Pacific Community
Communauté du Pacifique

Processing of food data collected in household income and expenditure surveys in the Pacific region:

Essential guidelines for food data processing

PACIFIC COMMUNITY
GUIDANCE NOTES SERIES

Guidance Note

Processing of food data collected in household income and expenditure surveys in the Pacific region: Essential guidelines for food data processing

This Guidance Note is adapted to the Pacific region and is based on the UNCEAG document "Processing food consumption data from Household Consumption and Expenditure Surveys (HCES)" which was drafted by a team of experts from Statistics Norway, SPC, World Bank and FAO.

N. Troubat, A. Borlizzi

Pacific Community
Communauté du Pacifique

Noumea, New Caledonia
December 2024

Cover picture: SPC

Prepared for publication at SPC's Headquarters,
B.P. D5, 98848, Noumea Cedex, New Caledonia

Table of Contents

List of Acronyms

AE	Adult Equivalent
CAPI	Computer-Assisted Personal Interview
COICOP	Classification Of Individual Consumption According to Purpose
CPI	Consumer Price Index
DEC	Dietary Energy Consumption
FAFH	Food Away From Home
FCDB	Food Composition Data Base
FCT	Food Composition Table
FPI	Food Price Index
GN	Guidance Note
HCES	Household Consumption and Expenditure Survey
HIES	Household Income and Expenditure Survey
IQR	Inter-Quartile Range
ISU	International System of Units
KCAL	Kilocalories
LCU	Local Currency Unit
LSMS	Living Standards Measurement Survey
MDG	Millennium Development Goal
NCT	Nutrient Conversion Table
NSO	National Statistics Office
NSU	Non-Standard Unit
PFCT	Pacific Islands Food Composition Table
PICTs	Pacific Island Countries and Territories
PNDB	Pacific Nutrient database
PPPD	Per person per day
PSU	Primary Sampling Unit
SDG	Sustainable Development Goal
SPC	Pacific Community
UNCEAG	Committee of Experts on Food Security, Agricultural and Rural Statistics of the United Nations
UNSC	United Nations Statistical Commission
WB	World Bank

About this series

The SPC Guidance Notes series is designed to offer information on methodological aspects relevant for Pacific statistics stakeholders in a concise and easily digestible way. The expected audience is Pacific Government Statisticians, Methodologist, and Analysts, who work with survey and census data to undertake economic, poverty, and food security analysis. The guidance notes series is a work in progress with new notes being added over time on methodological aspects, ranging from sampling to HIES methods, and to other topics of interest for methodologists and policy makers in the Pacific region.

1. Introduction and main objectives

This guidance note (GN) is an adaptation to the Pacific region of a more comprehensive document[1] entitled "Processing food consumption data from Household Consumption and Expenditure surveys (HCES[2]): Guidelines for countries collecting data in line with the United Nations Statistical Commission-endorsed guidelines on food data collection in HCES", which were[3] endorsed by the UN Statistical Commission at its 55th session. This GN has been specifically designed for HIES implemented in the Pacific region and which have adopted the standard food consumption modules – both in-house and away from home – developed by the Pacific Community (SPC) and which follow the 2018 WB/FAO guidelines on food data collection. These modules were first tested in 2018 in the Republic of Marshall Islands (RMI) and have since been adopted by most Pacific Island countries and territories (PICTs).

The food consumption modules collect the monetary value and quantity of each food item consumed by each surveyed household over a 7-day period, as well as the different sources the household obtained the food from. These modules are part of a larger survey that captures information on consumption of other goods and services. Considering this, it is therefore important that the preparation of the food data is fully integrated into the overall survey data processing.

The primary goal of this GN is to provide a standard approach to follow when preparing the food data, to transform them into a format that can be used for poverty analysis, and to inform the consumer price index and national accounts, and for understanding food consumption patterns. The main output of this process is a dataset with estimates of **quantity (in grams), dietary energy, and monetary value for every food item consumed by the household, from every source of consumption**.

Background

Until 2018, Household Income and Expenditure Surveys (HIES) implemented in PICTs were mainly capturing food acquisition[4] through a 14-day paper-based diary where households were asked to report on their daily food acquisition from cash purchases, and their daily consumption of food that were own-produced or received as gift. Beyond uses in national accounts and CPI, the collected food data were often not used for further analyses because of shortcomings occurred during the field data collection and data processing. When the food data were used for poverty and nutrition applications, they were processed differently across countries and based on the needs of the users, ending up with conflicting estimates. Data processing was not always standard and transparent.

---

[1] https://unstats.un.org/UNSDWebsite/statcom/session_55/documents/BG-3i-ProcessingFoodComsuptionData-E.pdf

[2] Household consumption and expenditure surveys (HCES) are also referred to as a variety of other names, including household income and expenditure surveys (HIES), household budget surveys (HBS) or living standards measurement surveys (LSMS). These surveys are conducted on a nationally representative sample to characterize important aspects of household socio-economic conditions. In the Pacific region, these surveys are referred to as HIES.

[3] In 2018, the UN Statistical Commission endorsed data collection guidelines (hereafter referred to as the '2018 WB/FAO guidelines') to better capture food consumption in HCES and to improve the quality of statistics used to inform poverty, food security, and nutrition analyses, while maintaining the traditional purpose to derive the weights for use in the rebase of consumer price indices (CPIs) and in the compilation of national accounts. The 2018 WB/FAO guidelines, besides providing recommendations on how to better capture food consumption data, aim to harmonize survey design worldwide to derive indicators that can be compared between countries, and over time. The guidelines also bring up the issue of harmonizing the processing of food data resulting from HCES.

[4] Food data collected in HCES can refer to food either acquired or consumed over a certain reference period. Food acquired corresponds to all the food purchased, own-produced, or received as gift, over the survey reference period. Food consumed correspond to the foods purchased, own-produced or received as agift, and consumed during the reference period. The difference between "acquisition" and "consumption" lies in stocks. In acquisition surveys, the quantities acquired over a certain reference period will be 0 if the households exclusively consumed from its own stocks, while the apparent consumption would be overestimated if the household acquired food to build stocks over the reference period. See more on the difference between acquisition and consumption survey in Conforti et al. (2017).

After 2018 and the adoption of a harmonized survey questionnaire aligned to the 2018 WB/FAO guidelines, food data processing in the Pacific region adopted a standard process (called "survey harmonisation"), which served as the basis for the drafting of the food data processing guidelines endorsed at the 55th session of the UNSC in February 2024. The overall process of survey harmonization in the Pacific is an 8-step process, as depicted in Figure 1. The 8 steps describe how to derive the cleaned food consumption data, from the raw form to transformed data ready for statistical analysis. These steps are referred to as 'food data processing'. At the end of the process, each food item reported in the survey, from every source (purchase, own production, and gifts) and for each household will have estimates of food consumption:

- quantity (in grams);
- monetary value (referred to as 'value'); and
- dietary energy (in kilocalories).

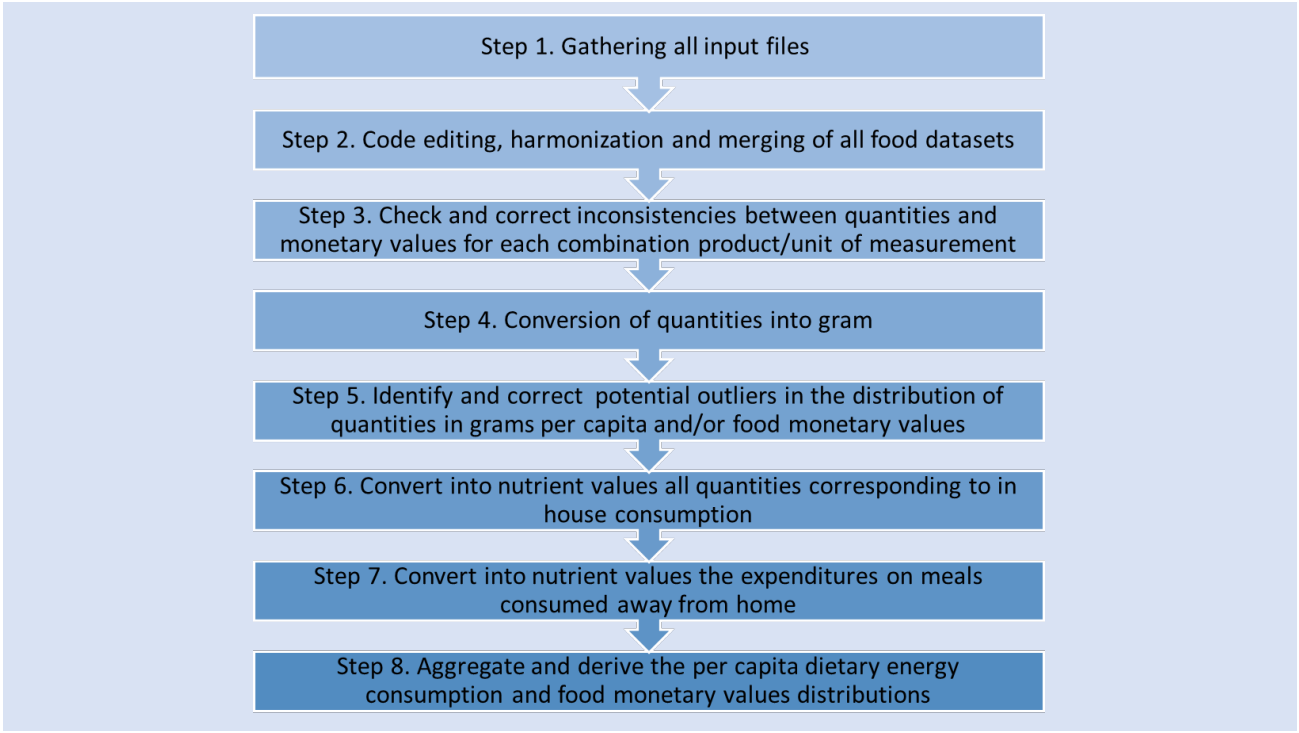


Figure 1: Main steps of food data processing in PICTs

These three variables are needed to inform poverty or undernourishment estimates[5] and can be further used to inform food consumption patterns. This guidance note stop at the dietary energy consumption distribution. Once all the quantities of consumed food items are converted into grams, it is then possible to merge the database with food composition tables used for the estimation of dietary energy consumption.

