Measuring dietary diversity with high frequency mobile phone interviews

Thomas Assefa, UGA, thomas.assefa@uga.edu; Ellen McCullough, UGA, emccullough@uga.edu;
Tamara McGavock, Grinnell College, mcgavock@grinnell.edu

Selected Paper prepared for presentation at the 2022 Agricultural & Applied Economics Association Annual Meeting, Anaheim, CA; July 31-August 2

Measuring dietary diversity with high frequency mobile phone interviews

Thomas Assefa
UGA

Ellen McCullough
UGA

Tamara McGavock
Grinnell College

May 18, 2022

Abstract

We conducted an experiment to validate a novel method of collecting dietary diversity through high-frequency phone call surveys in Ethiopia. We found that Household Dietary Diversity Score (HDDS) measured based on an in-person one-time survey is significantly higher compared to one measured based on high-frequency phone call surveys. We also found that the Women's Dietary Diversity Score (WDDS) measured based on a high-frequency phone call survey is significantly higher than one measured based on an in-person one-time survey. Our results suggest that respondents behave differently in terms of over-reporting and underreporting food consumption when the recall period is longer and shorter.

1 Introduction

Dietary diversity scores are widely used to proxy for diet quality and food intake in nutrition, agriculture and health research. Dietary diversity scores are easier and less costly to collect than a standard 24-hour food consumption recall, which also includes quantities consumed (Gibson and Ferguson, 2008), and they are highly correlated with food consumption for infants and young children (Daniels et al., 2009; Kennedy et al., 2007; Moursi et al., 2008; Steyn et al., 2006), adult women (Arimond et al., 2010; Diop et al., 2020; Nguyen et al., 2018) and households (Hoddinott and Yohannes, 2002; Mekonnen et al., 2020).

One key design feature in dietary diversity modules is the reference period over which data are collected. Dietary diversity modules are usually based on the respondent's recall over a reference period spanning the previous 1-3 days, though a 7 day period is also commonly used, and periods of up to 15 days have been reported (FAO, 2013; Ruel, 2003). Dietary diversity data can be difficult to collect from respondents due to the large cognitive burden associated with recalling which food items were consumed. Data quality are threatened by recall errors such as reversion to "usual" practices, telescoping, or under/over-reporting of items consumed. While shorter recall periods generally decrease the cognitive burden that respondents face in reporting data, they are not without problems. First, respondents are more likely to erroneously include consumption episodes occurring outside the recall period (i.e., telescoping) when the recall period is short (Eisenhower et al., 2004). Moreover, shorter recall periods can result in larger within-person error, if the aim is to estimate "usual" diets, when some foods, such as animal sourced foods or fruits, are infrequently consumed (Thorne-Lyman et al., 2014). Shorter recall periods can also be problematic in contexts where food consumption is highly cyclical, e.g., due to market days, weekly religious fasting periods (such as the Wednesday and Friday fasting among Orthodox Christian Ethiopians), or seasonal patterns that can affect the timing of meals, types of foods being consumed, and the intra-household allocation of foods.

Our study contributes to a growing body of research seeking to develop improved survey tools and data collection methods to measure diet quality in the field (Ameye et al., 2020; De Weerdt et al., 2016; Friedman et al., 2017; Gicevic et al., 2020; Hanley-cook et al., 2020; Herforth et al., 2020) and mode of data collection (Caeyers et al., 2012; Lamanna et al., 2019). We complement these efforts by using mobile tools to bound recall periods and extend the total reference period over which dietary diversity indicators are constructed. We also contribute to the broader body of research seeking to develop and validate new mobile-based survey methods in order to greatly reduce the costs of collecting micro data in LMIC settings, especially when multiple observations are required over time from a respondent. Due to the decreasing operating costs, cell phone ownership is rapidly increasing across Africa (Stork et al. 2017, GSMA 2018) and mobile phone methods are becoming much more popular. We also contribute to the evidence base addressing the role of mobile tools in improving data quality by reducing respondents' cognitive burden (Ashman et al., 2017; Kong et al., 2017).

