# Assessing a proxy for the variability in usual dietary energy intake with household-level data

Food Security and Nutrition Statistics, Statistics Division, Food and Agriculture Organization of the United Nations. Rome, Italy[1].

[1] Authors: Cristina Álvarez-Sanchez, Ana Moltedo, Nathalie Troubat and Carlo Cafiero

## 1. Introduction

To assess the prevalence of dietary energy or nutrient inadequacy in a population it is essential to have information on the distribution of usual intake — defined as the long-run average daily intake of a dietary component by an individual — rather than the distribution over a small number of days. This is because nutrient recommendations are intended to be met over time, not necessarily daily, and diet–health relationships are often the product of long-term exposures. However, observed consumption based on self-report quantitative dietary survey data are affected by several types of measurement errors, especially random error, which, if left unaddressed, would bias estimations of the proportion of a population at risk of insufficiency (Carriquiry, 2017).

The most common type of individual quantitative dietary surveys conducted in several countries is the 24-hour recall (24HR). Various analytic methods have been developed to statistically treat individual-level food consumption data collected with 24-hour recalls (24HR), removing day-to-day variation in intake (because what we eat and drink tends to change from day to day) and other random measurement error, to describe usual intake distributions (Laureano *et al.*, 2016). The United States National Cancer Institute Method (NCI-Method) is one of the latest and most sophisticated ones (Tooze *et al.*, 2010). Due to the paucity of individual-level data across countries and time, food consumption data from Household Consumption and Expenditure Surveys (HCES) are increasingly used (Russell *et al.*, 2018). However, surprisingly, many analysists and researchers use short-term household-level data to assess dietary energy or nutrient inadequacy without treating the data statistically to remove excess variation due to measurement error.

The Statistics Division at FAO has for long employed food consumption data from national HCES to derive a proxy for the variability in usual dietary energy intake (DEI) (e.g., Borlizzi *et al.*, 2017), which is one of the parameters[2] of the Prevalence of Undernourishment (PoU) indicator[3] (Cafiero,

---

[2] The other parameter of the PoU at national level is the mean, which is based on a country's mean per capita dietary energy intake. The use of HCES and Food Balance Sheet data to estimate a proxy for the mean of the distribution of usual intake is due to the lack of nationally representative individual-level food consumption surveys across countries and time.

[3] The PoU was adopted for tracking the Millennium Development Goals and it is one of the two food security indicators for monitoring progress on the Sustainable Development Goals Target 2.1.

2014; FAO *et al.*, 2021). If the effect of systematic errors is across the sample, they may not affect the estimate of the variability in dietary energy intake; however, when the survey is not well-designed to capture all sources of food, such as not collecting information on school meals, there might be an effect on the estimate giving the expected higher impact on poor households. The statistical approach developed at the FAO Statistics Division (the FAO-CV-Approach), which removes excess variation due to random measurement error has been evolving based on new methods and available survey data.

The purpose of this paper is two-fold: (1) to describe the latest version of the FAO-CV-Approach for estimating the variability in the distribution of usual DEI at national level using household-level data; and (2) to compare the variability estimate based on the FAO-CV-Approach to a variability estimate from individual-level data, from the same survey, treated with the NCI-Method.

It is worth noting that being the household-level data so different from individual-level data, the statistical techniques applied to each type of data for estimating a proxy of the variability in the distribution of usual dietary energy intake are completely different.

The data set used is the 2015 Bangladesh Integrated Household Survey (BIHS). It includes a household-level 7-day recall (7DR) food consumption module (we refer to this as household-level data), which is a typical food consumption module included in HCES, and two household-level food consumption rounds of a 24HR survey with information on intra-household food distribution (we refer to this as individualized or individual-level data), applied to the same households.

### 1.1. Important differences in dietary energy assessment using household-level and individual-level data

It is important to highlight some aspects about food consumption modules in HCES in comparison to individual-level quantitative dietary surveys. HCES is a term given to a family of surveys (such as Household Budget Surveys, Household Income Expenditure Surveys, and Living Standard Measurement Surveys) developed to inform economic policy. HCES typically collect information on food acquired and/or consumed (referred to as apparently consumed) at the household level to measure one component of households' expenditure to be used in poverty assessment exercises. In food consumption modules included in HCES, one respondent reports for the whole household.

The reference period is typically the last 7 or 14 days (in some cases 30 days). Food that is wasted or given away is not accounted for, typically resulting in an overestimation of consumption (Fiedler *et al.*, 2012). This contrasts with individual-level quantitative dietary surveys, such as 24HR, which are purposefully designed to collect information on individuals' food consumption. Individual-level 24HR surveys are self-reported and provide a very detailed recollection of all foods and beverages consumed during the prior 24-hours, cooking method, location of consumption, etc.

Another important aspect is the assessment of food prepared and consumed away from home (FCAH), because whereas individual-level surveys are designed to appropriately capture it, HCES are rarely designed to capture its quantity dimension. HCES typically capture the monetary value dimension on a handful of sentinel foods or dishes prepared and consumed away from home (e.g., lunch, beverage, hamburger), but not the quantity. Therefore, the dietary energy contributed by FCAH is estimated based on their monetary value and not on the amount of food. This can result in either an over- or underestimate of consumption (Moltedo *et al.*, 2018) and may be an important reason for divergences in average DEI between individual- and household-level surveys.

For several reasons, including the ones listed above, individual-level quantitative dietary surveys are considered superior tools for the assessment of food consumption. However, in many studies 24HR have been found to systematically underestimate dietary energy intake (Ravelli and Schoeller, 2020). On the other hand, the 24HR included in the 2015 BIHS is not a true individual-level 24HR; in this module, the female household member in charge of food preparation responded by proxy for the consumption of each of the other household members. So, it is bound to inaccuracies, especially in capturing food prepared and consumed away from home. Nevertheless, although this was not a "true" individual-level survey, we used it for this exercise because the BIHS is one of the very few surveys in the world including a typical household consumption module and a second module that collects data that can be individualized with a relative high degree of accuracy.

On a different note, it is worth highlighting that there is scant literature on food consumption measurement error treatment, from HCES data, for producing food and energy estimates. Thus, in this paper we heavily rely on literature from the area of dietary assessment using individual-level data. Furthermore, for simplicity, herein, we use the term “consumption” to refer to both the household and individual-level dietary data. Our use of the term “consumption” differs from that of economists in welfare analyses, who refer to “consumption” as food and non-food expenditures (World Bank, 2021).

### 1.2. The FAO methodology to estimate the Prevalence of Undernourishment indicator

The PoU estimated by FAO and used to monitor global hunger is a measure of chronic food insecurity; therefore, it refers to usual and not short-term food consumption. FAO defines the PoU as the probability that a randomly selected individual from a population has a usual food consumption inadequate to satisfy his/her dietary energy requirements consistent with long-term and good health (Cafiero, 2014). Empirically, the FAO methodology is based on modelling the population’s distribution of usual DEI with two parameters, the mean and the coefficient of variation (CV), under the assumption of lognormality. Given that the methodology is based on a single distribution for the whole population, the distribution corresponds to the daily per capita usual dietary energy intake of an average person representative of the population. The PoU is obtained as the cumulative probability that the usual DEI (x) is below a threshold (the minimum dietary energy requirement [MDER]) as shown in Equation 1.

$$PoU = \int_{x<MDER} f(x|\mu; CV)dx \qquad \text{[Equation 1]}$$

where MDER is the lowest limit of the range of energy requirements (based on a range of acceptable body weights and different physical activity levels compatible with long-term good health) for the population; $\mu$ and CV are, respectively, the mean and the coefficient of variation (which reflects variability in access to food) that define the log-normal distribution of usual DEI.

## 2. Sources of variation in usual and short-term dietary energy intake

### 2.1. Intrinsic variation in the distribution of usual dietary energy intake

Without constraints to access food, in the long run, an individual’s dietary energy intake is highly correlated with him/her energy requirements. The reason being that the physiological control system that regulates food intake, referred to as the energy homeostasis system, responds primarily to the available bodily energy reserves (Morton *et al.*, 2014).

In normal-weight individuals, energy intake matches energy expenditure over long periods of time. Energy expenditure comprises the energy spent in the maintenance of the basal metabolism (basic metabolic functions, determined mainly by the individual's age, gender, body size [weight] and body composition) and a level of necessary and desirable physical activity consistent with long-term good health (FAO, 2004). Additional energy is required by children for growth and development, and by pregnant and lactating women for the deposition of foetal tissue and the secretion of milk (FAO, 2004). FAO, WHO, and the United Nations University determined that, for any given sex-age group, there is a range of body weights that is consistent with good health, and similarly there is a range of physical activity levels that may be considered economically necessary and socially desirable (FAO, 2004). It then follows that for any given sex-age group there is a range of plausible energy requirements compatible with individuals' long-term good health. Energy homeostasis may be dysregulated during certain diseases, in obesity (Elia, 2005; Morton *et al.*, 2014), and in response to diet and low levels of physical activity (Shook *et al.*, 2015; Romieu *et al.*, 2017). However, as the influence of these on energy homeostasis at a population level is not known they are typically not considered in population assessments.