With a goal is to enable more timely, consistent, and reliable statistics derived from food consumption data, while also improving the quality and transparency of the data processing, this GN provides a series of steps to follow while pointing out issues to be aware of during each phase. The GN is developed on the assumption that the food data were collected following the 2018 WB/FAO guidelines on food data collection. It is therefore assumed here that the food data refers to the amount of food *consumed* by the household over the reference period, and not the amount of food *acquired*.[6] This is an important point to keep in mind.[7]

For effective food data processing, it is recommended that:

[5] Dietary energy serves as basis estimate undernourishment, and the cost of the food consumption baskets used in monetary-based poverty analysis
[6] See the 2018 WB/FAO guidelines for discussion on acquisition versus consumption.
[7] Acquisition surveys need to collect changes in stocks over the reference to calculate consumption. If stocks are not captured, the result may be a very high or low estimates of dietary energy consumption (DEC) at the household level. Bulk acquisitions could also impact the outlier identification process.

- The process be transparent and replicable, documenting each step, including each decision taken, or adjustment made to the data. Analytical programs and all ad-hoc manipulations should be made available to the user.
- While this GN was designed for the use after data collection, it is recommended that national statistics offices (NSOs) also refer to it during survey design phase, including training of field staff. Understanding the steps to follow when preparing the data will improve the collection of high-quality data, because survey stakeholders will understand why the data should be collected in a certain way.

The next sections of this GN are organized as follows: section 2 describes the food data module; section 3 describes the 8 steps of food data processing; and section 4 provides some concluding remarks. The annexes provide more detailed insights into some methodological aspects of food data processing.

2. The harmonized food consumption modules

Harmonization of the HIES in the Pacific region

Household Income and Expenditure Surveys were traditionally used to update the Consumer Price Index (CPI) and National Accounts. The need to monitor global targets (the Millenium Development Goal (MDGs) and now the Sustainable Development Goals (SDGs)) raised interest in HIES for informing poverty and food security indicators. The quality of the food data collected in HIES, however, has been found to be inadequate to properly inform global indicators. An assessment conducted in 2014 on 100 survey questionnaires[8] found that only 9% of the assessed HIES were collecting the information needed to provide reliable estimates of poverty, food security, and nutrition. Following this study, further research was conducted to assess the impact of survey design on food consumption measurement and, in 2018, the guidelines on food data collection were drafted under the umbrella of the UN Committee of Experts on agricultural and food security statistics (UNCEAG) and further endorsed by UNSC (i.e., the 2018 WB/FAO guidelines). The 2018 WB/FAO guidelines provide a set of recommendations on how to best capture food data in HIES. The guidelines were further tested in the Republic of Marshall Islands in 2018 through a survey experiment[9]; it was found that a 7-day recall CAPI survey was providing very similar estimates to those obtained with a well monitored 14-days CAPI diary (benchmark of the experiment). The results of the experiment were presented to the Pacific Statistics Methods Board (PSMB) in May 2019, and the 2018 WB/FAO guidelines were endorsed.

Box 1. The 2018 WB/FAO guidelines recommendations

- Data collection should preferably be done through recall rather than diary.
- Recall period should be 7 days.
- Collect quantities of food consumed, rather than acquired.
- Use predefined list of food items.
- Use of local units of measurement and a market survey to collect weights in gram of one unit.
- Include a module on non-household member meal partakers.
- All sources of consumption should be well identified.[10]

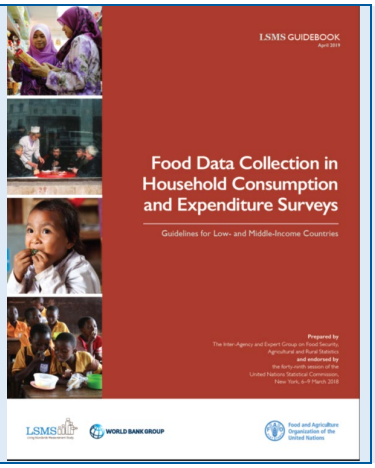
[8] Smith, L.C., Dupriez, O. and Troubat, N., 2014. Assessment of the reliability and relevance of the food data collected in national household consumption and expenditure surveys. *International Household Survey Network*, *82*. (https://www.ihsn.org/sites/default/files/resources/IHSN_WP008_EN.pdf).

[9] In the Republic of Marshall Island, the experiment on food data collection was collected in 5 different ways: 1) a 7-day recall Computer Assisted Personal Interview (CAPI) module with one visit (recommendation of the 2018 WB/FAO Guidelines), 2) a 7-day recall CAPI module with two visits, 3) a 14-day PAPI diary not well monitored (method of data collection in PICTs prior to 2018); 4) a 14-day PAPI diary well monitored; and 5) a 14-day CAPI diary well monitored (benchmark of the experiment).

[10] Households can obtain foods from different sources, and it is important to assess how external shocks may affect sources and consumption. For example, a price hike may have a lower impact on households consuming food from their own production than on households mainly purchasing food. The main sources of consumption are: cash purchases, foods received as gift or as in-kind

The 2018 guidelines also recommend the inclusion of a module on food consumed away from home, which is to:

- Be administered to individually to each household member.
- Use the same recall period as the in-house food consumption module.
- Organized by meal event (e.g., breakfast, lunch, snack, dinner).
- Collect the amount spent to consume the meals.
- Collect information on meals both purchased and received for free.



Following the PSMB recommendation, a standard HIES design, aligned with the 2018 WB/FAO guidelines, was developed by SPC for adoption by PICTs. This HIES design was adopted by most PICTs since 2019 (hereafter referred to as *harmonized HIES*).

The adoption of the harmonized HIES presents many opportunities, as listed below:

- Cost efficiency: it costs 3 times less to conduct a 7-day recall survey with one visit compared toa 14-day diary with 7 visits (which is the way food data were collected by PICTs prior to 2018).[11]
- Timeliness: the development of standard data processing allows to considerably reduce the time between survey development and dissemination of the survey results.
- Comparability of results, between countries and over time.
- Inform national and regional food consumption patterns.[12]
- Knowledge sharing and capacity development because the same module is implemented in all PICTs, which allows for standardization of training materials, sharing of experiences, and reduced cost of training as regional technical workshops targeting more countries can be organized, in place of more expensive and time-consuming national workshops.

The harmonized survey

Compared to the standard food module recommended by the World Bank LSMS group[13], the food consumption modules developed for the Pacific region ask respondents not only to report the amount spent to purchase the food they consumed in the reference period, but also to provide an estimate of the amount they would spend on food they received for free or consumed via their own account production. This form was adopted following the various issues faced in imputing the food monetary values of non-market consumption collected in the Vanuatu 2019/20 National Sustainable Development Plan Survey that adopted the food module recommended by the LSMS group. In addition to collecting the quantity and amount of food consumed from cash purchases, gifts, and home production, where exchange/barter represents an important source of acquisition, consumption quantity and amount from this source is also included.

Following the 2018 FAO/WB guidelines, an individual module on meals consumed away from home was added as a core module to the harmonized HIES (hereafter referred to as FAFH). This information is collected in a different section of the questionnaire and is administered at the individual, rather the household, level. It

payments, foods consumed from own production, and food taken from stocks (only in the case of an acquisition survey). Of note in some PICTs, exchange or barter represent an important source of food acquisition for many households and should hence be included in the food consumption module.

[11] See "Effects of Data Collection Methods on Estimated Household Consumption and Survey Costs Evidence from an Experiment in the Marshall Island", World Bank, 2022. https://openknowledge.worldbank.org/server/api/core/bitstreams/b822d3b7-b904-5b75-93d4-e7ba840516e8/content.

[12] Of note, it is important to remind readers that HIES does not measure dietary intake as recipes and cooking methods are not collected. This is normally done through nutrition surveys.

[13] See. "Capturing what matters – essential guidelines for designing household surveys". WB LSMS guidebook. 2nd edition. 2021. https://www.worldbank.org/en/programs/lsms/publication/CapturingWhatMattersEssentialGuidelinesforDesigningHouseholdSurveys.

collects the information on the number of meals (breakfast, lunch, dinner) consumed away by each household member (either purchased or received for free) and the total amount spent to consume such meals (or the estimated value in case of FAFH meals received as gift).

The main differences between the new survey design based on 7-day recall and the past survey design based on a 14-day diary are further described in Table 1 below.