2 Methods and data

We conduct an experiment in rural Ethiopia to develop and validate a novel survey method to resolve the tradeoffs between cognitive burden and recall. Our new method seeks to extend respondents' reference period without exacerbating cognitive burden associated with a longer recall period. Using the novel method, we collect diet data through phone calls several times a day over a 7-day window, with each call corresponding to a short, bounded recall period. By comparing women's dietary diversity measures collected using this new method to dietary diversity measures collected through traditional methods (in person interview with recall over 24-hour and 7-day periods), we formally test whether the new method affects recall bias (e.g. telescoping, cognitive burden) and within-household measurement error. We disentangle the effects of call bounding on the incidence of telescoping, following Abate et al. (2020), who found that telescoping error was greatly reduced by visiting households at the start of their 7-day recall period (thus aiding respondents in bounding the reference period). We also measure the effect of extending the reference period on reducing within-individual 24-hour dietary diversity measurement error given the observed weekly and seasonal food intake patterns.

Our study population includes poor women in northern Ethiopia who are participating in multi-faceted graduation-from-poverty programs. Our respondent sample includes over 600 poor households from 60 randomly sampled livelihoods groups. We randomly assign each household to one of two survey methods. The first method ("phone") uses multiple phone interviews, each covering a bounded period, to construct individual and household dietary diversity measures. Phone respondents report food intake twice per day for a 7-day period, once in the morning, and once in the afternoon, with each recall period clearly bounded between the previous call and current call. The second method ("in person") collects a 24-hour women's dietary diversity measure (as well as a 7-day household level dietary diversity measure) through one, traditional in person interview. Respondents in the "in person" group are randomly assigned to one of 7 days of the week for their face-to-face interview, which allows us to credibly assess day-of-week survey effects on diet quality measures. Because phone ownership is low in our study population, we issue each respondent, regardless of data collection method, a free cell phone.

We analyze the effect of survey method on two key outcome variables:

1. Women's 24-hour dietary diversity score ($y_{ivv} = WDDS_{iv}$)

2. Household level 7-day dietary diversity score ($y_{iv} = HDDS_{iv}$),

Differences in these dietary outcome variables are compared across the treatment arms using the following specification:

$$y_{iv} = \beta T_{iv} + X'_{iv}\gamma + \epsilon_{iv} \tag{1}$$

where T_{iv} is the individual's treatment assignment (receiving high-frequency cell phone survey takes the value of 1), X'_{iv} is a vector of controls, ϵ_{iv} is a normally distributed error term, iv stands for individual i in village or VESA v .

3 Results

We start with checking if randomization was successful in creating comparable groups of households. Table 1 presents the means of key covariates disaggregated by treatment status on columns 1 and 2. Column 3 shows the differences between the means of the covariates of treated and control group households. None of the differences were found to be statistically significant based on a t-test. This suggests, that except for the difference in the treatment status, treated and control groups of households are similar in terms of the covariates listed. These covariates are used as controls in the regressions later.

Table 1: Comparison of the treatment and control groups based on covariates

	(1)		(2)		(1-2)[#]
	Treated [N=208]		Control [N=434]		T - C
Variables	Mean	[Sd]	Mean	[Sd]	Difference
Male-headed household	0.837	[0.371]	0.834	[0.372]	0.002
Number of children aged under 24 month age	0.255	[0.437]	0.276	[0.458]	-0.022
Number of children aged between 2 - 5 years	0.587	[0.608]	0.553	[0.633]	0.034
Number children aged between 5 - 15 years	1.601	[1.021]	1.691	[1.034]	-0.090
Number of children aged between 15 - 18 years	0.317	[0.525]	0.385	[0.524]	-0.067
Number of Adults aged between 18 - 49 years	1.630	[0.769]	1.645	[0.709]	-0.015
Number of Adults aged above 50 years	0.505	[0.688]	0.541	[0.732]	-0.037
Dependency ratio	104.593	[63.367]	107.347	[67.305]	-2.753
Average number of days in PSNP participation	36.974	[28.617]	40.122	[38.712]	-3.148
Total land owned by the household in hectare	0.711	[0.584]	0.710	[0.507]	0.002
Total Value of household asset (Non-Land) in USD	29.932	[26.782]	32.805	[64.512]	-2.873
Total value of livestock owned at present in USD	387.525	[296.497]	390.627	[279.628]	-3.102