In a hypothetical healthy population in a steady state of equilibrium and without constraints to access food, variation in usual DEI would solely reflect differences in energy requirements. In such a population, we may expect all individuals to be perfectly nourished, including those on the left tail of the DEI distribution, because their lower DEI would simply reflect the fact that they have lower energy requirements. However, people do face physical and socio-economic constraints in accessing food — there are also certain personal preferences and cultural practices that influence what and how much people eat, but these may not result in a suboptimal DEI in the long-term. Thus, in virtually all populations, but particularly in resource-constrained contexts, variation in DEI not only reflects differences in individuals' energy requirements but also an inequality dimension in access to food. Summarizing, given the high correlation between energy intake and requirements, when referring to the variation in the distribution of usual DEI we should use the term variability and not inequality, as it is the case for the distribution of usual nutrient intake given that nutrient intake is not correlated to nutrient requirements.

## 2.2. Additional variation in the distribution of short-term dietary energy intake from survey data

Most instruments to assess food consumption, either at the individual- or household level, capture only short-term consumption, given the short reference periods used for data collection. Also, measurement in surveys is never perfect – at every phase, there is potential for introducing error. Measurement error generated during the design of the questionnaire (e.g., poorly worded questions), data collection (e.g., inadequately trained field staff, incorrect answers given by respondents, fatigue of both enumerators and respondents), and data processing and analysis (e.g., errors made during data entry), can generate variance inflation in the data (United Nations, 2005). Most of these measurement errors can be minimized prior to data collection (e.g., using a well-designed survey, counting with well-trained enumerators, using computer assistance personal interview systems) and with reliable data processing (e.g., identifying, and imputing outliers). The sources of additional, undesirable, variation in short-term data from individual- and household-level surveys are described more in detail below.

### 2.1.1. Individual-level quantitative dietary data

Assessing usual intake of individuals is challenging because it would require compiling information on intake during many days spread throughout the year (Willet, 2013). Instead, individual-level quantitative dietary surveys only compile intake information during one or two/three days with short-term instruments (e.g., 24HR), and then the resulting short-term dietary data are treated statistically to derive distributions of usual intake.

The variation from the short-term distribution is composed of within-person variation (that considers jointly day-to-day variation in consumption[4] and variation due to random measurement error) and between-person variation (due to differences in intake between individuals). The variability in the distribution of usual dietary energy intake (due to differences in energy requirements and to inequality in access to food) corresponds to the between-person variation.

[4] The day-to-day variation is unique to individual level data collected with 24HR surveys. Food consumption data from HCES do not permit ascertaining and removing individual-level day-to-day variation due to the nature of the data.

Thus, to estimate the variability in the distribution of usual intake, the day-to-day variation and variation due to random errors should be eliminated applying modelling statistical techniques (Dodd *et al.*, 2006). These techniques could be applied when a 24HR survey includes at least two rounds. For analytical purposes, it is assumed that a 24HR survey is an unbiased estimator of the mean intake over many administration days, that is, systematic error is assumed to be zero, although it is known that energy intake measured with this tool tends to be underreported.

**Seasonal** differences in food availability can have a strong effect on daily consumption of certain foods and micronutrients. DEI may also fluctuate across different seasons of a year (e.g., between dry and raining seasons or during religious festivities (e.g., Ramadan, Christmas). Seasonality may also influence food consumption via seasonal fluctuations in income (e.g., expensive foods could be afforded only sparingly). Some individual-level large-scale food consumption surveys (like the US National Health and Nutrition Survey) are designed so that all days of the week and seasons of the year are represented in the sample, thus capturing variation in consumption due to those factors.

### 2.1.2. Household-level data

Food consumption data from HCES suffer also from measurement error (United Nations, 2005), which is not totally avoidable. The effect of measurement error is much greater for estimates at household level than at population level — for example, random measurement errors are expected to cancel out when averaging to infer the mean for the population. However, data should be statistically treated to remove excess variation due to random measurement error before inferring the variability in the distribution of usual consumption.

In HCES, seasonality can be dealt with in the design phase by distributing the sample equally and uniformly during the year. However, many HCES do not cover a full year mainly due to lack of human and/or economic resources. Not distributing data collection across the year might over- or under- estimate average apparent consumption (Conforti *et al.*, 2017).

The number of food partakers (i.e., the presence of guests/workers who consumed food and the absence of household members during the reference period) is other source of variation in short-term household apparent consumption. This source of variation is due to the nature of the data, aggregated at the level of the household (i.e., there is no information on intra-household food distribution). Finally, there is also between-household variation in apparent consumption due to differences in households' composition (in terms of age, sex, physiological status, physical activity level, and body weight).

Figure 1 summarizes the expected sources of variation in the usual DEI distribution at population level and the sources of variation in short-term dietary energy consumption from individual- and household-level surveys. The sources of variation are classified into those that influence energy requirements and food access (i.e., explain usual dietary energy intake), and those due to survey design that introduce variance inflation.

Figure 1 Expected sources of variation in the distribution of usual per capita DEI used for estimating the PoU based on the FAO approach, and sources of variation in short-term consumption from individual- and household-level data.

| **Expected** sources of variation in the distribution of usual per capita[a] DEI used for estimating the PoU based on the FAO approach | Sources of variation in **short-term individual-level** dietary energy consumption as assessed with 24HR | Sources of variation in **short-term household-level** dietary energy apparent consumption as assessed with HCES |
|---|---|---|
| ***At the individual-level (determining energy requirements)*** [a,b]<br>• Body weight<br>• Physical activity level | ***At the individual-level*** [b] ***(determining energy requirements)***<br>• Sex<br>• Age<br>• Physiological status<br>• Body weight<br>• Physical activity level | |
| ***Factors determining access to food*** [c]<br>• Socio-economic characteristics (e.g., income, education, occupation)<br>• Geographic location | ***Factors determining access to food*** [c]<br>• Socio-economic characteristics (e.g., income, education, occupation)<br>• Geographic location | ***Factors determining access to food*** [c]<br>• Socio-economic characteristics (e.g., income, education, occupation)<br>• Geographic location |
| | ***Induced by the survey design***<br>• Individual day-to-day variability in intake and random measurement error (within-person variation)<br>• Seasonality [d] | ***Induced by the survey design***<br>• Measurement error<br>• Seasonality [d]<br>• Household composition: in terms of sex, age, physiological status, body weight, and physical activity level<br>• Number of food partakers |

Abbreviations: DEI, dietary energy intake; HCES, Household Consumption and Expenditure Surveys.

[a] Energy requirements are determined by individuals' sex, age, physiological status (pregnant, lactating, or none of these), body weight and physical activity level. The per capita distribution refers to an average person of the population in terms of sex, age, and physiological status (pregnancy/lactation); for analytical purposes it is considered as a single homogeneous group. Thus, variation due to sex, age and physiological status are not reflected in the variability of the distribution of usual DEI.

[b] Energy requirements and/or homeostasis may be altered during certain diseases, in obesity, and in response to diet and low levels of physical activity. However, as the influence of these on energy

requirements at a population level is not known they are not considered.

[c] There are socio-cultural factors that influence food consumption, such as preferences, religion, and family traditions. However, the nature of the data precludes assessing their effect, thus, for the purpose of our analyses we assume that usual DEI is not affected by any of these factors.

[d] Some surveys are designed so that all seasons of the year are represented in the sample, thereby given the possibility to adjust for seasonality.

Source: authors' own preparation.

## 3. FAO's approach for estimating a proxy of the variability in usual DEI from household-level data

As it was previously explained, the FAO methodology for estimating the PoU is based on the distribution of daily per capita usual dietary energy intake of an average person representative of the population. The FAO Statistics Division has developed an approach (FAO-CV-Approach) to estimate a proxy for the variability in this distribution using household-level data. This approach, assumes that most of the variation in per capita consumption is explained by: (a) physiological and lifestyle factors that determine energy requirements (body weight and physical activity level), and (b) factors that affect access to food (socio-economic characteristics and geographic location) (FAO, 1996).

Empirically, the variability in usual DEI is proxied with the coefficient of variation ($CV_{usual}$), which is estimated as the sum of two components ($CVusual \cong \sqrt{CV|r^2 + CV|y^2}$): the CV due to factors affecting energy requirements (i.e., given by differences in body weight and physical activity level) $(CV|r)$ and the CV due to factors affecting food accessibility $(CV|y)$.

$$CV_{usual} \cong \sqrt{CV|r^2 + CV|y^2} \quad \text{[Equation 2]}$$

The **CV|y** represents the component of variability that is due to inequality in access to food, which exists between individuals owing to their differences in socio-economic characteristics and geographic location (FAO, 1996). A higher CV|y implies higher inequality in the distribution and reflects a constraint in accessing food to the point that a proportion of the population may not meet their dietary energy needs even if there is enough dietary energy available in the country.

The CV|y is obtained from the short-term household-level data. The first steps are removing between-household differences in the number of food partakers and adjusting for seasonality. Variation between-households' consumption due to different number of food partakers is removed by computing per capita households' apparent consumption, that is, dividing total household's apparent consumption by household size as a proxy for food partakers. The potential effect of seasonality in daily per capita households' dietary energy consumption and income is removed by adjusting them using a seasonal factor for dietary energy and another one for income. These factors indicate how much the regional average monthly dietary energy or income tends to be above or below their respective average for the survey reference period.