Table 1. Differences between old and new survey design

	Old design	Harmonised survey
Mode of data capture	PAPI (but recent HIES are using CAPI)	CAPI
Mode of data collection	Diary	Recall
Reference period	14 days	7 days
List of food products	Open (up to 16,000 names of different food products)	Closed (no more than 200 products)
Concepts collected	Acquisition from cash purchases Consumption from non-market sources	Consumption for all sources
Unit of measurement	Not pre-defined: standard and non-standard	Pre-defined: standard and non-standard. (with an ad-hoc category corresponding to "other unit" when the unit is not listed)
Food Away From Home	Collected in the in-house module One respondent on behalf of all household members Only amount spent	Collected in a separate module Individual administration Number of meals and amount spent (estimated amount for gifts) Organized by meal events (breakfast, lunch, dinner, snacks, water, hot drink, non alcoholic beverages)
Seasonality	12 month survey period	12 month survey period

The 7-day recall module on in-house food consumption is responded by one household member, who is usually the person most knowledgeable about food acquisition and preparation for the household. The food products included in the list are representative of the food products available in the country and are classified by food groups. It includes the following questions:

- Filter question on whether the household consumed the specific food product of question in the last 7 days; in the case of a negative answer, the enumerator goes to next food product and, in case of an affirmative answer, the enumerator goes to next question.
- Question on the total quantity of food consumed by the household in the last 7 days, and the corresponding unit of measurement in which the food was obtained.
- Filter question on whether the household purchased some of the food consumed the last 7 days. In the case of negative answer, the enumerator goes to next source of consumption and, in case of affirmative response, the household is asked to report on the quantity consumed from cash purchases, the unit of measurement in which the quantity was purchased[14], and the amount spent to purchase the quantity consumed
- Filter question on whether the household consumed the food product from home production. In case of negative answer, the enumerator goes to the next source of consumption and in case of affirmative answer the household is asked to report on the quantities consumed from home production, the unit of measurement, and to provide an estimate of the amount they would have spent to purchase the food at the market.

[14] In most recent surveys only the unit of measurement corresponding to the total quantity consumed is collected to facilitate crosschecking of quantities reported from all sources of consumption with total quantities consumed.

- Filter question on whether the household received the food as a gift. In case of a negative answer, the enumerator goes to the next question on the total quantity consumed for the next food product (or to the quantity, unit, and amount of consumption if exchange/barter is a source of consumption. In case of an affirmative answer, the household is asked to report on quantity of that food item consumed in the last 7 days as a gift, the unit of measurement, and to provide an estimate of the amount they would have spent to purchase the food at the market.

In countries where food is obtained through barter, additional questions on quantities of food received in exchange of other foods is also asked, as well as the type of product exchange to obtain the food and the estimated amount that would be spent if this food was bought at the market.

The 7-day recall module on meals consumed away from home (FAFH) is filled by each member of the household, with a proxy response for children. This module collects information on number of meals consumed away from home the last 7-days and the amount spent when the meals were purchased, or amount that would have been spent when the meals were received as gift – the collection of home produced items that are consumed away from home is not captured in HIES. The module is organized around meal events such as breakfast, lunch, snack, and dinner, and also collects information on the number and value of hot drinks, non-alcoholic beverages, and bottle of waters consumed.

See Annex 1 for the PAPI version of the in-house and FAFH food consumption modules.

In addition to the in-house and away-from-home food consumption modules, two other modules are included in the Pacific harmonized survey.

- One module collects information on the number of guests (i.e., non-household members who consumed food in the household being surveyed during the survey reference period) and meals they consumed with the household over the last 7 days. This information is needed to estimate the total number of people who consumed food in the household (known as 'partakers') during the recall period. The final number of partakers may not correspond to the household size. For example, one member of the household might be absent during the whole reference period, or the household may have had visitors who ate with them during the reference period. In the former case, the quantity of food consumed per person in that week will be underestimated if the total amount of food consumed is divided by the household size. Similarly, if the household had a visitor, the average amount of food consumed by the household will be overestimated if that visitor is not considered in the total number of partakers. In the HIES conducted in the Pacific region, this information is collected through a standard module asking respondents to report on the number of meals consumed by visitors in the last 7 days. Assuming the household consumes three meals a day (i.e., breakfast, lunch, dinner), the number of partakers is then estimated as:

  Household members present in the household the last 7 days +
  *+ (number of meals consumed by visitors the last 7 days / (3*7))*

- Another module that is administered at a geographical level (e.g., enumeration area, cluster, or census block) is the market survey, which is included in the HIES to collect information on the price and weight in grams of one unit of the most consumed foods in the country. This information is needed later to convert all the quantities collected in non-standard units into grams (see box 2).

Box 2. Unit of measurement

'Standard units' refer to units of measurement included in the International System of Units (ISU), i.e., those units for which there is an established and agreed international weight equivalent in kilograms (e.g., grams, pounds, ounces) and litres. Other units of measurement for weight and volume are considered non-standard units (NSU). In these guidelines, the unit of measurement which all other units are converted into is "gram", because most food composition tables provide the nutrient value per 100 grams of edible quantity of food item. As discussed later, food composition tables are used in the food

data processing when all quantities of food consumption are converted into grams which are then converted into nutrient values to estimate the energy content expressed in calories (kcal), which is the main unit of interest for poverty and food security analysts.
Information about the *gram-equivalent factor* is crucial to convert non-standard units of measurement into grams. Common non-standard units are bag, bunch, heap, piece, and bucket. Most countries in the Pacific have their own system of units. For instance, in Kiribati rice is mainly reported in kilograms, while in Tuvalu it is mainly reported in bags.
Further, the weight in grams for a non-standard unit may fluctuate over time in a given location; it should therefore be collected through a market survey, to be conducted in parallel with each HIES. The data on weight in grams collected through market surveys should be cleaned and aggregated to an appropriate level. Often the mean or median weight in grams for each item/NSU in each region is used (and season if this is relevant).

- The market survey module is standard; the information is collected for the most consumed food products and organized by food groups (cereals, meat, fish, dairy products, fruits, vegetables, non-alcoholic beverages, snacks, spices, take-away food). An extract of the PAPI version of the market survey is shown in Annex 1 (Table A1.c) for the group "bread, grains and cereal products" and using rice and flour as an example.

Development of the Pacific Nutrient Database

In addition to the harmonization of the survey design, another important initiative was undertaken in 2018 by SPC in collaboration with the University of Wollongong and FAO to speed up the food data processing. This initiative consisted in the development of a Nutrient Conversion Table (NCT) linking food products collected in the Pacific HIES to food products from various reference Food Composition Tables and Databases (FCT/FCDB)[15], using the same food classification (based on COICOP 2018[16]). This table is named the Pacific Nutrient Database (PNDB[17]) as it is a mix of various FCTs; it contains, however, only a few nutrients compared to the reference FCT/FCDBs and focuses on food items consumed in the Pacific. In the PNDB, all foods products are coded using the COICOP 2018 classification, as in the HIES. This means that the match between the food data collected in the HIES and the food products included in PNDB is automatic. The development of the PNDB presents an important improvement in food data processing as it considerably speeds up the allocation of nutrient values to the food products collected in HIES (this process is referred to as "food matching"[18]). The time is reduced because of two main factors: in the recall, the list of food items is closed and reduced compared to dairy; secondly, the food match between the PNDB and the HIES is automatic, while previously the match between the available FCTs/FCDBs and HIES was performed manually and could take up to one month.

3. Food data processing

As shown in Figure 2 below, the food data processing consists in 8 main steps:

1) First, gather all the information that will be needed throughout the process: food data from all the different survey modules, information on socioeconomic, demographic and geographic

[15] Food Composition Tables and Food Composition Databases provide information on the amount of macronutrients (protein, fat, carbohydrate, and fiber), micronutrients (such as vitamins, iron, folate, zinc) and minerals (such as calcium, magnesium, potassium) contained in different foods. Most FCTs/FCDBs provide the nutrient composition in 100 grams of edible food (that is, after removing the non-edible part of the food such as skin, bones, peel, seeds). More than 40% of the food profiles of the Pacific Nutrient Data Base (PNDB, see footnote 20) comes from the Pacific Islands Food Composition table (PFCT); other FCTs used are the Australia Food Composition Database, the USDA National Nutrient Database for Standard Reference, and other regional Food Composition Databases. More information on food matching and FCT/FCDB can be found at the FAO INFOODS website: https://www.fao.org/infoods/infoods/en/.

[16] Available at: https://unstats.un.org/unsd/classifications/unsdclassifications/COICOP_2018_-_pre-edited_white_cover_version_-_2018-12-26.pdf.

[17] See: https://sdd.spc.int/digital_library/pacific-nutrient-database-pndb.

[18] Food matching consists in matching the food product collected in the survey with the similar food in the Food Composition table to obtain the nutrient value per 100 grams of product. See section 3 for further details.

characteristics of the household, information on food prices, and information on the nutrient values(which is coming from the PNDB in the case of the Pacific region).

2) Perform a pre-editing of the raw data to check coding, missing values, duplicates and to prepare the working data file in a format convenient for the analysis.
3) For each combination product and unit of measurement, check for inconsistencies between the quantities reported and the corresponding monetary value.
4) Convert all quantities reported into grams.
5) For each food product, perform an outlier detection on the distribution of quantities expressed in gram per capita.
6) Convert all the quantities expressed in grams into dietary energy.
7) Convert all the foods for which a quantity in grams is not available (mainly foods consumed away from home) into dietary energy.
8) For each household, aggregate the total dietary energy and total monetary value and express values in quantity per capita per day.

The process can be performed using any analytical software (e.g., Stata, R, SPSS). For the PICTs, the process is run through a set of syntaxes developed in Stata. Of note, for the process to be transparent and replicable it is recommended to explain and document as much as possible each step of the process.