[#] We conducted a t-test on the differences of the means of covariates, and *s represents significance levels with * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Figure 1 shows the kernel density distribution of Household Dietary Diversity Score (HDDS) for treated and control households separately. The distributions are similar with slightly smaller values from phone surveys. The three large modes of the distribution of HDDS for treated households are 4, 5, and 6, while it is 5,6, and 7 for control households. Figure 2 presents the kernel density distribution of the Women's Dietary Diversity Score (WDDS). The distributions are similar with slightly higher values from the phone surveys. Figure 3 shows similar distributions of WDDS for treated and control households when phone survey WDDS is computed by taking average WDDS from all of the phone survey days instead of just the first day.

Figure 1: Distribution of HDDS by Treatment Status

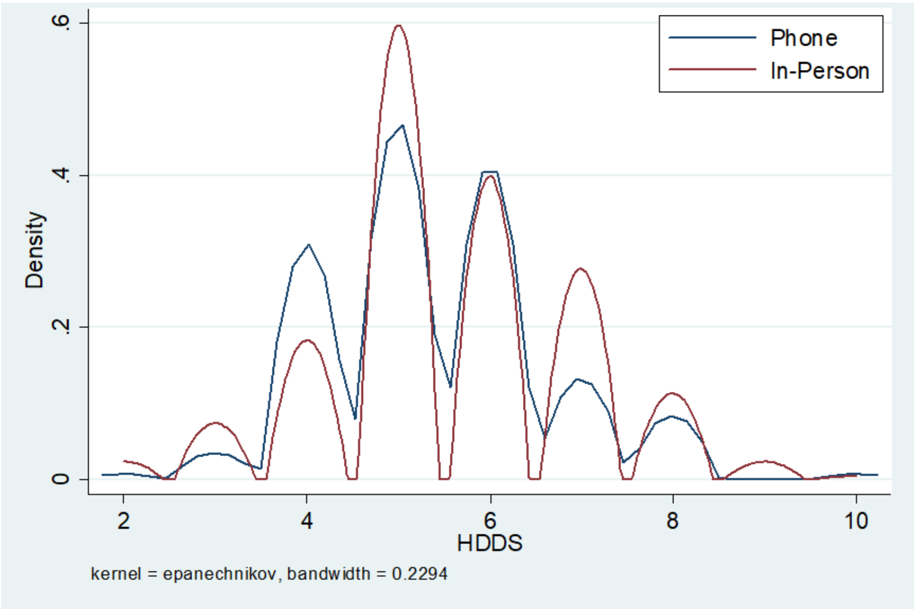


Figure 2: Distribution of WDDS by Treatment Status

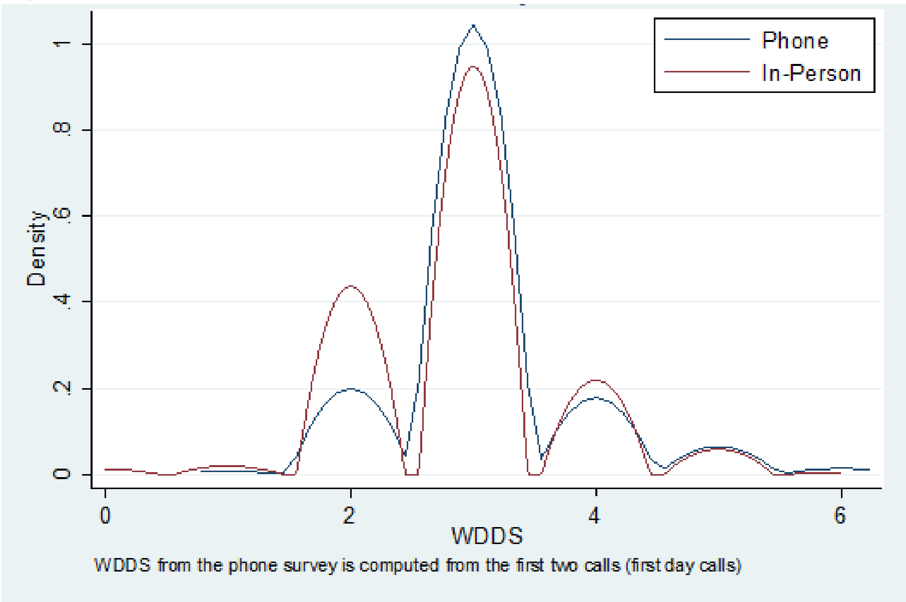
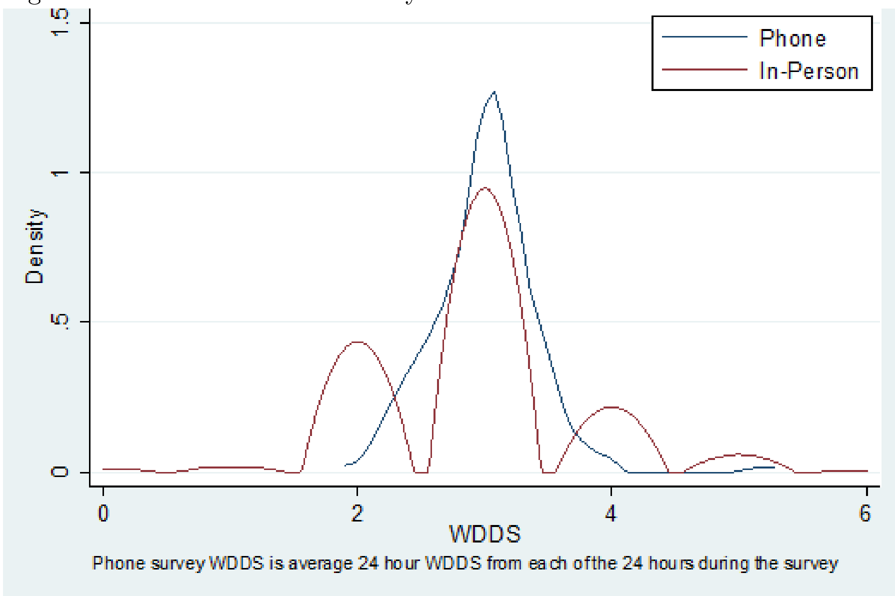


Figure 3: Distribution of WDDS by Treatment Status



We estimated equation 1 using OLS. But the outcome variables are counts of the different food groups households or individuals consumed. The distributions in Figures 1, 2, and 3 also show accumulated values at integer values, suggesting OLS may not be appropriate and a Poisson regression might be more appropriate. Therefore, we also estimate equation 1 using a Poisson regression model.

Table 2 presents the results for HDDS. Columns 1, 2, and 3 show the estimation results without any control, with just VESA fixed effects and with even additional controls listed in Table 1, respectively. The coefficient of interest is negative consistently across the different models. It becomes statistically significant, at a 10 percent level of significance, when we control for VESA fixed effects and additional covariates. This suggests that HDDS measured using high-frequency phone surveys is significantly lower than HDDS measured using in-person one-time surveys. This happens when respondents report consumption of food groups by mistake when asked in person due to recall error while it is less likely to do so on high-frequency phone surveys as recall error is dramatically reduced. The reference period on in-person one-time surveys is seven days and it is likely that respondents forgot what type of food they consumed and may report food groups that were not consumed. This misdating of consumption episodes, commonly known as telescoping, has been documented by others. Our results are in line with what Abate et al., (2020) found that reported food consumption is 16 percent higher in the unbounded single visit recall relative to the two-visit bounded recall.