Then, the variation due to socio-economic characteristics and geographic location (between-household variation) is isolated by removing variation due to random measurement error and between-household differences in households' composition. This is done applying a regression ([Equation *3*), which models household per capita dietary energy consumption as a function of per capita income, region of residence, and area of residence (urban or rural). The regression parameters of the equation are estimated with ordinary least squares. Other socio-economic characteristics were tested as independent variables, but they did not prove better at predicting DEI.

$$\begin{aligned} DEC_h = {} & \ln(Income_h) + \ln(Income_h)^2 + Region_h + Urban_h + Region_h * Urban_h \\ & + \ln(Income_h) * Urban_h + \ln(Income_h)^2 * Urban_h + \varepsilon_h \end{aligned} \qquad \text{[Equation 3]}$$

where *h* refers to household; DEC is the daily per capita dietary energy apparent consumption in household *h*; Income is the daily per capita income in household *h*; *Region* is a set of dummy variables indicating the region or province in which the household *h* is located; *Urban* is a dummy variable indicating whether the household *h* is in urban area (rural area serve as reference category); ε is the error term.

The CV of the predicted dietary energy values is used as a proxy for the CV|y, which is estimated as the ratio of the inferred parameters[5] standard deviation and mean of the predicted values

[5] Inferred using population weights (household weights*the number of household members).

([Equation *4*). Nevertheless, it should be clear that each of the predicted dietary energy points do not represent estimates of household's usual dietary energy consumption.

$$CV|y \cong \frac{sd_{pfv}}{mean_{pfv}} \qquad \text{[Equation 4]}$$

Where $CV|y$ is the coefficient of variation due to socio-economic characteristics and geographic location; $sd_{pfv}$ and $mean_{pfv}$ are, respectively, the inferred standard deviation and the inferred mean of the distribution of the predicted fitted values of dietary energy consumption.

Summarizing, the regression not only removes the non-desirable variation due to measurement error but also between-household variation due to differences in household composition (in terms of age, sex, physiological status, physical activity level, and body weight). However, the FAO methodology for estimating the PoU is based on the distribution of daily per capita usual dietary energy intake of an average person representative of the population. As such, this constructed average person, reflects all sexes, ages, and physiological statuses. To better understand this idea, the distribution of usual dietary energy intake, where the reference unit is the average person, should be seen as a distribution of individuals of the same sex and pregnant/lactating status, and similar age. Therefore, the distribution of per capita usual dietary energy intake does not have variation due to differences in sex, age, and pregnant/lactating status, but it has between-individual variation due to differences in body weights and physical activity levels. This variation is estimated with the CV|r.

The **CV|r** reflects the per capita variation in DEI due to energy requirements given by different body weights and physical activity levels. It is estimated using [Equation *5*.

$$CV|r = \frac{\sigma}{\mu} \qquad \text{[Equation 5]}$$

Where $CV|r$ is the coefficient of variation due to differences in body weight and physical activity levels; $\mu$ is the population mean; and $\sigma$ is the population standard deviation (SD).

Assuming that energy requirements are normally distributed, the mean and SD can be estimated if at least two percentiles and their values in the distribution are known. The two percentiles chosen are the 1st and 99th and for each sex-age group they are approximated with the minimum dietary energy requirement (MDER) and the maximum dietary energy requirement (XDER) for that group. The MDER and the XDER are estimated using FAO, WHO, and the United Nations University equations for human energy requirements (FAO, 2004) (See S1 Appendix). The MDER and the XDER of the whole population are computed as a weighted average of the MDERs and XDERs, respectively, of 62 sex-age groups (FAO, 2004). The weights are the proportion of individuals in each group, thus, the value of the CV|r also depends on the sex-age structure of the population.

The mean and SD can be estimated relying on the definition of the z-score: Z=(x-μ)/σ (where x is one of the values of the distribution, μ is the mean of the distribution, and σ is the standard deviation (SD) of the distribution) and solving a two-equation system ([Equation *6* and [Equation *7*):

$$z(0.01) = \frac{(MDER - \mu)}{\sigma} \qquad \text{[Equation 6]}$$

$$z(0.99) = \frac{(XDER - \mu)}{\sigma} \qquad \text{[Equation 7]}$$

Where $z$ is the z-score; *MDER* is the minimum dietary energy requirement; *XDER* is the maximum dietary energy requirement; $\mu$ is the population mean; and $\sigma$ is the population standard deviation.

If there were no constraints in access to food in a population, the CV due to differences in socio-economic characteristics and geographic location (CV|y) would be zero and the CV of the distribution of usual DEI would equal the CV due to differences in body weight and physical activity (CV|r).

## 4. Estimating usual dietary energy intake from individual-level data using the NCI method

Several statistical methods have been developed to address the issue of removing within-person variation from short-term individual-level consumption data collected with 24HR surveys with at least a second round, which could be performed to all the individuals or a random subset of the population. The method developed by the US NCI is considered as one of the most sophisticated ones. It was originally meant for nutrition survey data from complex surveys in the US but has since been used to analyze data from many other countries. The NCI-method to derive percentiles of the distribution of usual intakes is implemented in two SAS macros (mixtran and distrib) developed by researchers at the NCI (National Cancer Institute, 2009).

The NCI-Method estimates usual dietary intake for foods and nutrients (including energy) based on a mixed effects model. It separates intake into two parts: the probability of consumption (for episodically consumed foods and nutrients) and the amount consumed (of the food or nutrient) on a "consumption day" (Tooze *et al.*, 2010). Given that, dietary energy is not episodically but daily consumed considering only the amount part of the model is sufficient. Covariates aid in the model fitting process and the estimation of the within- and between-person variation (Kipnis *et al.*, 2009; Tooze *et al.*, 2010).

A detailed description of the NCI-Method has been provided by Tooze and colleagues (2006, 2010). It can be summarized in three main steps:

1. **Fit model and data transformation to approximate normality.** First, a model is fit to the data. The main purpose of the model is to estimate how much of the variation in the reported intakes is within-person variation[6], so that this can then be omitted to subsequently simulate the usual intake distribution. The rest of the variation, the between-person variation, is a combination of "structured" variability that is explained by the covariates (this is the fixed effect part of the model) and of "unstructured" or residual-between person variation from all

[6] Within-person variance for a person could be estimated from the sample, but there are few degrees of freedom and the estimate is imprecise, so data are pooled and common variance is assumed. In this way a pooled within-person variance estimate is obtained, which is more precise.

other unmeasured characteristics (the random effect part of the model). Given the skewed nature of dietary intake distributions, data are transformed using a Box-Cox transformation, which includes the logarithmic transformation as a limiting case. The Box-Cox parameter (lambda) is estimated during the model fitting procedure at the same time the covariate effects are estimated so that the best transformation is chosen after adjusting for these effects. The within-person variation and residual between-person variation are also estimated during the model fitting. Both the within- and between-person variation are modelled as a normal distribution (after transformation) with a mean of zero.

2. **Simulation of usual intakes based on model fitted.** Percentiles of the distribution of usual nutrient intake are estimated by using a Monte Carlo simulation procedure. The method predicts consumption for each individual in the sample based on the estimated intercept and parameters for each covariate. Then, to add an estimate of the person-specific random effects, the method generates simulated intake values for 100 pseudo-persons for each sampled individual. Each of the 100 pseudo-persons has the same structured variation of their correspondent individual in the sample, but different simulated person-specific effects, so the covariate pattern of the simulated intakes is proportional to that found in the sample. The within-person variation is not included in the simulated intakes since, by definition, it does not contribute to usual intakes.

3. **Back-transform to original scale.** Lastly, the simulated pseudo-person intakes are back transformed to the original scale (or units) to give a simulated population usual intake distribution. A population mean and percentiles of intake are derived empirically from this representative sample of the back-transformed values. Individuals' sample weights are considered to ensure the results represent the population. In the distribution of usual intake obtained, the points of the distribution do not represent the "true" usual intake of individuals. However, the mean and spread of the distribution of the group are expected to represent the parameters of the distribution of usual intake in the population.

The NCI-Method assumes that 24HR intake is an unbiased estimator of individual usual intake on the original scale. However, 24HR are known to under report energy intake; therefore, if systematic errors are not previously corrected the NCI-Method does not remove their potential effect when estimating the mean usual intake (Ahluwalia *et al.*, 2016).

## 5. Methods

### 5.1. Data

The Bangladesh Integrated Household Survey (BIHS) is a three-round panel household survey representative of rural Bangladesh, which collects information on household-level and individual-level food consumption from the same households. The current study is an analysis of the second round conducted from January to June in 2015. The 2015 BHIS covered a total of 6,715 households (IFPRI, 2016a). We used the BIHS sample representative of rural Bangladesh, which included 5,569 households and 23,135 individuals. S2 Appendix presents some characteristics of the survey data.

We used two food consumption modules that are based on interviews with the female household member in charge of cooking, supervising, and serving. The 7DR module contains information on household food consumption (using a 321-item food list) in the previous 7 days. The 24HR module, collected information on household consumption of prepared foods over the prior 24 hours, the description of each prepared dish, and the intra-household food allocation. The 24HR was applied a second time to 10% of the households. Both 24HRs were used in the analysis.

For the 7DR module, we excluded 124 households identified with incomplete surveys (refused, migrated or were not at home during survey) by data owners. We also excluded household members less than two years of age because a large number were breastfed. The final analytical sample for the 7DR consisted of 5,427 households with 22,319 household members. In the 24HR module, we excluded individuals with missing 24HR data that were away from home or did not take any food. The final analytical sample for the first 24HR consisted of 5,424 households and 21,310 individuals. The two samples did not differ in key socio-demographic characteristics (sex, region, and income). Additional data were obtained from other 2015 BIHS files: household's socio-economic characteristics, household composition in terms of sex and age, pregnant and lactating status, and height for all household members. All files were downloaded from the open-source research data repository Harvard Dataverse (IFPRI, 2016b) on January 2019. A detailed description of the sampling procedures, survey methodology, and questionnaires has been described elsewhere (IFPRI, 2016a).