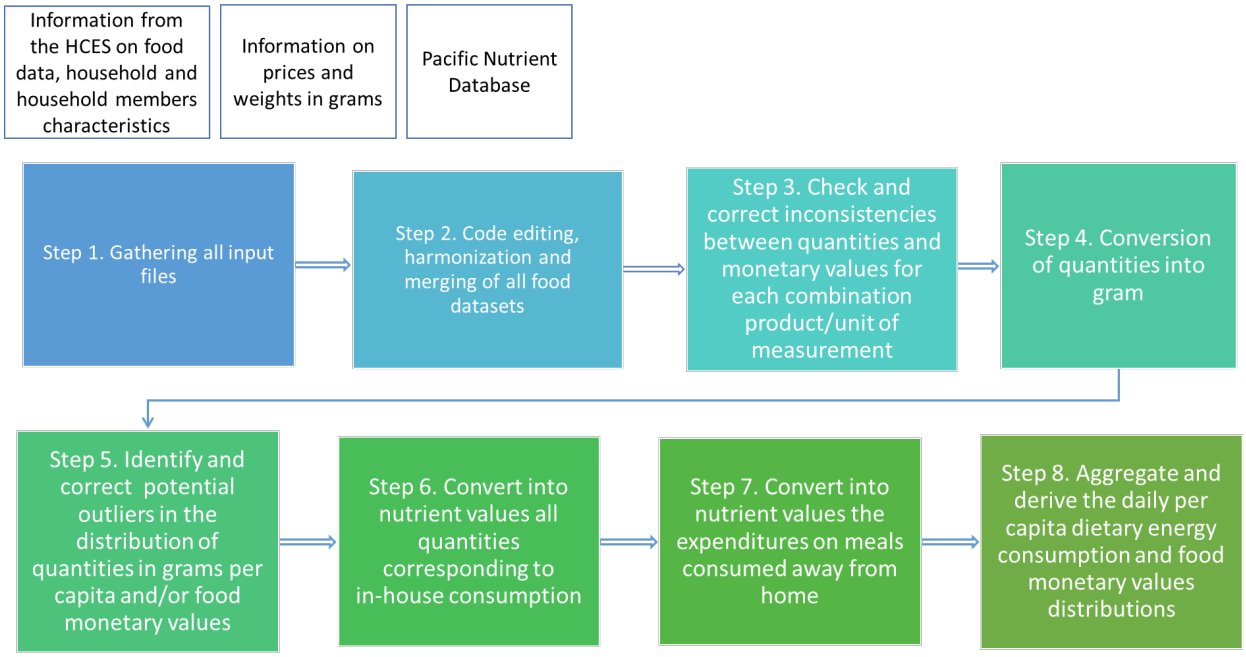


Figure 2: Steps of food data processing in the Pacific region

Step 1. Gathering all the information needed

The first step consists of preparing the working file. All the datasets containing data obtained from the HIES food consumption modules discussed in section 2 need to be assembled, reshaped and edited.

Two types of information are needed: information directly related to the household (collected within the survey) and auxiliary information. The former includes household characteristics such as location, size of the household, sampling weights, and information on food consumption collected in various modules of the survey, including school meals, tobacco and alcohol (see check list in Box 3).[19] Information on the nutrient content of the food is also needed and, in the Pacific region, this is drawn from the PNDB.

[19] Tobacco is not a food but is quite often analysed together with food because of the negative health consequences from its consumption.

Box 3. Check list of data needed

From the HIES

Household level

- ✓ Food consumption quantities, monetary values, and units of measurement corresponding to cash purchases, own production, gifts and exchange/barter (if collected) for each food product as reported in the 7-day in-house food consumption recall module.
- ✓ Household non-food expenditure distribution, or any qualitative information collected in the survey related to the welfare level of the household.
- ✓ Household size (obtained as the sum of the individuals listed in the roster).
- ✓ Number of guests and number of meals consumed with the household in the last 7 days for each of the main meal events (breakfast, lunch and dinner).
- ✓ Number of household members absent during the reference period for food consumption data collection (this information is usually collected in the person roster).
- ✓ Sampling weight, so the statistics can be extrapolated to represent the entire population.
- ✓ Geographic location, such as region, province, district and urban and rural areas, and the Primary Sampling Units.
- ✓ Month and year of each household's interview.

Individual level

- ✓ The number of meals consumed away from home and the corresponding monetary value for each meal event (breakfast, lunch, snack, and dinner) and other FAFH common expenditure items (bottle of water, non-alcoholic drinks, and hot drinks) collected in the individual 7-day recall module on meals consumed away from home.
- ✓ Quantity and monetary value of tobacco and alcohol consumed from cash purchases, own production, and received as gift, as collected in the module on individual expenditures.
- ✓ Amount spent on school meals, usually collected in the individual module on education.
- ✓ Household member ID.*
- ✓ Age and sex of each household member.*

Market level

- ✓ Price per unit of product.
- ✓ Weight in gram for each non-standard unit of product.

External to the HIES

- ✓ Nutrient composition (expressed as macronutrient values per 100 edible grams of food) and refuse factors (available in the Pacific Nutrient Database (PNDB)).**
- ✓ Density*** (gram-equivalent factor of one millilitre of product) for all products measured in volume.
- ✓ Consumer Price Index (CPI) and Food Price Index (FPI) for the period corresponding to the HIES.
- ✓ If a market survey was not conducted, ad-hoc information on weight in grams for the most reported product-unit of measurement combinations and, if not available, at least the price in grams of the most consumed foods. This information is crucial in step 4 of the process.

* This information is needed only to estimate the adult equivalent (AE) that allocates the value of 1 to adults and 0.5 to children. This factor is mainly used for poverty analysis.

** The refuse factor is the non-edible portion of the food, such as bones, peels, skin, seeds, etc.

*** The density is mainly relevant for liquid or semi-liquid products, but can also be of help to identify non-standard units when the volume of the container is known.

Step 2. Code editing, harmonization, merging, and reshaping the files

Raw data and code editing

This step consists in reviewing the data to see if it contains only eligible transactions and if codes appropriately allocated. Data entry error of the unit of measurement for a specific food item (for example, one litre of bread, or one meter of rice) or the wrong use of numerical code of missing values (e.g., 99999) are easy to detect and correct. On the other hand, systematic errors[20,21, 22], such as wrong coding (for example, the code corresponding to "can" is entered for an important number of quantities corresponding to a food like "fresh reef fish") should be identified through visual inspection and corrected. In the glossary of the Generic Statistical Data Editing Model (GSDEM)[23], this step brings the data from 'raw' to 'edited DOS'.

Checks to be performed on the food data.

1. **Check for negative or zero values for quantities and monetary values.**

    (a) If a zero quantity of a food product is associated with a non-zero value of monetary expenditure for that product (and vice versa) it does not explicitly mean a '0' consumption; it should hence be replaced by missing values and imputed at a later stage (Step 4).

2. **Check the filter questions** ('yes' or 'no' to having consumed a food item from a given food group).

    (a) If a household reports not having consumed a food item ('no' in the filter question) but at the same time reports a quantity or food monetary value, then this observation cannot be dropped. It should be further investigated if this is a true or false report. If the enumerator wrongly entered 'no' instead of 'yes', the observation should be kept. If instead there was an error in reporting (for example, the enumerator realised the information was entered for the wrong food item and afterwards changed the 'yes' to 'no'), then the observation should be dropped. These types of errors require further interaction with the enumerator when possible. If not, an *a priori* decision should be made that a non-missing value (if relevant) may prevail to the answer to the filter question.

    (b) Similarly, if a household reports having consumed a food item in the last 7 days ('yes' in the filter question) but both quantities and food monetary values are left empty, it is likely that this is an entry error issue. In such a case, it can be decided to drop this specific occurrence.

3. **Identify duplicates.**

To identify duplicates in in the 7-day recall module, it is recommended to look at the number of observations for each combination household id/product/unit of measurement/source/quantity. These combinations should be unique, however when they are not, they obviously point to duplicates, which should be dropped.

4. **Check the consistency** reported between the total quantity consumed and the sum of quantities reported from cash purchases and/or own production and/or received as gift for each food product.

5. **Check the validity of the codes applied for the food items and the units**

---

[20] A systematic error is an error that is predictable, that is generally constant or proportional to the true value. A systematic error introduces a constant bias in the distribution that needs to be corrected. Systematic errors can occur through a variety of processes, including mistakes in questionnaire design, programming of data entry application, misunderstanding of the questionnaire, data coding or data transformation. The amount of work required to detect and correct systematic errors depends on: a) the mode of data collection, and b) how much cleaning has already been done. See Guidance note on food data cleaning for more details.

[21] Non-systematic errors are more difficult to detect at this stage, especially on large samples. However, the process involves various stage of data cleaning during which it will be possible to detect and correct any non-systematic errors. Non-systematic errors quite often refer to random errors that affect the measurements in unpredictable ways: the measurements are equally likely to be higher or lower than the true values. A random error cannot be predicted and therefore can't be replicated. See Guidance note on food data cleaning for more details.

[22] Here we consider as "systematic" a pattern that is observed on a number of observations that is large enough to determine a cluster.

[23] See: https://statswiki.unece.org/display/sde/2+Introducing+the+GSDEM.

The unit in which the quantity is reported is important information when converting quantities into grams in in Step 4. If we know in which unit the quantity of a food product was procured and the equivalence in grams of that unit, then the conversion of the quantity into grams is straightforward. However, if there are doubts about the unit used to report the quantity, then the conversion of that quantity into grams becomes problematic. That is why it is important at this stage to address all the potential issues related to the unit of measurement and their coding. There are two main types of coding errors:

(a) Invalid combinations of food items and units (for example, one entry for rice in metres). These errors can be easily detected looking at all the combinations product/unit of measurement. In cases where it is obvious that the issue is in the unit of measurement (say, one metre of rice) but it is not possible to allocate a new unit of measurement to the quantities collected, then it is recommended to set the quantities of these specific transactions reported in implausible unit to missing. Quantities will then be converted into grams using monetary values and the price of one gram (see Step 4). In some cases, it may not be possible to detect if the issue is in the food item code or the unit code. For example: some quantities of "fresh reef fish" are reported in "can"; the issue can be that for these specific transactions the food product was not well coded and was referring to "canned fish" instead of "fresh fish". In such case, it is recommended to look at the price per unit of product (for example, the price of one can of "canned fish" can be compared to the price of the quantities of fresh fish reported in can).