Table 2: Treatment effects on HDDS

	OLS			Poisson regression		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment [1 if HFPS]	-.203	-.214*	-.216*	-.0369	-.039*	-.0392*
	(.126)	(.127)	(.123)	(.0229)	(.0221)	(.0212)
VESA fixed effects	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
_cons	5.59***	5.18***	4.78***	1.72***	1.64***	1.57***
	(.137)	(.0422)	(.27)	(.0244)	(.00719)	(.0473)
Control group mean	*5.592 (0.065)*					
	642	642	639	642	642	639

Standard errors in parentheses and are clustered by VESA; * p<0.10, ** p<0.05, *** p<0.01

Table 3 presents the results for the estimation of equation 1 when we are using WDDS as an outcome variable. We present estimation results based on OLS and Poisson regressions. The results are consistent across the different models. We found positive and significant coefficients suggesting that the number of food groups reported as consumed by the respondent woman are larger on high-frequency phone surveys compared to one-time in-person surveys. This happens when respondents forget to include a food group consumed during the reference period due to a recall error. High-frequency phone surveys help respondents to bound and shorten the reference period so that a relatively accurate value of food consumption is reported. The results suggest that high-frequency phone surveys helped to capture the consumption of food groups that would be unreported through the in-person one-time survey.

The difference in treatment effect on HDDS and WDDS warrants an explanation. We found a negative treatment effect on HDDS while the treatment effect on WDDS is positive. One possible explanation is the difference in the reference period used. HDDS uses seven days while WDDS uses a 24-hours reference period. With a longer reference period like the seven days, respondents may tend to think most types of food have been consumed when they don't remember what was consumed while they tend to think the reverse when the reference period is shorter. The logic behind such a tendency is - longer periods allow consumption of different types of foods while it is likely only a few types of food are consumed within 24 hours. Respondents with recall errors may follow the above logic in providing their responses which will result in a negative coefficient for HDDS and a positive coefficient for WDDS.

Table 3: Treatment effects on WDDS: based on the first two calls (firs day)

	OLS			Poisson regression		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment [1 if HFPS]	.187**	.174**	.173**	.0623**	.0581**	.0588**
	(.0796)	(.0813)	(.0805)	(.0265)	(.0261)	(.0252)
VESA fixed effects	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
_cons	2.9***	3.39***	3.53***	1.07***	1.22***	1.26***
	(.0582)	(.0271)	(.148)	(.02)	(.00905)	(.0468)
Control group mean	*2.903 (0.039)*					
N	646	646	643	646	646	643

Standard errors in parentheses and are clustered by VESA; * p<0.10, ** p<0.05, *** p<0.01

In Table 3, phone survey WDDS is computed using the first two phone surveys which collects information about food consumption in the first 24 hours of the phone surveys. We conduct phone surveys for more than 24 hours, extending the reference period up to seven days. The choice of the first day (the first two phone calls) is arbitrary. One can use any of the 24 hours from the seven days. Alternatively, we can construct an average WDDS from all of these 24 hours within the seven days. Table 4 presents an estimation of equation 1 when phone survey WDDS is measured by taking the average WDDS from all of these 24 hours within the seven days. The results confirm our findings in Table 3 that high-frequency phone surveys allowed capturing food groups that would be unreported through in-person one-time surveys.

Table 4: Treatment effects on WDDS: based on average 24 hours WDDS from all phone survey days

	OLS			Poisson regression		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment [1 if HFPS]	.0931*	.0792*	.0838*	.0316**	.0268*	.0291**
	(.0464)	(.0455)	(.0431)	(.016)	(.015)	(.0139)
VESA fixed effects	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
_cons	2.9***	3.31***	3.38***	1.07***	1.19***	1.22***
	(.0582)	(.0152)	(.123)	(.02)	(.00511)	(.0395)
Control group mean	*2.903 (0.039)*					
N	646	646	643	646	646	643

Standard errors in parentheses and are clustered by VESA; * p<0.10, ** p<0.05, *** p<0.01

4 Conclusion

We conducted an experiment to validate a novel cell phone-based method of collecting dietary diversity in low-income country settings. We intended to increase the reference period of the data while decreasing the respondent's cognitive burden. We do so by using high-frequency phone survey interviews, twice a day for seven days, that help respondents to bound and shorten the reference period.

Our results show that HDDS measured using one-time in-person surveys are significantly higher than what would be found through high-frequency phone surveys, suggesting possible telescoping effects by respondents. We also found that WDDS measured using high-frequency phone surveys is significantly higher than what would be found through one-time in-person surveys.