### 5.2. Preparation of the food consumption data

Food consumption data from the 7DR were prepared following the approach outlined by Moltedo and colleagues for HCES data processing (Moltedo, A. Troubat, N. Lokshin, M. Sajaia, 2014). The 24HR was processed analogously, with some exceptions. The steps followed to process both modules were the following: 1) Matching of food items to food composition data and standardization of quantities into grams. (2) For the 7DR only, estimation of the dietary energy of food consumed away from home (FCAH) using median calorie monetary unit values. (3) Outlier detection and imputation.

#### 5.2.1. Source of food composition data and non-edible portions

For both, the 24HR and the 7DR, the dietary energy content and non-edible portions by food were obtained from the Food Composition Table for Bangladesh (Shaheen, N. Rahim and Banu, C. Bari, L. Tukun, A. Mannan, M. Bhattacharjee, L. Stadlmayr, 2013). Where necessary, it was supplemented with values from the Indian FCT (Longvah, T. Ananthan, R. Bhaskarachary, K. and Venkaiah, 2017), the ASEAN Food Composition Database (Institute of Nutrition, 2014), and the USDA Food Composition Databases (Laboratory, 2018). In the absence of non-edible portion information for a given food (e.g., for some types of fish), we used the respective value of a similar item.

#### 5.2.2. Food matching and unit standardization

**24HR dataset**

The food matching was based considering cooking methods (raw, boiled, fried, etc.). Nevertheless, it is expected that the type of food matching (i.e., based on the nutrient content in raw or cooked form of the food) has no impact on average apparent dietary energy estimates (Moltedo *et al.*, 2021). There were five approaches, to convert quantities into Kcal, involving single-ingredient foods and multi-ingredient dishes:

- *Single-ingredient foods consumed raw* (e.g., apple) were matched with a raw item in the FCTs; raw weights were used; adjustment for non-edible portions was applied as necessary.
- *Single-ingredient foods* consumed *cooked* (e.g., rice) were matched with a raw item in the FCTs; retention factors were applied to account for alterations in nutrient content during

cooking; raw weights were used; adjustment for non-edible portions was applied as necessary.

- *Multi-ingredient dishes* consumed *cooked* with *no list of ingredients* (e.g., burger) were matched with cooked mixed dishes in the FCTs; cooked weights were used.
- *Multi-ingredient dishes* consumed raw with a *list of ingredients* (e.g., salad); each item/ingredient was matched with a raw item in the FCTs; raw weights were used; adjustment for non-edible portions was applied as necessary.
- *Multi-ingredient dishes consumed cooked with* a *list of ingredients* (e.g., bhuna curry); each item/ingredient was matched with a raw item in the FCTs; retention factors were applied to account for alterations in nutrient content during cooking; raw weights were used; adjustment for non-edible portions was applied as necessary.

**7DR dataset**

The 7DR module collect information on quantities only for single-ingredient foods; the dietary energy content of these foods was calculated as described for the 24HR (the same nutrient conversion table was used for both modules). Where necessary, quantities were adjusted for non-edible portions. This module did not collect quantities for mixed dishes consumed away from home.

### 5.2.3. Estimation of dietary energy content in FCAH from the 7DR

As in the 24HR, the 7DR module collected information on FCAH (over 20 foods); however, it did not collect information on quantities consumed, only on the associated expenditures. Therefore, we estimated the dietary energy of these foods using information on their monetary values along with median (at region-income quintile level) dietary energy at-home unit cost (Moltedo, A. Álvarez-Sánchez, C. Troubat, N. Cafiero, 2018). This approach assumes that the implicit price of the calories in the FCAH is equal to that of food prepared at home. We are cognizant that this is not an adequate assumption, as the price of FCAH is likely different (and possibly higher) than the price of home-cooked food. However, we decided against removing FCAH from our analysis, because of the increasing importance FCAH is taking in many countries and because that could have resulted in an even larger bias.

### 5.2.4. Outlier detection and imputation

**24HR dataset**

Outlier detection and imputation was performed on the dietary energy consumption values, separately for each of the sex-age groupings defined[7]. We identified outlying values that fell outside the interval given by the 25$^{th}$ percentile minus twice the interquartile range (IQR) to 75$^{th}$ percentile plus twice the IQR on the logarithmic scale. Outliers on the left and right tails were imputed with the quantity corresponding to the 25$^{th}$ or 75$^{th}$ percentile, respectively, of the distribution of dietary energy consumption of individuals of the same sex-age group.

The potential effect of seasonality in dietary energy consumption was removed by adjusting the values using a seasonal factor.

**7DR dataset**

Outlier detection and imputation was conducted on food quantities per capita for each food item, as it is typically done with household-level food consumption data. We identified outlying values that fell outside the interval given by the 25$^{th}$ percentile minus twice the IQR to 75$^{th}$ percentile plus twice the IQR on the logarithmic scale. For each food item, outliers on the left and right tails were imputed with the quantity corresponding to the 25$^{th}$ or 75$^{th}$ percentile of the distribution, respectively. We did not conduct outlier detection on the dietary energy consumption values because the nature of the data prevents identifying consumption by sex-age group.

Missing information on height was imputed with Bangladesh's median height of the corresponding sex-age group (height information was needed to calculate dietary energy requirements for the estimation of the CV|r).

## 5.3. Statistical methods

### 5.3.1. Individual-level data (24HR)

Usual dietary energy consumption was modelled with the NCI-Method implemented in SAS macros (National Cancer Institute, 2009) using the two rounds of the 24HR. We used the 1-part

[7] Fifty-eight groups in total, based on the groupings proposed by FAO (2004) to estimate dietary energy requirements, and excluding infants less than two years of age.

nonlinear mixed model because dietary energy is consumed daily by everyone (Tooze *et al.*, 2010). The NCI-Method requires a sufficiently large sample size to provide stable estimates of the variance components, which may be achieved by pooling categories of important covariates (e.g., age and sex groups) (Krebs-Smith *et al.*, 2010). Davis and colleagues (Davis *et al.*, 2019) studied the effect of stratification versus pooling in nutrient consumption with the NCI-Method and found that while there were no meaningful differences in mean consumption, some of the percentiles on the tails for some age groups differed largely. To assess any potential impact, on the CV, of data stratification compared to pooling we run three different models: (1) pooled sample including sex and age group as covariates, (2) stratified sample into two sex groups and including age group as a covariate, (3) stratified sample into 58 sex-age groups. In the three models we incorporated covariates to account for sequence effect and weekend/weekday effect. Household income (decile) and the region as geographic location were included as covariates in each of the models — we also run analyses with household per capita income, but result did not differ. From each of the three models, we derived the distribution of usual dietary energy consumption at national (rural) level. The points of the distribution do not represent individuals' consumption; however, the CV of the distribution could be used a proxy of the CV in the distribution of usual dietary energy intake. The CV was calculated as the standard deviation divided by the mean of the distribution, using weights.

While for estimating a proxy of the CV in the distribution of usual dietary energy intake we used the two rounds of the 24HR, for estimating the CV in the distribution of short-term dietary energy consumption we used only data from the first 24HR. The CV in the short-term distribution, as derived from the observed self-reported quantities (empirical CV), was estimated as the population standard deviation divided by the population mean.

### 5.3.2. Household-level data (7DR)

An empirical (unadjusted) CV of the short-term apparent dietary energy consumption distribution (as derived from the observed self-reported quantities), at national (rural) level, was estimated as the population standard deviation divided by the population mean.

A proxy for the CV in the distribution of usual DEI was estimated as described in Section 3. Additional specific information on the computation of the CV|r and CV/y using the 2015 BIHS dataset is provided below. In the estimation of the component of variation due to differences in body weight and physical activity level (CV|r), the MDER and the XDER for the average person of the population were computed, using the weighted average of the MDERs and XDERs of 58 sex-age groups[8] implementing Equation 5 through Equation 7. S1 Appendix shows the equations used for calculating the MDER and the XDER. S3 Appendix is an Excel that shows a numerical example of the calculation of the MDER and XDER for the 2015 BIHS (where children less than 2 years were excluded from the analysis). The Excel file also includes an example when children less than 2 years are not excluded from the analysis.

The component of the variation due to factors affecting food accessibility (CV|y) was obtained applying the procedures and equations described in Section 3. Urban/rural location was not included in the regression because the survey is only representative at the rural level (Equation 8).

$$DEC_h = \ln(Income_h) + \ln(Income_h)^2 + Region_h + \ln(Income_h) * Region_h + \ln(Income_h)^2 * Region_h + \varepsilon_h \quad \text{[Equation 8]}$$

where *h* refers to household; DEC is the daily per capita dietary energy apparent consumption in household *h*; Income is the daily per capita income in household *h*; *Region* is a set of dummy variables indicating the region or province in which the household *h* is located; ε is the error term.

Finally, the CV|y was computed as the population SD divided by the population mean of the predicted daily per capita dietary energy consumption. Information on households' income (proxied with information on the aggregate household's expenditures) and sampling weights were received from IFPRI staff responsible for the survey. Income information was used in the regression to obtain the CV|y; and in classifying household into income quintiles, needed for estimating dietary energy content from FCAH[9].

