
	n *	**0**	**± 1**	**± 2**	**± 3 or more**	**Observed agreement (%)**	**Agreement expected by chance (%)**	**κ_w (SE)**	***p*-value**
Rice (6 photographs)	71	33.8	50.7	11.3	4.2	94.7	91.2	0.391 (0.105)	<0.001
Vegetable curry (4 photographs)	49	46.9	40.8	10.2	2.0	89.0	80.7	0.430 (0.139)	0.001
Dal (spiced lentil soup) (3 photographs)	34	47.1	41.2	11.8	not possible	77.9	76.9	0.045 (0.063)	0.238

* Recalls were excluded if respondents used multiplication or division factors (e.g. if someone reported having two servings of a portion image) or if respondents recorded recalls that were in between two portion size images.

REFERENCES

1. Black RE, Allen LH, Bhutta ZA *et al.* (2008) Maternal and child undernutrition: global and regional exposures and health consequences. *Lancet* **371**, 243-260.

2. World Health Organization (2011) Global Status Report on Non-Communicable Diseases 2010. http://www.who.int/nmh/publications/ncd_report_full_en.pdf (accessed Nov 2015)

3. National Institute of Population Research and Training (NIPORT), Mitra and Associates, ICF International (2011) Bangladesh Demographic and Health Survey: Key Indicators. Dhaka, Bangladesh and Calverton, Maryland, USA.

4. International Institute for Population Sciences (IIPS), Macro International (2007) National Family Health Survey (NFHS-3), 2005–06: India: Volume I, . Mumbai.

5. Gibson RS (2005) *Principles of Nutritional Assessment*: Oxford University Press.

6. 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.

7. Gibson RS, Ferguson EL (1999) *An interactive 24-hour recall for assessing the adequacy of iron and zinc intakes in developing countries*: International Life Sciences Institute Washington^ eD. C DC.

8. Bingham SA (1991) Limitations of the various methods for collecting dietary intake data. *Ann Nutr Metab* **35**, 117-127.

9. Poslusna K, Ruprich J, de Vries JH *et al.* (2009) Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. *Br J Nutr* **101**, S73-S85.

10. Margetts BM, Nelson M (1997) *Design concepts in nutritional epidemiology*: Oxford University Press.

11. Thompson FE, Byers T (1994) Dietary assessment resource manual. *J Nutr* **124**, 2245s-2317s.

12. Panter-Brick C (1993) Seasonality of energy expenditure during pregnancy and lactation for rural Nepali women. *Am J Clin Nutr* **57**, 620-628.

13. Bains K, Kaur B, Mann S (2002) Measurement of energy cost of selected household and farm activities performed by rural women. *Food Nutr Bull* **23**, 274-279.

14. Gittelsohn J (1991) Opening the box: intrahousehold food allocation in rural Nepal. *Soc Sci Med* **33**, 1141-1154.

15. Parry JP (2013) *Caste and kinship in Kangra.* London, New York: Routledge.

16. Lüthi D (2010) Private Cleanliness, Public Mess: Purity, Pollution and Space in Kottar, South India. In *Urban Pollution: Cultural Meanings, Social Practices*, vol. 15, pp. 57 [E Dürr and R Jaffe, editors]. New York: Berghahn Books.

17. Cypel YS, Guenther PM, Petot GJ (1997) Validity of portion-size measurement aids: a review. *J Am Diet Assoc* **97**, 289-292.

18. Vereecken C, Dohogne S, Covents M *et al.* (2010) How accurate are adolescents in portion-size estimation using the computer tool Young Adolescents' Nutrition Assessment on Computer (YANA-C)? *Br J Nutr* **103**, 1844-1850.

19. Lanerolle P, Thoradeniya T, Silva A (2013) Food models for portion size estimation of Asian foods. *J Hum Nutr Diet* **26**, 380-386.

20. Godwin SL, Chambers IV E, Cleveland L (2004) Accuracy of reporting dietary intake using various portion-size aids in-person and via telephone. *J Am Diet Assoc* **104**, 585-594.

21. Bernal-Orozco MF, Vizmanos-Lamotte B, Rodríguez-Rocha NP *et al.* (2013) Validation of a Mexican food photograph album as a tool to visually estimate food amounts in adolescents. *Br J Nutr* **109**, 944-952.

22. Thoradeniya T, de Silva A, Arambepola C *et al.* (2012) Portion size estimation aids for Asian foods. *J Hum Nutr Diet* **25**, 497-504.

23. Kulsoom T, Mushtaq S, Hakeem R (2008) Evaluation of a Pakistani Food Photographs Atlas. *Nurture*, 30.

24. Ovaskainen M, Paturi M, Reinivuo H *et al.* (2007) Accuracy in the estimation of food servings against the portions in food photographs. *Eur J Clin Nutr* **62**, 674-681.

25. Subar AF, Crafts J, Zimmerman TP *et al.* (2010) Assessment of the accuracy of portion size reports using computer-based food photographs aids in the development of an automated self-administered 24-hour recall. *J Am Diet Assoc* **110**, 55-64.

26. Robinson F, Morritz W, McGuiness P *et al.* (1997) A study of the use of a photographic food atlas to estimate served and self ‐ served portion sizes. *J Hum Nutr Diet* **10**, 117-124.

27. De Keyzer W, Huybrechts I, De Maeyer M *et al.* (2011) Food photographs in nutritional surveillance: errors in portion size estimation using drawings of bread and photographs of margarine and beverages consumption. *Br J Nutr* **105**, 1073-1083.