The differences in treatment effect on HDDS which uses seven days reference period and on WDDS which uses a 24 hours reference period suggest that respondents cope with recall error differently for longer and shorter reference periods. Faced with uncertainties, respondents may tend to assume most types of foods must have been consumed in a longer reference period while they tend to assume fewer types of foods must have been consumed in a shorter reference period.

References

Abate, G. T., de Brauw, A., Gibson, J., Hirvonen, K., & Wolle, A. (2020). *Telescoping causes overstatement in recalled food consumption: Evidence from a survey experiment in Ethiopia*. *November*.

Ameye, H., De Weerdt, J., & Gibson, J. (2020). Measuring macro- and micronutrient intake in multi-purpose surveys : Evidence from a survey experiment in Tanzania. In *LICOS Discussion Paper Series* (Vol. 421, Issue 0).

Arimond, M., Wiesmann, D., Becquey, E., Carriquiry, A., Daniels, M. C., Deitchler, M., Fanou-Fogny, N., Joseph, M. L., Kennedy, G., Martin-Prevel, Y., & Torheim, L. E. (2010). Simple food group diversity indicators predict micronutrient adequacy of women's diets in 5 diverse, resource-poor settings. *Journal of Nutrition*, *140*(11), 2059–2069. https://doi.org/10.3945/jn.110.123414

Ashman, A. M., Collins, C. E., Brown, L. J., Rae, K. M., & Rollo, M. E. (2017). Validation of a smartphone image-based dietary assessment method for pregnant women. *Nutrients*, *9*(1), 1–17. https://doi.org/10.3390/nu9010073

Caeyers, B., Chalmers, N., & De Weerdt, J. (2012). Improving consumption measurement and other survey data through CAPI: Evidence from a randomized experiment. *Journal of Development Economics*, *98*(1), 19–33. https://doi.org/10.1016/j.jdeveco.2011.12.001

Daniels, M. C., Adair, L. S., Popkin, B. M., & Truong, Y. K. (2009). Dietary diversity scores can be improved through the use of portion requirements: An analysis in young Filipino children. *European Journal of Clinical Nutrition*, *63*(2), 199–208. https://doi.org/10.1038/sj.ejcn.1602927

De Weerdt, J., Beegle, K., Friedman, J., & Gibson, J. (2016). The challenge of measuring hunger through survey. *Economic Development and Cultural Change*, *64*(4), 727–758. https://doi.org/10.1086/686669

Diop, L., Becquey, E., Turowska, Z., Huybregts, L., Ruel, M. T., & Gelli, A. (2020). Standard Minimum Dietary Diversity Indicators for Women or Infants and Young Children Are Good Predictors of Adequate Micronutrient Intakes in 24–59-Month-Old Children and Their Nonpregnant Nonbreastfeeding Mothers in Rural Burkina Faso. *The Journal of Nutrition*, 1–11. https://doi.org/10.1093/jn/nxaa360

Eisenhower, D., Mathiowetz, N. A., & Morganstein, D. (2004). RECALL ERROR: SOURCES AND BIAS REDUCTION TECHNIQUES. In *Measurement Errors in Surveys*. John Wiley & Sons, Inc.

FAO. (2013). *Guidelines for measuring household and individual dietary diversity*. https://doi.org/613.2KEN

Friedman, J., Beegle, K., De Weerdt, J., & Gibson, J. (2017). Decomposing response error in food consumption measurement: Implications for survey design from a randomized survey experiment in Tanzania. *Food Policy*, *72*(October), 94–111. https://doi.org/10.1016/j.foodpol.2017.08.016

Gibson, R. S., & Ferguson, E. L. (2008). An interactive 24-hour recall for assessing the adequacy of iron and zinc intakes in developing countries. In *HarvestPlus Technical Monograph 8*. ILSI Press.