(b) Systematic mistakes in coding that affect a significant number of observations, such as mistakes from one team in coding a unit of measurement differently from other teams. This type of error can be identified looking at the number of occurrences for each combination food product/unit of measurement. For instance, salt consumption is mainly reported in "grams" and for some transactions it is reported in "kg", cooking oil is mainly reported in "litres" and for some transactions it is reported in "grams". When it is a reasonable to suspect that this can be an error of an enumerator, one can further look at the product-unit combinations of measurement/enumerator. It is recommended to further investigate with the enumerator to correct the issue.

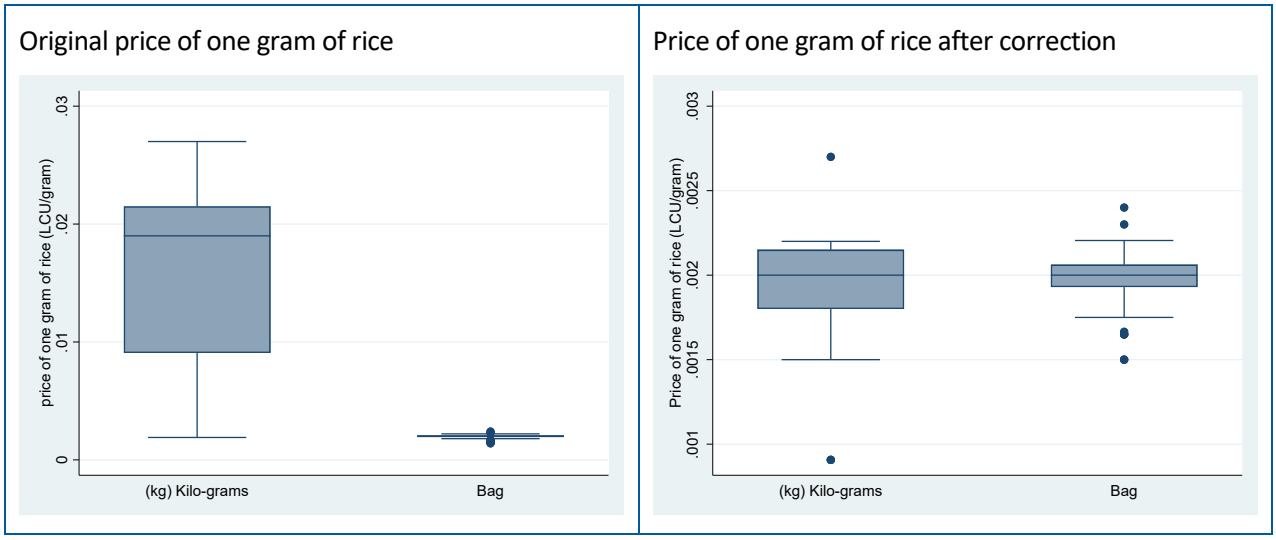
Checks on the market survey:

Before using prices or weights that are collected in the market survey for the analysis, it is important to assess their validity. After all the prices are converted into grams using the weight in standard unit per one unit of product, identification of invalid prices can be performed applying the z-score or interquartile range[24] approach on the price per gram of each product. If the number of observation is large enough, it is recommended to look at the distribution in the price in grams by area of residence (urban/rural) or region. Any value outside an "acceptable range" identified by either of these two methodologies can be either dropped or replaced by the median of the price of that product in that area or region. The median is preferred to other values as we are not interested in the distribution of the price in grams but rather a price to be eventually used in Step 4 of the process when we convert quantities into grams.

When the number of observations is not big enough to perform a robust outlier detection, a graphical approach can be used to compare prices in gram per unit of product. If for some records reported in a certain unit (e.g., one kg) the price per gram departs too much from the price per gram of the other transactions reported in other

[24] The Z-score is defined as the distance in standard deviations away from the mean (Z score = (x - mean) / std. deviation). The Z-score method assumes the variable is normally distributed. An observation is often considered as an outlier if its Z-score is higher than 3 or lower than -3. At this level, the probability of finding a value with a Z-score higher or lower than |3| is less than 0.3% (which corresponds to a significant level of 99.7%). The Tukey approach is another approach to detect outliers and it measures the distance from the median in reference to the interquartile range (IQR = Q3-Q1) where Q1 and Q3 correspond respectively to the lowest value of the first quartile and to the highest value of the third quartile. Observations that lie outside the range [Q1-c*IQR, Q4+c*IQR] where c=1.5 are often considered as outliers. This is equivalent to a Z-score cut-off of +/-3 (that is 3 standard deviation from the mean under the assumption of normal distribution. The interquartile range approach may be preferred to the Z-score as it does not depend on the average of the distribution and is therefore more robust to outliers.

units, then there may be an issue either with the price itself or with the unit in which these records were reported. Figure 3 shows two box plots and the difference in prices per gram before and after the unit corresponding to "kg" was corrected into bag. In the two box plots on the left, the price in gram of one kg of rice is 10 times higher than the price of one bag of 10 kg. The unit seems to correspond to bags rather than to kg; after correction, i.e. after changing the unit from "kg" to "bag" (2 box plots on the right) the prices per gram are aligned.



Source: author's elaboration using data from the 2023 Tuvalu Long Form Census.

Figure 3: Difference in price of one gram of rice after and before correction of the unit

Merging and reshaping all food datasets

After this pre-editing, all datasets with food data are merged and reshaped from *wide form* (one line for each combination household/food product and as many columns as the number of food sources for which quantities and monetary values are collected) to *long form* (each row corresponds to a combination household/food product/source of consumption). See Annex 2 for an example of datasets transformed from wide to long form. The long form is preferred because it reduces the number of variables and facilitates data processing.

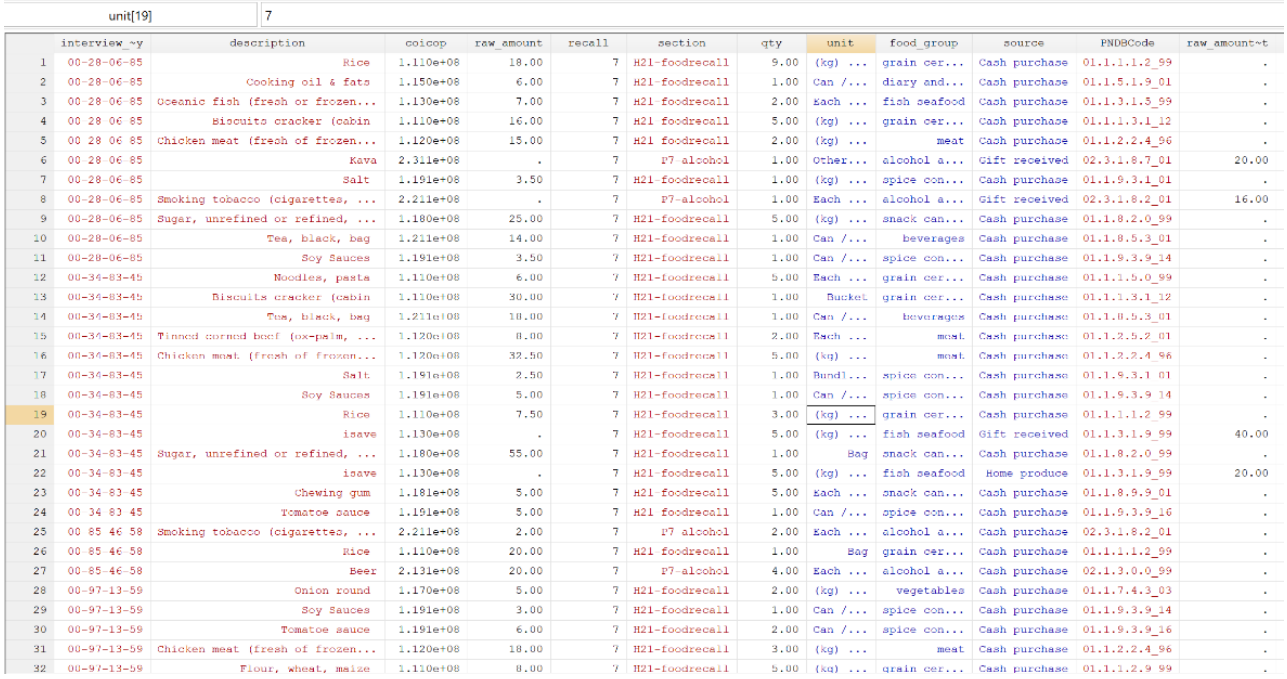
At the end of the merging and reshaping phase, the dataset contains the following information after dropping all non-relevant information:

- Household ID;
- Food item code;
- Food item name;
- Quantity consumed;
- Unit of measurement;
- Source of food (purchased and consumed in the house; own production; received as a gift and consumed in the house; consumed away from home from purchases; consumed away from home received for free; and received in exchange/barter of food, when relevant);
- Monetary value corresponding to the quantity consumed (direct report for cash purchase, and estimates for own production, food received for free, and exchange/barter, when relevant);
- Any other variable related to the food item (e.g., origin, type, place of acquisition).

Once the dataset is reshaped, the code of the food item is matched with the code of the food item in the PNDB. Most PICTs have been using the COICOP 1999 classification and in the PNDB there is a direct match between COICOP 1999 and COICOP 2018; therefore in many cases the food matching is straightforward.

At this stage, it is recommended to add information on the location of the household to the dataset, as this will be needed throughout the process.