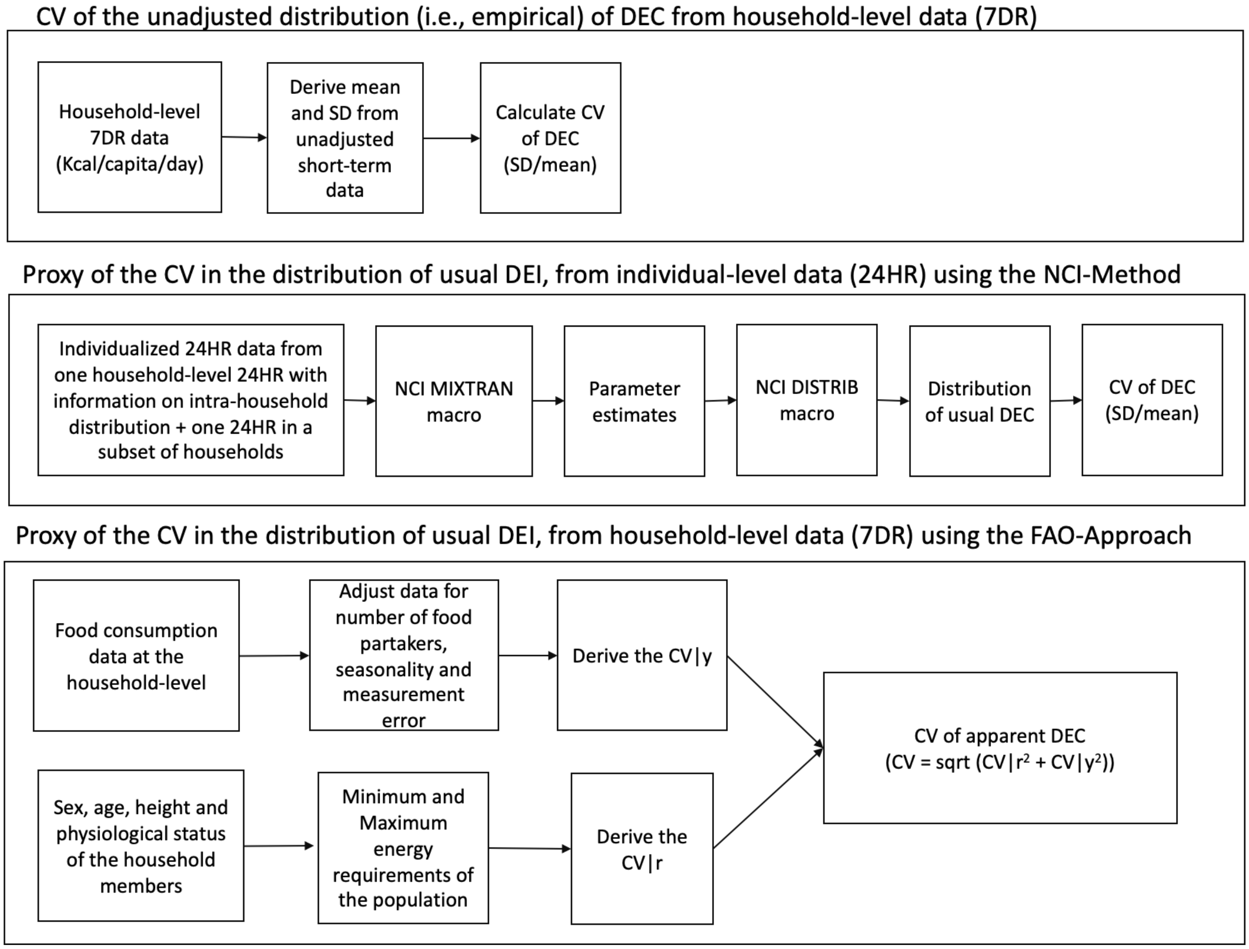
---

[8] The four groups corresponding to children less than two years were not considered because these children were excluded from the analysis.

[9] Households' total expenditure and food expenditure —used to estimate FCAH — were deflated with information on Consumer and Food Price Indices, respectively, from FAOSTAT (FAO, 2021a) as described in Moltedo *et. al*. (2014).

All statistics were inferred to the population using sampling weights. Figure 2 presents the main steps in involved in the estimation of the coefficient of variation from different data types.

Figure 2 Illustration of the main steps in estimating the coefficient of variation used in the analysis




Notes. CV, coefficient of variation; DEC, dietary energy consumption; DEI, dietary energy intake; NCI, National Cancer Institute; SD, standard deviation; 24HR, 24-hour recall; 7DR, 7-day recall; CV|y, CV due to socio-economic characteristics and geographic location; CV|r, CV due to differences in body weight and physical activity level.

# 6. Results

## 6.1. Individual-level data (24HR)

The empirical CV of the short-term distribution of dietary energy consumption (i.e., not adjusted for within-person variation), at national (rural) level, based on the first 24HR was 36.8%. It was estimated based on a standard deviation of 751 (kcal/capita/day) and an average of 2040 Kcal/capita/day. Table 1 shows the mean, standard deviation, and coefficient of variation (CV) of the distribution of usual dietary energy consumption based on individual-level data (24HR), as estimated with three different models: (1) pooled sample, (2) stratified by sex, (3) stratified into 58 sex-age groups. The three models produced comparable average consumption (2040, 2041 and

2041 Kcal/capita/day, respectively for models 1, 2, and 3) and standard deviations (449, 451, and 451 Kcal/capita/day, respectively for models 1, 2, and 3). After outlier treatment, the ratios of the within-person to between-person variance for all sex-age groups, in the sex-age stratified sample, were similar confirming the appropriateness of pooling the sex-age groups (Davis *et al.*, 2019). Therefore, we selected the CV from the pooled sample model for comparison to the CV obtained with the FAO-CV-Approach applied to household-level data. Nevertheless, the calculated CVs from the three models are very similar, diverging only in one decimal point (22.0%, 22.1%, and 22.1 % respectively for models 1, 2, and 3).

Table 1 Mean, standard deviation, and coefficient of variation of the distribution of usual dietary energy consumption based on individual-level data (24HR)

| | **Covariates included in the model** | **Mean (Kcal/capita/day)** | **Standard deviation (Kcal/capita/day)** | **CV of the distribution of usual DEI (%) (SD/mean)** |
|---|---|---|---|---|
| **Pooled sample** | Sex, age, income, region, sequence and weekend/weekday[a] | 2040 | 449 | 22.0 |
| **Stratified by sex** | Age, income, region, sequence and weekend/weekday[b] | 2041 | 451 | 22.1 |
| **Stratified by sex and age** | Income, region, sequence and weekend/weekday[c] | 2041 | 451 | 22.1 |

Notes. SD, standard deviation; CV, coefficient of variation; Kcal, kilocalories. The CV is calculated as the SD divided by the mean. Sequence refers to the order of administration of the 24-hr recall: first or second.

The mean and the SD were derived after treating data with the NCI-Method.

[a] We included one dummy variable for sex, 28 dummy variables for age, 6 dummy variables for region and 9 dummy variables for income decile groups.

[b] We included 28 dummy variables for age, 6 dummy variables for region and 9 dummy variables for income decile groups.

[c] We included 6 dummy variables for region and 9 dummy variables for income decile groups.

### 6.2. Household-level data (24HR)

The empirical CV of the unadjusted short-term distribution of dietary energy consumption based on household-level data (7DR) is 32.3 % (Table 2). This CV is about 10 percentage points higher than the adjusted CV (22.9 %) estimated from household-level data (7DR) applying the FAO-CV-Approach (Table 3). The adjusted CV, used as a proxy of the CV in the distribution of usual DEI distribution, is very similar to the CV of usual dietary energy consumption from individual-level data obtained applying the NCI-Method (22.0%).

Table 2 Mean, standard deviation, and empirical coefficient of variation of the distribution of short-term dietary energy consumption based on household-level data (7DR)

| **Mean (Kcal/capita/day)** | **Standard deviation (Kcal/capita/day)** | **Empirical CV of the distribution of short-term DEI (SD/mean) (%)** |
|---|---|---|
| 2649 | 856 | 32.3 |

Notes. SD, standard deviation; CV, coefficient of variation; Kcal, kilocalories. The CV is calculated as the population SD divided by the population mean.

[a] The CV Empirical is the CV of the distribution of short-term per capita dietary energy consumption from 7DR data.

Table 3 Proxy for the CV in the distribution of usual dietary energy intake based on household-level data (7DR) derived with the FAO-CV-Approach

| **Mean (Kcal/capita/day)** | **CV due to body weight and PAL (CV\|r) [a] (%)** | **CV due to socio-economic characteristics and geographic location (CV\|y) [b] (%)** | **CV of the distribution of usual DEI [c] (%)** |
|---|---|---|---|
| 2649 | 9.0 | 21.1 | 22.9 |

Notes. CV, coefficient of variation; DEI, dietary energy intake; Kcal, kilocalories; PAL, Physical Activity Level.

[a] The CV|r was calculated using the population's Minimum and Maximum Energy Requirements.

[b] The CV|y was calculated, using the regression's predicted values, as the population standard deviation divided by the population mean.

[c] The CV is calculated as CV = sqrt ($CV|r^2 + CV|y^2$).

## 7. Discussion

This paper describes the FAO-CV-Approach for estimating a proxy of the variability in the distribution of usual DEI at national level using HCES data, and compares the CV obtained with this approach to an estimate derived from individual-level data (from the same households) treated with the NCI-Method.