Gicevic, S., Mou, Y., Bromage, S., Fung, T., & Walter, W. (2020). Development and Evaluation of a Novel Diet Quality Screener for Global Use. *Current Developments in Nutrition*, *4*(Supplement_2), 1168–1168. https://doi.org/10.1093/cdn/nzaa056_015

Hanley-cook, G. T., Tung, J. Y. A., Sattamini, I. F., Marinda, P. A., Thong, K., Zerfu, D., Kolsteren, P. W., Tuazon, M. A. G., & Lachat, C. K. (2020). 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. *Nutrients*, *12*(7), 1–13. https://doi.org/10.3390/nu12072039

Herforth, A., Beal, T., & Rzepa, A. (2020). *Global Diet Quality Project*. GALLUP.

Hoddinott, J., & Yohannes, Y. (2002). Dietary Diversity as a Food Security Indicator; Food Consumption and Nutrition Division Discussion Paper. *International Food Policy Research Institute*, *24*(2–3), 163–180.

Kennedy, G. L., Pedro, M. R., Seghieri, C., Nantel, G., & Brouwer, I. (2007). Dietary diversity score is a useful indicator of micronutrient intake in non-breast-feeding Filipino children. *Journal of Nutrition*, *137*(2), 472–477. https://doi.org/10.1093/jn/137.2.472

Kong, K., Zhang, L., Huang, L., & Tao, Y. (2017). Validity and practicability of smartphone-based photographic food records for estimating energy and nutrient intake. *Asia Pacific Journal of Clinical Nutrition*, *26*(3), 396–401. https://doi.org/10.6133/apjcn.042016.05

Lamanna, C., Hachhethu, K., Chesterman, S., Singhal, G., Mwongela, B., Ngendo, M., Passeri, S., Farhikhtah, A., Kadiyala, S., Bauer, J. M., & Rosenstock, T. S. (2019). Strengths and limitations of computer assisted telephone interviews (CATI) for nutrition data collection in rural Kenya. *PLoS ONE*, *14*(1), 1–20. https://doi.org/10.1371/journal.pone.0210050

Mekonnen, D. A., Talsma, E. F., Trijsburg, L., Linderhof, V., Achterbosch, T., Nijhuis, A., Ruben, R., & Brouwer, I. D. (2020). Can household dietary diversity inform about nutrient adequacy? Lessons from a food systems analysis in Ethiopia. *Food Security*, *12*(6), 1367–1383. https://doi.org/10.1007/s12571-020-01056-5

Moursi, M. M., Arimond, M., Dewey, K. G., Trèche, S., Ruel, M. T., & Delpeuch, F. (2008). Dietary diversity is a good predictor of the micronutrient density of the diet of 6- to 23-month-old children in Madagascar. *Journal of Nutrition*, *138*(12), 2448–2453. https://doi.org/10.3945/jn.108.093971

Nguyen, P. H., Huybregts, L., Sanghvi, T. G., Tran, L. M., Frongillo, E. A., Menon, P., & Ruel, M. T. (2018). Dietary diversity predicts the adequacy of micronutrient intake in pregnant adolescent girls and women in Bangladesh, but use of the 5-group Cutoff Poorly identifies individuals with inadequate intake. *Journal of Nutrition*, *148*(5), 790–797. https://doi.org/10.1093/jn/nxy045

Ruel, M. T. (2003). Operationalizing Dietary Diversity: A Review of Measurement Issues and Research Priorities. *Journal Nutrition*, *133*, 3911s-26s.

Steyn, N., Nel, J., Nantel, G., Kennedy, G., & Labadarios, D. (2006). Food variety and dietary diversity scores in children: are they good indicators of dietary adequacy? *Public Health Nutrition*, *9*(5), 644–650. https://doi.org/10.1079/phn2005912

Stork, C., Esselaar, S., & Chair, C. (2017). OTT - Threat or opportunity for African Telcos? *Telecommunications Policy*, *41*(7–8), 600–616. https://doi.org/10.1016/j.telpol.2017.05.007

Thorne-Lyman, A., Spiegelman, D., & Fawzi, W. W. (2014). Is the strength of association between indicators of dietary quality and the nutritional status of children being underestimated? *Maternal and Child Nutrition*, *10*(1), 159–160. https://doi.org/10.1111/mcn.12104