Figure 4 shows an example of the structure of the food consumption file after reshaping it to long form.



	unit[19]	7									
	interview_~y	description	coicop	raw_amount	recall	section	qty	unit	food_group	source	PNDBCode	raw_amount~t
1	00-28-06-85	Rice	1.110e+08	18.00	7	H21-foodrecall	9.00	(kg) ...	grain cer...	Cash purchase	01.1.1.1.2_99	.
2	00-28-06-85	Cooking oil & fats	1.150e+08	6.00	7	H21-foodrecall	1.00	Can /...	diary and...	Cash purchase	01.1.5.1.9_01	.
3	00-28-06-85	Oceanic fish (fresh or frozen...	1.130e+08	7.00	7	H21-foodrecall	2.00	Each ...	fish seafood	Cash purchase	01.1.3.1.5_99	.
4	00 28 06 85	Biscuits cracker (cabin	1.110e+08	16.00	7	H21 foodrecall	5.00	(kg) ...	grain cer...	Cash purchase	01.1.1.3.1_12	.
5	00 28 06 85	Chicken meat (fresh of frozen...	1.120e+08	15.00	7	H21 foodrecall	2.00	(kg) ...	meat	Cash purchase	01.1.2.2.4_96	.
6	00 28 06 85	Kava	2.311e+08	.	7	P7 alcohol	1.00	Other...	alcohol a...	Gift received	02.3.1.8.7_01	20.00
7	00-28-06-85	Salt	1.191e+08	3.50	7	H21-foodrecall	1.00	(kg) ...	spice con...	Cash purchase	01.1.9.3.1_01	.
8	00-28-06-85	Smoking tobacco (cigarettes, ...	2.211e+08	.	7	P7-alcohol	1.00	Each ...	alcohol a...	Gift received	02.3.1.8.2_01	16.00
9	00-28-06-85	Sugar, unrefined or refined, ...	1.180e+08	25.00	7	H21-foodrecall	5.00	(kg) ...	snack can...	Cash purchase	01.1.8.2.0_99	.
10	00-28-06-85	Tea, black, bag	1.211e+08	14.00	7	H21-foodrecall	1.00	Can /...	beverages	Cash purchase	01.1.8.5.3_01	.
11	00-28-06-85	Soy Sauces	1.191e+08	3.50	7	H21-foodrecall	1.00	Can /...	spice con...	Cash purchase	01.1.9.3.9_14	.
12	00-34-83-45	Noodles, pasta	1.110e+08	6.00	7	H21-foodrecall	5.00	Each ...	grain cer...	Cash purchase	01.1.1.5.0_99	.
13	00-34-83-45	Biscuits cracker (cabin	1.110e+08	30.00	7	H21-foodrecall	1.00	Bucket	grain cer...	Cash purchase	01.1.1.3.1_12	.
14	00-34-83-45	Tea, black, bag	1.211e+08	10.00	7	H21-foodrecall	1.00	Can /...	beverages	Cash purchase	01.1.8.5.3_01	.
15	00-34-83-45	Tinned corned beef (ox-palm, ...	1.120e+08	8.00	7	H21-foodrecall	2.00	Each ...	meat	Cash purchase	01.1.2.5.2_01	.
16	00-34-83-45	Chicken meat (fresh of frozen...	1.120e+08	32.50	7	H21-foodrecall	5.00	(kg) ...	meat	Cash purchase	01.1.2.2.4_96	.
17	00-34-83-45	Salt	1.191e+08	2.50	7	H21-foodrecall	1.00	Bundl...	spice con...	Cash purchase	01.1.9.3.1_01	.
18	00-34-83-45	Soy Sauces	1.191e+08	5.00	7	H21-foodrecall	1.00	Can /...	spice con...	Cash purchase	01.1.9.3.9_14	.
19	00-34-83-45	Rice	1.110e+08	7.50	7	H21-foodrecall	3.00	(kg) ...	grain cer...	Cash purchase	01.1.1.1.2_99	.
20	00-34-83-45	isave	1.130e+08	.	7	H21-foodrecall	5.00	(kg) ...	fish seafood	Gift received	01.1.3.1.9_99	40.00
21	00-34-83-45	Sugar, unrefined or refined, ...	1.180e+08	55.00	7	H21-foodrecall	1.00	Bag	snack can...	Cash purchase	01.1.8.2.0_99	.
22	00-34-83-45	isave	1.130e+08	.	7	H21-foodrecall	5.00	(kg) ...	fish seafood	Home produce	01.1.3.1.9_99	20.00
23	00-34-83-45	Chewing gum	1.181e+08	5.00	7	H21-foodrecall	5.00	Each ...	snack can...	Cash purchase	01.1.8.9.9_01	.
24	00 34 83 45	Tomatoe sauce	1.191e+08	5.00	7	H21 foodrecall	1.00	Can /...	spice con...	Cash purchase	01.1.9.3.9_16	.
25	00 85 46 58	Smoking tobacco (cigarettes, ...	2.211e+08	2.00	7	P7 alcohol	2.00	Each ...	alcohol a...	Cash purchase	02.3.1.8.2_01	.
26	00-85-46-58	Rice	1.110e+08	20.00	7	H21 foodrecall	1.00	Bag	grain cer...	Cash purchase	01.1.1.1.2_99	.
27	00-85-46-58	Beer	2.131e+08	20.00	7	P7-alcohol	4.00	Each ...	alcohol a...	Cash purchase	02.1.3.0.0_99	.
28	00-97-13-59	Onion round	1.170e+08	5.00	7	H21-foodrecall	2.00	(kg) ...	vegetables	Cash purchase	01.1.7.4.3_03	.
29	00-97-13-59	Soy Sauces	1.191e+08	3.00	7	H21-foodrecall	1.00	Can /...	spice con...	Cash purchase	01.1.9.3.9_14	.
30	00-97-13-59	Tomatoe sauce	1.191e+08	6.00	7	H21-foodrecall	2.00	Can /...	spice con...	Cash purchase	01.1.9.3.9_16	.
31	00-97-13-59	Chicken meat (fresh of frozen...	1.120e+08	18.00	7	H21-foodrecall	3.00	(kg) ...	meat	Cash purchase	01.1.2.2.4_96	.
32	00-97-13-59	Flour, wheat, maize	1.110e+08	8.00	7	H21-foodrecall	5.00	(kg) ...	grain cer...	Cash purchase	01.1.1.2.9_99	.

Figure 4: Structure of the food consumption file in long form

Step 3. Check and correct inconsistencies between quantities and monetary values for each combination of product-unit of measurement

In this step, data editing is done for each combination of food item and unit of measurement to detect and correct inconsistencies between quantities and monetary values. This is recommended to do before quantity data is converted into grams for the following reasons:

- The quantities expressed in their original unit (for instance, bag) do not suffer from systematic errors that can be introduced when transforming the quantities into grams. For example, when a wrong conversion factor is used to convert the quantities from NSU to grams, this introduces an error that would not appear in the non-transformed quantities.
- It is recommended to detect extreme quantities before they are converted into grams and concealed by the other quantities. For example, a household reports having consumed 21 cups of rice; when compared with the overall distribution of rice reported in cups, this food transaction appears as an outlier, but once converted into grams, this quantity might not be detected as an outlier and you risk to carry over a wrong quantity for this specific household[25].
- Prices are likely to differ across units. For example, food items bought in large quantities are likely to be cheaper than the same foods bought in smaller quantities. Thus, unit values may provide better estimates than those at gram level.
- Unit values per unit of measurement can be estimated and used to impute monetary values before food quantities are translated into grams. These monetary values can later be used to impute quantities in grams when conversion factors are missing for some units (which is a common situation). This is further explained in Step 6 of the process.

[25] This is further discussed in the guidance note on food data cleaning. But let's assume the average consumption of rice in a country corresponds to 200 grams per capita per day and the average weight of one cup of rice is 250 grams. A quantity of rice reported by a household of 3 people and corresponding to 21 cups may appear as an outlier in a distribution where the median quantity of cups of rice is 4 per household but once converted into grams this quantity does not appear anymore as an outlier even if it was identified as "extreme" within the distribution of rice in cup.

Outlier detection

As we are mainly interested in looking at inconsistencies between quantities and reported monetary values (e.g., the quantity is too high when compared with the monetary value, or vice versa), a "multivariate"[26] approach, i.e. an approach looking simultaneously at the distribution of quantity, monetary value and unit price, is adopted. The unit price is constructed as: monetary value divided by quantity (expressed in each unit of measurement for each food product). Although some quantities, values or unit prices will be detected as outliers when looked at independently, the decision on whether to correct a quantity and a monetary value will depend on the position of the other variables within their distribution. This process can be described as a 'consistency check' (see table in Annex 3).

For example, a quantity detected as an outlier for a combination food item/unit of measurement will not be corrected if it is associated with a unit price that is not an outlier within its distribution for that combination. In this example, no inconsistency between quantity and monetary value is evidenced and therefore there is no need to correct the quantity. The decision to correct, or not to correct, a quantity or monetary value reported by a household for a specific combination product and unit of measurement, is further explained in the decision matrix shown in Table 2. Annex 3 provides an example of how it works.

If the sample size allows, it is recommended to perform the check for each combination product-unit of measurement by area (e.g., urban/rural), as the quantity or unit value per unit may be different in urban and rural areas (e.g., a heap of tomatoes may be smaller in urban areas than in rural areas, or the value per unit of a heap may also differ). In such case, the median used to replace quantities or monetary values detected as outliers corresponds to the median of the food product expressed in that unit of measurement in this area.

Table 2: Decision matrix, multivariate approach – for each combination product/unit of measurement/area

Monetary value	Unit value	Quantity	Action
Outlier	Outlier	Outlier	First correct food expenditure by using median food expenditure, then correct the quantity using median unit price and median food expenditure
Outlier	Outlier	No	Correct the monetary value using the quantity and the median unit price
No	Outlier	Outlier	Correct the quantity using the monetary value and the median unit price
No	Outlier	No	Correct the quantity when the unit price deviates too much from the median price (we use less than 0.25 times the median as arbitrary threshold) and correct the monetary value when the unit price is more than 4 times the median

For the last case (where the unit value is detected as an outlier), thresholds of 4 and 0.25 deviation from the median were set arbitrarily and are relatively "conservative". It is recommended to be careful in cases the unit value of a product is relatively uniform (no variability) as a small deviation of the unit value from the median unit value is identified as an outlier[27] (this is often observed for rice in some PICTs).