The study findings confirmed that the empirical distribution of apparent dietary energy consumption from household-level data has excess variation that is not part of the variability in the distribution of usual DEI, which should be statistically corrected before estimating the prevalence of dietary energy inadequacy. The analysis also found that the FAO-CV-Approach produced close estimates of the variability in usual DEI to the NCI-Method. Using information from the 2015 BHS, the CV obtained with the FAO-CV-Approach, applied to household-level data, was 22.9% and the CV obtained after applying the NCI-Method to individual-level data was 22.0%. In comparison, the CV of the 7DR empirical distribution (unadjusted for measurement errors) was substantially higher (32.3%), by about 10 percentage points, than the CV obtained with the other two methods.

In the interpretation of the results of this study, it is important to note that the assumptions and statistical frameworks of the FAO-CV-Approach and NCI-Methods are by no means the same. The NCI-Method uses individual-level data, and it is based on the decomposition of overall variance in within- and between-person, and removal of the within-person variation. Whereas the FAO-CV-Approach uses household level data and proxies the variation in usual DEI considering key factors that affect access to food and energy requirements. However, our study found that the resulting variation obtained from both methods converges, thus confirming, that the FAO-CV-Approach is suitable for estimating a proxy for the variation in the distribution of usual DEI from household-level data. This is likely to be a proxy of the between-individual variation that exists in the true population distribution of usual DEI. Nonetheless, it should be kept in mind that the true usual DEI distribution remains unknown because of the lack of objective validation data.

The CVs of DEI distributions, empirical and/or corrected for variance inflation, are rarely published. There are only a handful of studies that have published CVs of dietary energy for specific sex-age groups based on 24HR data (Marr and Heady, 1986; Harbottle and Duggan, 1994;

Nyambose, Koski and Tucker, 2002; Verly-Junior *et al.*, 2010; Willett, 2012), but no study based on individual level data has published CVs at national level.

The CV of usual DEI at national level is fundamental for estimating the PoU for global monitoring, which is the Sustainable Development Goal 2.1.1 indicator. The FAO Statistics Division has worked on estimating a proxy for the CV of the usual DEI distribution at national level from HCES for more than two decades, testing various statistical approaches along the way. There could be alternative approaches to proxy the variability in the distribution of usual DEI from HCES; however, the FAO-CV-Approach represents the best one found by the FAO Statistics Division until now. Further research on data collection and methods for improving estimates of dietary energy consumption from foods with no information on quantities (most prepared and consumed away from home) are needed. Since 2000, the FAO Statistics Division has published the CVs of usual DEI, for countries, which can be downloaded from the FAO Food security indicators site (FAO, 2021b). The data from the 2015 BIHS were representative for rural Bangladesh only, therefore we could not compare the CV to the estimates published by FAO for global monitoring, which refer to the whole population.

An important point of discussion, although not directly related to the objective of the study, is that of the difference in average dietary energy consumption obtained from the 24HR and the 7DR data. Using data from a single 24HR and a 7DR from the BIHS 2011, Karageorgou and colleagues (Karageorgou *et al.*, 2018) found that the average apparent consumption from the 7DR was 241 Kcal (excluding the food consumed away from home collected in the 7DR module [personal communication]) higher than that of a single 24HR; and concluded that the 7DR module "substantially overestimated" individual consumption. However, this is a debatable conclusion. While it may be tempting to make this judgement, considering that individual-level food consumption data are typically regarded as of higher quality, it is important to note a few aspects about the 24HR instrument included in the BIHS, and about 24HRs in general. In the BIHS, the 24HR is not a standard individual-level food consumption survey. It is rather a household-level 24HR with information on intra-household food distribution, where the female household member in charge of shopping and cooking reports food consumption for all the other household members. While responding by proxy is a standard approach for small children, whose food consumption depends on their caregivers (Livingstone and Robson, 2000), it may lead to important reporting biases in adults. As per FCAH, the numerators asked all household members present during the survey, if and what they ate outside, however, sometimes those members were not present and the proxy respondent could not report if and what they had eaten. Furthermore, it is widely acknowledged that 24HR systematically underreport dietary energy intake (Ravelli and Schoeller, 2020). The problem of underreporting of energy intake in individual-level surveys is so generally recognized, that a method has been proposed to adjust distributions of food consumption based on 24HR data for this systematic bias using biomarker data, in addition to adjusting for within-person variation (Yanetz *et al.*, 2008). Unlike the standardized 24HR in individual-level surveys, a wide spectrum of survey design exists for HCES that depend on the purpose of the survey, but dietary energy or nutrient consumption analysis are never the primary purpose of the surveys (FAO and The World Bank, 2018). Nevertheless, an analysis of 81 HCES revealed that the survey characteristics have an impact on the reported level of food consumption and so on average dietary energy consumption, but not on dietary energy variability estimates after having been adjusted for measurement error (Conforti *et al.*, 2017). Considering all the above, it is impossible to categorically conclude that average apparent dietary energy consumption estimates based on household-level data from the BIHS 2015 are more “biased” than average dietary energy consumption derived using data from the 24HR module. Nevertheless, even a small difference in the proxy used for estimating the mean of the distribution of usual DEI might have a significant impact on the PoU estimates (Conforti *et al.*, 2017); therefore, PoU estimates calculated using these two data sources would largely differ. This highlights the importance of clearly report the data source used (i.e., individual, household or food balance sheet data) to estimate a proxy of the mean in the distribution of usual DEI, which is a parameter needed to estimate the PoU. Furthermore, due to the different levels of average dietary energy consumption estimates between using individual and household level data, this analysis is a clear example of the importance of using the CV, and not the standard deviation, as a measure of variability in the distribution of usual DEI because of its independence from the mean.

## 8. Conclusion

The study found that the empirical short-term distribution of dietary energy consumption from household-level data has excess variation that is not part of the variability in the distribution of usual dietary energy intake. Therefore, household-level data should be statistically adjusted for measurement error before estimating a proxy for the variation in the distribution of usual dietary energy intake at the population level. This also confirms that dietary energy consumption at household level is prone to measurement error. Therefore, dietary energy values at the household level should be adjusted before being used for analysis such as comparing them with dietary energy requirements of household's members. The analysis also found that the FAO-CV-Approach produced close estimates of the variability in usual dietary energy intake to the NCI-Method.

## 9. Recommendations

This analysis was based on only one survey representative of a rural population. Further analysis of this type for the whole population (i.e., representative at national level) are needed to support the conclusions.

## Acknowledgment

The authors gratefully acknowledge the financial support from the European Union through the INFORMED project; the contributions received from Abdul Sattar and Filippo Gheri; and the support provided by Akhter Ahmed and Carlo Azzarri for a better understanding of the microdata.

## References


Ahluwalia, N., Dwyer, J., Terry, A., Moshfegh, A. and Johnson, C. (2016) 'Update on NHANES Dietary Data: Focus on Collection, Release, Analytical Considerations, and Uses to Inform Public Policy', *Advances in nutrition (Bethesda, Md.)*. American Society for Nutrition, 7(1), pp. 121–134. doi: 10.3945/an.115.009258.

Cafiero, C. (2014) *Advances in hunger measurement: Traditional FAO methods and recent innovations, ESS Working Paper No. 14-04*. Available at: http://www.fao.org/3/a-i4060e.pdf.

Carriquiry, A. (2017) 'Understanding and assessing nutrition', *Annual Review of Statistics and Its Application*, 4, pp. 123–146. Available at: http://www.annualreviews.org/doi/pdf/10.1146/annurev-statistics-041715-033615.

Conforti, P., Grünberger, K. and Troubat, N. (2017) 'The impact of survey characteristics on the measurement of food consumption', *Food Policy*, 72, pp. 43–52. doi: https://doi.org/10.1016/j.foodpol.2017.08.011.

Davis, K. A. *et al.* (2019) 'Early Experience Analyzing Dietary Intake Data from the Canadian Community Health Survey—Nutrition Using the National Cancer Institute (NCI) Method', *Nutrients*. Multidisciplinary Digital Publishing Institute, 11(8), p. 1908. doi: 10.3390/nu11081908.

Dodd, K. W. *et al.* (2006) 'Statistical methods for estimating usual intake of nutrients and foods: a review of the theory', *J Am Diet Assoc*, 106(10), pp. 1640–1650. doi: 10.1016/j.jada.2006.07.011.

Elia, M. (2005) 'Insights into energy requirements in disease.', *Public health nutrition*. England, 8(7A), pp. 1037–1052. doi: 10.1079/phn2005795.

FAO (1996) *The Sixth World Food Survey*. Rome. Available at: http://www.fao.org/docrep/012/w0931e/w0931e.pdf.

FAO (2004) *Human energy requirements. Report of a Joint FAO/WHO/UNU Expert Consultation. FAO Food and Nutrition Technical Report Series No. 1*. Rome: FAO. Available at: http://www.fao.org/3/a-y5686e.pdf.

FAO (2021a) *FAOSTAT - Consumer Price Indices*. Available at: https://www.fao.org/faostat/en/#data/CP (Accessed: 1 June 2021).

FAO (2021b) *FAOSTAT - Suite of Food Security Indicators*. Available at: https://www.fao.org/faostat/en/#data/FS (Accessed: 18 October 2021).

FAO and The World Bank (2018) *Food data collection in Household Consumption and Expenditure Surveys. Guidelines for low- and middle-income countries.* Rome. Available at: http://www.fao.org/3/CA1561EN/ca1561en.pdf.