28. Foster E, Matthews JN, Nelson M *et al.* (2006) Accuracy of estimates of food portion size using food photographs–the importance of using age-appropriate tools. *Public health nutr* **9**, 509-514.

29. Lillegaard I, Øverby N, Andersen L (2005) Can children and adolescents use photographs of food to estimate portion sizes? *Eur J Clin Nutr* **59**, 611-617.

30. Trolle E, Vandevijvere S, Ruprich J *et al.* (2013) Validation of a food quantification picture book targeting children of 0–10 years of age for pan-European and national dietary surveys. *Br J Nutr* **110**, 2298-2308.
31. Robson PJ, Livingstone MBE (2000) An evaluation of food photographs as a tool for quantifying food and nutrient intakes. *Public health nutr* **3**, 183-192.
32. Boateng G (2014) The Development of a Photographic Food Atlas with Portion Sizes of Commonly Consumed Carbohydrate Foods in Accra, Ghana. PhD, University of Ghana.
33. Venter C, MacIntyre U, Vorster H (2000) The development and testing of a food portion photograph book for use in an African population. *J Hum Nutr Diet* **13**, 205-218.
34. Huybregts L, Roberfroid D, Lachat C *et al.* (2008) Validity of photographs for food portion estimation in a rural West African setting. *Public health nutr* **11**, 581-587.
35. Korkalo L, Erkkola M, Fidalgo L *et al.* (2013) Food photographs in portion size estimation among adolescent Mozambican girls. *Public health nutr* **16**, 1558-1564.
36. Lazarte CE, Encinas ME, Alegre C *et al.* (2012) Validation of digital photographs, as a tool in 24-h recall, for the improvement of dietary assessment among rural populations in developing countries. *Nutr J* **11**, 61.
37. Steyn NP, Senekal M, Norris SA *et al.* (2006) How well do adolescents determine portion sizes of foods and beverages? *Asia Pac J Clin Nutr* **15**, 35.
38. Tueni M, Mounayar A, Birlouez-Aragon I (2012) Development and evaluation of a photographic atlas as a tool for dietary assessment studies in Middle East cultures. *Public health nutr* **15**, 1023-1028.
39. Turconi G, Guarcello M, Berzolari FG *et al.* (2005) An evaluation of a colour food photography atlas as a tool for quantifying food portion size in epidemiological dietary surveys. *Eur J Clin Nutr* **59**, 923-931.
40. Ministry of Health and Population (MOHP) [Nepal], New ERA, Inc. II (2011) Nepal Demographic and Health Survey 2011. Kathmandu, Nepal and Calverton, Maryland.
41. Nelson M, Haraldsdóttir J (1998) Food photographs: practical guidelines II. Development and use of photographic atlases for assessing food portion size. *Public health nutr* **1**, 231-237.
42. Akhter N (2013) Food and Nutrition Security in the Rural Plains of Nepal: Impact of the Global Food Price Crisis. Doctoral thesis. *UCL (University College London)*.
43. Sudo N, Sekiyama M, Maharjan M *et al.* (2005) Gender differences in dietary intake among adults of Hindu communities in lowland Nepal: assessment of portion sizes and food consumption frequencies. *Eur J Clin Nutr* **60**, 469-477.
44. Moshfegh AJ, Rhodes DG, Baer DJ *et al.* (2008) The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. *Am J Clin Nutr* **88**, 324-332.
45. Ferguson E, Gadowsky S, Huddle J *et al.* (1995) An interactive 24-h recall technique for assessing the adequacy of trace mineral intakes of rural Malawian women; its advantages and limitations. *Eur J Clin Nutr* **49**, 565-578.
46. Hartung C, Lerer A, Anokwa Y *et al.* (2010) Open data kit: tools to build information services for developing regions. *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*, 18.
47. Johansson G, Wikman Å, Åhrén A-M *et al.* (2001) Underreporting of energy intake in repeated 24-hour recalls related to gender, age, weight status, day of interview, educational level, reported food intake, smoking habits and area of living. *Public health nutr* **4**, 919-927.
48. Heerstrass D, Ocke M, Bueno-de-Mesquita H *et al.* (1998) Underreporting of energy, protein and potassium intake in relation to body mass index. *Int J Epidemiol* **27**, 186-193.
49. USDA (2013) USDA National Nutrient Database for Standard Reference, Release 26. *US Department of Agriculture, Agricultural Research Service Nutrient Data Laboratory Home Page*.
50. Public Health England (2015) McCance and Widdowson's The Composition of Foods Integrated Dataset 2015. *Institute of Food Research*.
51. Shaheen N, Rahim ATM, Mohiduzzaman M *et al.* (2013) Food Composition Table for Bangladesh. *Institute of Nutrition and Food Science, Centre for Advanced Research in Sciences, University of Dhaka*.
52. Nepal Nutrition Program (2012) Food Composition Table for Nepal. *Nepal Government, Ministry of Agriculture Development, Department of Food Technology and Quality Control*.
53. Sharavathy M, Urooj A, Puttaraj S (2001) Nutritionally important starch fractions in cereal based Indian food preparations. *Food Chem* **75**, 241-247.
54. Bland JM, Altman D (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **327**, 307-310.
55. Cohen J (1968) Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull* **70**, 213.
56. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* **8**, 135-160.
57. Nelson M, Haraldsdóttir J (1998) Food photographs: practical guidelines I. Design and analysis of studies to validate portion size estimates. *Public health nutr* **1**, 219-230.

58. Sim J, Wright CC (2005) The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* **85**, 257-268.