The multivariate approach:

- is relatively conservative and allows to correct only inconsistencies between quantity and amount reported. True outliers will be further detected in Step 5 of the process. The detection of outliers
- is based on the *interquartile range* method and is only valid when the number of observations for each combination "product-unit of measurement" or "product-unit of measurement by area" is larger than 5. Below that number, the detection will fail; potential outliers or inconsistencies will not be detected

[26] This term refers to the fact that three series are looked at simultaneously and should not be confused with the outlier detection methods based on regression that is also based on the use of many variables to detect outliers. Here this approach is called "multivariate" to distinguish with the "univariate" approach that focuses on one distribution only (e.g., consumption in grams).

[27] In a situation where there is no variation in the unit value, the mean, the median, the 25th and 75th percentiles of the distribution are very close and the IQR is close to 0 and therefore any point respectively higher than the 75th percentile or lower than the 25th percentile of the distribution of the unit value is detected as outlier.

at this step will be further detected in Step 5. The level at which outliers are looked at depends on the size of the sample; if the sample is large enough, it is recommended to detect outliers for each product-unit combination in urban and rural areas.

The median of the quantity, monetary value or unit value corresponding to the level at which the outlier detection was performed (that is, "product-unit or measurement" or "product-unit of measurement by area") is used to replace the value when it is detected as an outlier. The median, in fact, is considered a robust estimate of the central tendency. Of note, if the survey is conducted over a cycle of 12 months, it is recommended to deflate the monetary values using the food price index to account for fluctuations in price over time. Also, the detection and correction of inconsistent values are performed without using sampling weights and without using values *per capita*, as size effect is cancelled out in the estimation of the unit value.

Step 4. Conversion of quantities into grams

This step consists of transforming all the quantities reported by respondents into grams. In the HIESs of the Pacific region, less than 50% of the quantities reported are expressed in standard units like grams, kilograms and pounds; the majority of food items are reported in NSUs such as cups, bags, heaps, etc. However, the PNDB provides nutrient values per 100 grams of food. It is therefore important to transform all food consumption quantities into grams prior to using the information provided by the PNDB. Harmonizing all the quantities into single standard units not only allow for the estimation of the nutrients corresponding to the quantity consumed but also to perform further analysis on prices, total quantities of food consumed, etc. The conversion of quantities consumed into grams is a crucial part of the food data processing process, because the wrong use of conversion factors may influence the consumption distribution and therefore the analysis.

To transform all the quantities consumed into grams, it is important to know the weight in grams of one unit of product (for example: in Tuvalu 1 bag of rice weighs 10kg, one loaf of bread in Samoa weighs 545 grams) or the price per gram, when available. Such information can be obtained either from a market survey embedded to the HIES, or from external market surveys implemented by the NSO to inform the CPI, or from ad-hoc sources. In the Pacific, many HIES have a market survey incorporated in their core survey module; in most cases, however, the information collected by them is not broad enough to cover all the possible product-unit combinations reported in the survey. The weight in grams for specific combinations of food items-unit, moreover, may be available for some regions but not for others; in such cases, the national average weight in grams may be used, provided it can be assumed they are the same. In all cases, it is important to consult local experts to validate the weights in grams for each NSU, and to drop entries that are not possible or do not make sense (such as *milk* in baskets*)*. Using a wrong conversion factor, in fact, can create systematic bias in the final dataset; it is therefore important to choose reliable values.

Finding the weight in grams for each combination of food item-NSU can be challenging and time-consuming if a thorough market survey has not been undertaken in parallel with the HIES. When this is the case, priority should be given to finding the weight in grams for all the combinations of food items-NSUs that are commonly reported in the survey. As a rule of thumb, weights in grams should be obtained for all combinations of food items-NSUs which represent more than 30% of the quantities collected for a specific food product.

Figure 4 displays how this works using the case of quantities reported for rice, papaya, and cakes. The example shows the unit in which the product is reported, the weight in gram available, the number of times the product is reported in a specific unit and the number of times the product is reported in total. More than 95% of the quantities of rice in the example below can be converted directly into grams using available weights in grams. The remaining quantities can be safely converted into grams using the price of one gram, the estimate of which is relatively reliable as obtained on the 95% of the quantities collected. On the other hand, quantity of cakes consumed are mainly reported in NSUs for which no weight in gram is available from the market survey. It is therefore important to collect at least the weight in gram of one tray of cake to obtain at least two thirds of quantities of cakes in grams to estimate a relatively reliable price of one gram. In the example in Figure 5, for a more robust estimate of price in grams, it is recommended to also collect the weight in gram of one "each/piece" of cake (Figure 5). Similarly, for papaya, 86% of observations are reported in "each/piece" unit. It is therefore necessary to estimate the average weight in grams of a papaya to estimate a reliable price per gram.

To collect such information it is recommended to consult the NSO, which can either conduct an ad-hoc survey or fill the missing weighs using local expertise.

food description in the hies	unit	weight in grams of one unit (from standard unit or from	Total number of observations for this combination	Total number of times the product is reported	percent
Rice, not further specified	(c) Cups		10	571	2%
Rice, not further specified	(kg) Kilo-grams	1000	460	571	81%
Rice, not further specified	(g) Grams	1	1	571	0%
Rice, not further specified	(ltr) Litres		1	571	0%
Rice, not further specified	Bag	10000	89	571	16%
Rice, not further specified	Bucket		2	571	0%
Rice, not further specified	Bundle / Bunch	10000	6	571	1%
Rice, not further specified	Plate / Bowl		2	571	0%
Cake, not further specified	(kg) Kilo-grams	1000	1	39	3%
Cake, not further specified	Bundle / Bunch /		2	39	5%
Cake, not further specified	Each / Piece		11	39	28%
Cake, not further specified	Plate / Bowl		1	39	3%
Cake, not further specified	Tray		24	39	62%
Papaya	(kg) Kilo-grams	1000	6	76	8%
Papaya	Bundle / Bunch /		3	76	4%
Papaya	Each / Piece		65	76	86%
Papaya	Plate / Bowl		2	76	3%

In the case of rice we have more than 95% of the quantities that can be converted into grams. It will be easy to convert the remaining quantities into grams using either the market price or the price estimated from the survey.

In the case of cakes only 3% of the quantities can be converted into grams. This product is reported by around 40 households so because of its importance we need to get the weight in grams of one tray and one piece of cake as a direct estimate may be more reliable than and indirect estimate based on only 3% of the quantities!

In the case of papaya, 86% of the quantities are reported in each/piece. It is essential to obtain the weight in grams of one piece of papaya to be able to convert all the remianing quantities. With 94% of the quantities in grams it will then be possible to estimate a relatively reliable price of one gram of papaya from the survey

Figure 5. Example of unit conversion table

The file containing the information on the weight in grams for all (or as many as possible) combinations of food items/units of measurement should include the following variables:

- Food product code and name of the product (same as in the HIES).
- Unit of measurement (same as in the HIES).
- Weight in grams per NSU for each valid combination of food item-NSU reported in the survey.
- Price in local currency units (LCU) of one unit for each valid combination food item-NSU (if collected).
- Geographical level (for example, state/province, urban/rural) to be matched with geographic variables in the HIES data.

To transform the quantities into grams certain principles need to be adopted, including:

- Weight in grams of one unit of product should be prioritised when available and reliable.
- Median prices should be preferred to average as they represent a more robust estimate of the central tendency.
- Quantities converted using market prices should not be used to estimate the median price from the survey.
- Only quantities reported in standard unit or converted into grams using relevant weights in grams should be used to estimate the median price per one gram of product. And the median price per one gram of product should be estimated only when more than 60% of the quantities of that product are available in grams.
- If, at the end of the process, some quantities cannot be converted into grams, these cases should not be dropped; the remaining quantities will then be estimated using prices based on direct and indirect transformation of quantities. If quantities still cannot be converted into grams, a weight in gram can be obtained from secondary sources such as internet, market surveys, expert knowledge, data from neighbouring countries, trade data etc.

The conversion of quantities into grams follows various steps and decision tree that are further illustrated in Figure 6.