FAO, IFAD, UNICEF, WFP and WHO (2021) *The State of Food Security and Nutrition in the World 2021. Transforming food systems for food security, improved nutrition and affordable healthy diets for all*. Rome. Available at: https://doi.org/10.4060/cb4474en.

Fiedler, J. L., Lividini, K., Bermudez, O. I. and Smitz, M. F. (2012) 'Household Consumption and Expenditures Surveys (HCES): a primer for food and nutrition analysts in low- and middle-income countries', *Food Nutr Bull*. 2012/12/01. HarvestPlus, IFPRI, 2033 K Street NW, Washington, DC 20006, USA. j.fiedler@cgiar.org, 33(3 Suppl), pp. S170-84. doi: 10.1177/15648265120333S205.

IFPRI (2016a) '000_IFPRI-PRSSP_Description of the BIHS datasets.pdf', *Bangladesh Integrated Household Survey (BIHS) 2015*. V3 edn. Edited by I. F. P. R. I. (IFPRI) A.-U. S. A. for I. D. (USAID). Harvard Dataverse (Household- and Community-level Surveys). doi: doi:10.7910/DVN/BXSYEL/Z0CJD3.

IFPRI (2016b) 'Bangladesh Integrated Household Survey (BIHS) 2015'. Harvard Dataverse (Household- and Community-level Surveys). doi: doi:10.7910/DVN/BXSYEL.

Institute of Nutrition, M. U. (2014) 'ASEAN food composition database V 1.0'. Thailand. Available at: http://www.inmu.mahidol.ac.th/aseanfoods/doc/OnlineASEAN_FCD_V1_2014.pdf.

Karageorgou, D., Imamura, F., Zhang, J., Shi, P., Mozaffarian, D. and Micha, R. (2018) 'Assessing dietary intakes from household budget surveys: A national analysis in Bangladesh', *PLoS One*. 2018/08/27, 13(8), p. e0202831. doi: 10.1371/journal.pone.0202831.

Kipnis, V. *et al.* (2009) 'Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes.', *Biometrics*, 65(4), pp. 1003–1010. doi: 10.1111/j.1541-0420.2009.01223.x.

Krebs-Smith, S. M., Guenther, P. M., Subar, A. F., Kirkpatrick, S. I. and Dodd, K. W. (2010) 'Americans do not meet federal dietary recommendations', *J Nutr,* 140. doi: 10.3945/jn.110.124826 LB - Krebs-Smith2010.

Laboratory, N. D. (2018) 'USDA national nutrient database for standard reference, Release 28'. Riverdale, MD: US Department of Agriculture, Agricultural Research Service. Available at: https://ndb.nal.usda.gov/.

Laureano, G. H., Torman, V. B., Crispim, S. P., Dekkers, A. L. and Camey, S. A. (2016) 'Comparison of the ISU, NCI, MSM, and SPADE Methods for Estimating Usual Intake: A Simulation Study of Nutrients Consumed Daily', *Nutrients*. 2016/03/15, 8(3), p. 166. doi: 10.3390/nu8030166.

Livingstone, M. B. and Robson, P. J. (2000) 'Measurement of dietary intake in children.', *The Proceedings of the Nutrition Society*. England, 59(2), pp. 279–293. doi:

10.1017/s0029665100000318.

Longvah, T. Ananthan, R. Bhaskarachary, K. and Venkaiah, K. (2017) *Indian Food Composition Table*. Available at: http://www.ifct2017.com/frame.php?page=food (Accessed: 21 February 2019).

Moltedo, A. Troubat, N. Lokshin, M. Sajaia, Z. (2014) *Analyzing Food Security Using Household Survey Data: Streamlined Analysis with ADePT Software*. Washington (DC): World Bank. Available at: https://openknowledge.worldbank.org/handle/10986/18091.

Moltedo, A., Alvarez-Sanchez, C., Troubat, N. and Cafiero, C. (2018) *Optimizing the use of ADePT-Food Security Module for Nutrient Analysis: ADePT-FSM Version 3*. Edited by S. D. FAO. Rome (Italy): FAO. Available at: http://www.fao.org/fileadmin/templates/ess/foodsecurity/Optimizing_the_use_of_ADePT_FSM_for_nutrient_analysis.pdf.

Moltedo, A., Jiménez, S., Álvarez-Sánchez, C., Manyani, T., Ramos, M. P. and Custodio, E. (2021) 'Raw versus cooked food matching: Nutrient intake using the 2015/16 Kenya Integrated Household Budget Survey', *Journal of Food Composition and Analysis*, 102, p. 103879. doi: https://doi.org/10.1016/j.jfca.2021.103879.

Morton, G. J., Meek, T. H. and Schwartz, M. W. (2014) 'Neurobiology of food intake in health and disease', *Nature reviews. Neuroscience*, 15(6), pp. 367–378. doi: 10.1038/nrn3745.

Must, A., Dallal, G. E. and Dietz, W. H. (1991) 'Reference data for obesity: 85th and 95th percentiles of body mass index (wt/ht2) and triceps skinfold thickness.', *The American journal of clinical nutrition*. United States, 53(4), pp. 839–846. doi: 10.1093/ajcn/53.4.839.

National Cancer Institute (2009) *Usual dietary intakes: SAS macros for the NCI method*. National Cancer Institute. Available at: https://epi.grants.cancer.gov/diet/usualintakes/macros.html (Accessed: 1 October 2019).

Ravelli, M. N. and Schoeller, D. A. (2020) 'Traditional Self-Reported Dietary Instruments Are Prone to Inaccuracies and New Approaches Are Needed', *Frontiers in nutrition*. Frontiers Media S.A., 7, p. 90. doi: 10.3389/fnut.2020.00090.

Romieu, I. *et al.* (2017) 'Energy balance and obesity: what are the main drivers?', *Cancer causes & control : CCC*, 28(3), pp. 247–258. doi: 10.1007/s10552-017-0869-z.

Russell, J., Lechner, A., Hanich, Q., Delisle, A., Campbell, B. and Charlton, K. (2018) 'Assessing food security using household consumption expenditure surveys (HCES): a scoping literature review', *Public Health Nutrition*. Cambridge University Press, 21(12), pp. 2200–2210. doi: 10.1017/S136898001800068X.

Shaheen, N. Rahim, A. M. and Banu, C. Bari, L. Tukun, A. Mannan, M. Bhattacharjee, L.

Stadlmayr, B. (2013) ‘Food composition table for Bangladesh’. Institute of Nutrition and Food Science, Centre for Advanced Research in Sciences, University of Dhaka. Available at: http://www.fao.org/fileadmin/templates/food_composition/documents/FCT_10_2_14_final_version.pdf.

Shook, R. P. *et al.* (2015) ‘Low levels of physical activity are associated with dysregulation of energy intake and fat mass gain over 1 year.’, *The American journal of clinical nutrition*, 102(6), pp. 1332–1338. doi: 10.3945/ajcn.115.115360.

Tooze, J. *et al.* (2010) ‘A mixed-effects model approach for estimating the distribution of usual intake of nutrients: the NCI method.’, *Statistics in medicine*. England, 29(27), pp. 2857–2868. doi: 10.1002/sim.4063.

Tooze, J. A. *et al.* (2006) ‘A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution.’, *Journal of the American Dietetic Association*, 106(10), pp. 1575–1587. doi: 10.1016/j.jada.2006.07.003.

United Nations (2005) *Designing Household Surveys: Practical guidelines. Studies in Methods Series F No.98*. Available at: https://unstats.un.org/unsd/demographic/sources/surveys/handbook23june05.pdf.

WHO (1983) *Measuring Change in Nutritional Status*. Geneva. Available at: https://apps.who.int/iris/bitstream/handle/10665/38768/9241541660.pdf?sequence=1.

WHO (2021a) *BMI-for-age (5-19 years)*. Available at: https://www.who.int/tools/growth-reference-data-for-5to19-years/indicators/bmi-for-age (Accessed: 1 October 2021).

WHO (2021b) *Weight-for-age*. Available at: https://www.who.int/tools/child-growth-standards/standards/weight-for-age (Accessed: 1 October 2019).

WHO (2021c) *Weight-for-length/height*. Available at: https://www.who.int/tools/child-growth-standards/standards/weight-for-length-height (Accessed: 1 October 2019).

World Bank (2021) *Global Consumption Database*. Available at: https://datatopics.worldbank.org/consumption/ (Accessed: 14 October 2021).

Yanetz, R., Carroll, R. J., Dodd, K. W., Subar, A. F., Schatzkin, A. and Freedman, L. S. (2008) ‘Using Biomarker Data to Adjust Estimates of the Distribution of Usual Intakes for Misreporting: Application to Energy Intake in the US Population’, *Journal of the American Dietetic Association*. Elsevier, 108(3), pp. 455–464. doi: 10.1016/j.jada.2007.12.004.

## Supplementary information

### S1 Appendix. Equations to calculate the Minimum Dietary Energy Requirement and the Maximum Dietary Energy Requirement

This Appendix presents the equations used to estimate the Minimum (MDER) and the Maximum (XDER) Dietary Energy Requirements. They are based on the FAO publication "Human and Energy Requirements" (FAO, 2004). Five parameters are used in the equations:

1. Reference Body Weight (RBW)

In general, for the calculation of the MDER, the RBW at the lower bound of normal body weight (for example, BMI 18.5 for adults) is used. Whereas for the calculation of the XDER, RBW on the upper bound of normal body weight (for example, BMI 25 for adults) is used. Table S1.1 shows the RBW used for each age group.

Table S1.1 Reference body weights (RBW) used for the calculation of the MDER and XDER by age group.

| Age group | RBW values for MDER | RBW values for XDER | Reference | Remarks |
|---|---|---|---|---|
| Children <2 years [a] | Weight-for-age z-scores at -2 SDs | Weight-for-length z-scores at 2SD | (WHO, 2021b, 2021c) | For this analysis, we excluded children <2 years of age. |
| Children 2-4.9 years | Weight-for-age z-scores at -2 SDs | Weight-for-height z-scores at 2 SDs | (WHO, 2021c) | |
| School-age children and adolescents 5-19.9 years | BMI-for-age z-scores at -2SD for the MDER | BMI-for-age z-scores at 1SD for the XDER | (WHO, 2021a) | BMI is used to infer the weight in kilograms (kg) for the attained height, using the formula RBW in kg=BMI in kg/cm$^2$*(height in cm/100)$^2$. For this analysis, the height values used come from the 2015 BIHS. |
| Adults (>19.9 years) | BMI 18.5 | BMI 25 | (Must *et al.*, 1991) | BMI was used to infer the weight in kilograms (kg) for the attained height, using the formula RBW in kg=BMI in kg/cm$^2$*(height in cm/100)$^2$. For this analysis, the height values used come from the 2015 BIHS. |

Notes. BIHS, Bangladesh Integrated Household Survey; BMI, body mass index; cm, centimeters; MDER, Minimum Dietary Energy Requirements; SD, standard deviation; XDER, Maximum Dietary Energy Requirements.
[a] In this study we excluded infants with less than 2 years; therefore, their energy requirements were not considered when estimating the MDER and XDER of the population.

2. Weight Gain for age (WG)

WG is a coefficient of the expected weight deposition in children and adolescents (0 to <18 years). The source for the WG values for each age-sex group is the WHO Publication *Measuring Change in Nutritional Status* (WHO, 1983).

3. Energy Requirement per Weight Gain (ERwg)

ERwg is a coefficient of the energy required (in kcals) per gram of weight deposited. It is applied to the equations for children and adolescents (0 to <18 years). The source for the ERwg values is (FAO, 2004).

4. Multiplication Coefficient (MC)

The MC is applied to the equations for adolescents 10 to <18 years. It is 0.85 for MDER and 1.15 for XDER to reduce or increase, respectively, by 15 percent the requirement of groups that are less or more active than average (FAO, 2004).

5. Physical Activity Level (PAL)

Lifestyles are classified in relation to the intensity of usual physical activity. For adolescents and adults > 17.9 years the PAL values used are standard: 1.55 for the MDER and 2.10 for the XDER (FAO, 2004).

The 2015 BIHS collect information on the physiological status (pregnant/lactating) of women in reproductive age; therefore, as regards the additional energy required by pregnant women, they are allotted an additional 281 kcal/day. This is based on FAO (2004), which reports that increased needs in the first, second and third trimesters of pregnancy are, on average, 85, 285, and 475 kcal/day, respectively (or 281 kcal/day across all trimesters)[10]. For lactating women, if breastfed children (0-2 years) are included in the analyses and the children's needs factored in, the increased requirements of lactation are not considered, to avoid doubling counting them. If breastfed children are excluded, lactating women are allotted an additional 547 Kcal/day. This is based on FAO (2004), which recommends that during the first 6 months of lactation women should increase their food intake by 505 kcal/day (if well-nourished women with adequate gestational weight gain) to 675 kcal/day (undernourished women and those with insufficient gestational weight gain), and that from 6 months onwards they should get an additional 460 kcal/day. The 547 Kcal/day figure is an average of the three values proposed by FAO. For the analysis with 2015 BIHS data we considered 58 sex-age groups. We excluded 4 sex-age groups corresponding to children younger than 2 years of age, thus, we considered the additional energy needs to lactating women.

Table S1.2 presents the equations used by FAO Statistics Division to estimate the Minimum and Maximum Energy Requirements.

Table S1.2 Equations used by FAO Statistics Division to estimate the Minimum and Maximum Energy Requirements

| | |
|---|---|
| | **Less than 1 year (Age group 1) [a]** |
| Male and Female | If the country Under 5 Mortality Rate > 10 ‰:<br>TEE = (-99.4 + 88.6*RBW) +2* WG * ERwg |
| Male and Female | If the country Under 5 Mortality Rate <= 10 ‰:<br>TEE = (-99.4 + 88.6*RBW) + WG * ERwg |
| | **1 to 1.9 years (Age group 2) [a, b]** |
| | If the country Under 5 Mortality Rate > 10 ‰: |
| Male | TEE = 0.93 * (310.2 + 63.3*RBW - 0.263*RBW$^2$) + 2 * WG * ERwg |
| Female | TEE = 0.93 * (263.4 + 65.3*RBW - 0.454*RBW$^2$) + 2 * WG * ERwg |

[10] When the survey does not capture information on pregnancy status, the birth ratio is used to estimate the pregnancy allowance—it is calculated as extra energy = Birth Ratio * 210 Kcal/day.

| | |
|---|---|
| | If the country Under 5 Mortality Rate <= 10 ‰: |
| Male | TEE = 0.93 * (310.2 + 63.3*RBW - 0.263*RBW$^2$) + WG * ERwg |
| Female | TEE = 0.93 * (263.4 + 65.3*RBW - 0.454*RBW$^2$) + WG * ERwg |
| **2 to 9.9 years (Age groups 3 to 10)** | |
| Male | TEE = (310.2 + 63.3*RBW - 0.263*RBW$^2$) + WG * ERwg |
| Female | TEE = (263.4 + 65.3*RBW - 0.454*RBW$^2$) + WG * ERwg |
| **10 to 17.9 years (Age groups 11 to 18)** | |
| Male | TEE = MC * (310.2 + 63.3RBW - 0.263*RBW$^2$) + WG * ERwg |
| Female | TEE = MC * (263.4 + 65.3RBW - 0.454*RBW$^2$) + WG * ERwg |
| **18 to 19 years (Age groups 19 to 20)** | |
| Male | TEE = PAL * (692.2 + 15.057*RBW) |
| Female | TEE = PAL * (486.6 + 14.818*RBW) |
| **20 to 29.9 years (Age groups 21 to 22)** | |
| Male | TEE = PAL * (692.2 + 15.057*RBW) |
| Female | TEE = PAL * (486.6 + 14.818*RBW) |
| **30 to 59.9 years (Age groups 23 to 28)** | |
| Male | TEE = PAL * (873.1 + 11.472*RBW) |
| Female | TEE = PAL * (845.6 + 8.126*RBW) |
| **More than 59.9 years (Age groups 29 to 31)** | |
| Male | TEE = PAL * (587.7 + 11.711*RBW) |
| Female | TEE = PAL * (658.5 + 9.082*RBW) |

Notes.
TEE, Total Energy Expenditure in kcal; U5MR, under-five mortality rate; RBW, reference body weight; WG, weight gain for age (g/day); Erwg, energy required per gram of weight gain (kcal); MC, multiplication coefficient; PAL, Physical Activity Level.
[a] The equation to be used depends on the level of infant undernutrition and infection in the country; the under-five mortality rate (U5MR) indicator is used to define this level.
[b] The equation is multiplied by a factor of 0.93 to compensate the fact that the predicted values of total energy expenditure (TEE) were about 7 percent higher than the actual TEE measurements (FAO, 2004).
[c] In this study we excluded infants with less than 2 years; therefore, their energy requirements were not considered when estimating the MDER and XDER of the population.

## S2 Appendix. Characteristics of the 24HR and 7DR analytical samples

Table S2.1 Characteristics of the 2015 Bangladesh Integrated Household Survey (BIHS) data

| | **Analytical sample 24HR Un-weighted n and %** | **Analytical sample 7DR Un-weighted n and %** |
|---|---|---|
| **Sample size** | 21,310 individuals<br>5,424 households | 22,319 individuals<br>5,422 households |
| **Sex, individuals** | | |
| Male | 10,013 (47.08) | 10,452 (47.02) |
| Female | 11,256 (52.92) | 11,778 (52.98) |
| **Age groups, individuals** | | |
| 2-5 years | 1,925 (9.05) | 2,026 (9.11) |
| 6-10 years | 2,718 (12.78) | 2,795 (12.57) |
| 11-18 years | 3,908 (18.37) | 4,127 (18.56) |
| 19-59 years | 10,809 (50.82) | 11,266 (50.68) |
| >=60 years | 1,909 (8.97) | 2,019 (9.08) |
| **Income, households in** | | |
| $1^{st}$ quintile | 1,040 (19.17) | 1,040 (19.18) |
| $2^{nd}$ quintile | 1,058 (19.51) | 1,057 (19.49) |
| $3^{rd}$ quintile | 1,083 (19.97) | 1,082 (19.96) |
| $4^{th}$ quintile | 1,120 (20.65) | 1,121 (20.67) |
| $5^{th}$ quintile | 1,123 (20.70) | 1,122 (20.69) |
| **Region, households in** | | |
| Barisal | 406 (7.49) | 407 (7.51) |
| Chittagong | 945 (17.42) | 945 (17.43) |
| Dhaka | 1,685 (31.07) | 1,682 (31.02) |
| Khulna | 535 (9.86) | 534 (9.85) |
| Rajshahi | 591 (10.90) | 591 (10.90) |
| Rangpur | 544 (10.03) | 544 (10.03) |
| Sylhet | 718 (13.24) | 719 (13.26) |