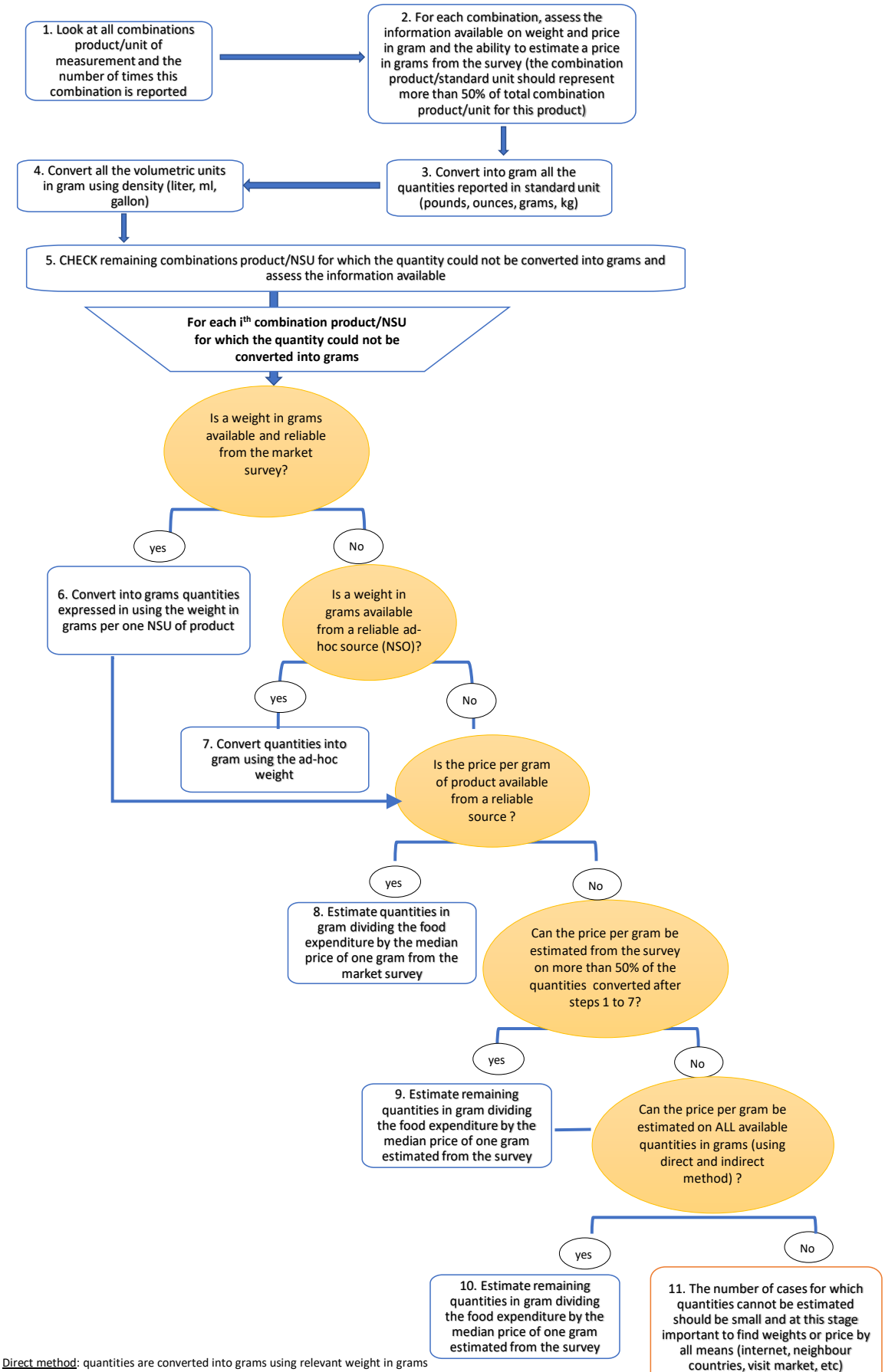
- Steps 1 & 2 – Prepare a table with all the combinations of product and unit of measurement and the number of times each combination is reported, and look at information on weight or price in grams available for all combinations. Prioritize the combinations for which a weight or a price in gram will be needed. For each product, priority should be given to the combination product-NSU representing more

than one third of all the quantities reported for this product. For these combinations, if a weight or price in gram is not available from a reliable source (such as a market survey), it is important to consult the NSO to obtain the weight or price in gram. The assessment of the information that will be needed to transform all the quantities into gram is crucial in the process.

- Step 3 – This step consists of simply converting into grams all the quantities reported in a standard unit of measurement for which there is a common and agreed standard system of conversion (such as grams, kilograms, pounds, etc.).
- Step 4 – At this stage, all the quantities reported in standard volumetric units are converted into grams using a density factor, which corresponds to the weight in grams of one unit of volume of the product. Some tables on density can be found in INFOODS website.[28] For example, 1 litre of water weighs 1 kg and therefore the density factor of water is 1, while one litre of cooking oil weighs 920 grams and the density factor of cooking oil is 0.92.
- Step 5 – Take stock of all the quantities collected in non-standard unit of measurement that cannot be converted into grams;[29] for all these combinations product-NSU, the information available is assessed and steps 6 to 11 are iterative and executed until all quantities are transformed into grams.
- Step 6 to 8 – For each combination product-NSU, we look first at the availability of a reliable weight in gram from a market survey for this specific combination; if available, we convert the quantities into grams using that weight and this information will be further used to estimate the price of one gram of product from the survey. If a reliable weight is not available, then we look at the availability of weights from other reliable sources such as from the NSO or ad-hoc surveys. We use this weight to convert the quantities into grams.
- Step 9 to 11 – In this iterative process, all the quantities that could not be converted into grams using a weight in gram from a valid source (direct approach) will be converted using the price of one gram (indirect approach) either collected from a reliable source or estimated from the quantities and monetary values reported in the HIES. Steps 9 to 11 iterate until all quantities are converted into grams.

[28] For density factors refer to: https://www.fao.org/infoods/infoods/tables-and-databases/faoinfoods-databases/en/.

[29] Excluding the meals consumed away from home and all prepared foods, for which only a monetary value is reported.



Direct method: quantities are converted into grams using relevant weight in grams
Indirect method: quantities are converted into grams using the price per one gram of product

Figure 6. Decision tree to convert quantities into grams

- If the price of one gram of product is available from a reliable source, we use this price (in Local Currency Unit (LCU)) to convert the quantity of food item i consumed by a generic household h into gram as follows:

$$Q(g)_{ih} = \frac{Value\ (LCU)_{ih}}{Price\ (LCU\ per\ g)_i}$$

If the price per gram is not available from the market survey, then we use the price of one gram estimated from the survey. In doing so, it is important to comply with the following rules:

- The price in grams of each product should be estimated only using quantities either reported directly in standard units or converted into grams using valid weights in grams.
- The price per gram of each product should be used only if estimated on at least 60% of quantities converted into grams using the direct approach (that is using quantities that were already expressed in standard unit or that were converted into grams using the weights in grams per unit from a valid source).[30]
- The price is estimated for each product and each household; the median of this price is used as robust estimate of the central tendency. The median should be estimated at the lowest representative geographical level and based on at least 5 observations[31].

The quantities are then converted into gram as follows:

$$Unit\ value\ (LCU\ per\ g)_{ihs} = Value\ (LCU)_{ihs} / Q\ (g)_{ihs}$$

Where h stands for households reporting quantities of product i in the unit s for which a weight in gram was available. The final conversion of quantities from NSU into grams is then performed as:

$$Q\ (g)_{ih(NSU)} = \frac{Value\ (LCU)_{ih(NSU)}}{Median\ Unit\ Value\ (LCU\ per\ g)_i}$$

Where NSU represents the units for which the quantity in grams could not be converted.

- Steps 10 & 11. If some quantities are yet to be converted into grams, there are three options:
  - Relax assumptions 1 to 3 to get a price per gram of product.
  - Conduct an ad-hoc survey to obtain the weight in grams of one unit of product.
  - Look at secondary sources (such as internet or data from neighbouring countries) for the weight in grams per unit.

At the end of this step, all the quantities corresponding to well-defined foods should already be converted into grams. The only two cases where quantities have not been converted into grams are:

1. Food items referring to a food group category (such as, 'other vegetables', which may include cabbages, lettuce and tomatoes, for example) for which there is no unit of measurement but for which a nutrient value exists in the PNDB corresponding to a similar food group (for instance the FAO-GIFT group "vegetables and their products").[32]

[30] If a large number of quantities could not be converted into grams, the price estimate based on unit values may not be reliable. There is no defined standard on the minimum number of observations; as a rule of thumb, if the food item is reported by more than 10 households and at least 60 to 70% of the quantities of that food item could be converted into grams, then the estimated unit value of one gram can be considered as relatively reliable.

[31] This threshold was set at a low value to account for the small samples of HIES conducted in the Pacific region. But it is recommended to increase this threshold to more than 10 when the sample is large enough.

[32] The food products were grouped according to FAO nutrition experts who developed the GIFT platform http://www.fao.org/gift-individual-food-consumption/data-and-indicator/en/ developed from the FoodEx2 classification.

2. Food products referring to 'take-away food consumed in the house' or 'prepared meals procured and consumed away from home' that cannot be quantified.

In the first case, when the food product refers to a food group category, quantities in grams may be imputed as follows:

- Assign the food product to a food category (for instance, the food classification used would allocate "other vegetables" to the food group "Vegetables and products"). In the case of PICTs, the FAO-GIFT classification is used for food consumption analyses and the COICOP classification for poverty analyses.
- Calculate the household unit value of one gram for that food category *j* by dividing the total monetary value spent on those products by total quantity. For example, all the foods that belong to the category 'vegetables' will be used to estimate the missing quantities in grams from the foods reported as 'other vegetables'.

$$HH\ unit\ value\ (LCU\ per\ gram)_{hj} = \frac{\sum_{ihj} Value\ (LCU)}{\sum_{ihj} Quantity\ (grams)}$$

where h refers to the household, j refers to the food category (say, 'vegetables') and i refers to the food items belonging to category j (say, 'lettuce', 'cabbage', 'tomatoes', etc.)

- Calculate the median household unit value per gram by food category using the most appropriate level of geographic disaggregation.
- Impute missing quantities in grams by dividing the monetary value of the food with the missing quantity by the median unit value per gram of the relative food category.

$$Quantity(gram)_{kh} = Value_{kh}\ (LCU) * \left(Median\ unit\ value\ (LCU\ per\ gram)_j\right)$$

where k refers to the unspecified food item for which quantities in grams are missing (say, 'other vegetables') and h refers to the household.

In the second case (take-away food and food away from home) it is not be possible to impute quantities in grams, and dietary energy will be directly imputed in Step 7 of the overall food data processing.

The approach discussed above is based on the assumption that a direct estimate (using weights from reliable sources) is always preferable to indirect estimates, obtained combining monetary value and price. If, for some transactions, the food monetary value is missing, this has to be imputed using the median value of one unit of product. As the standard HIES collects food monetary value for all sources of consumption, usually only a marginal number of transactions is reported with no food expenditure.

Step 5. Detect and correct outliers in the distribution of quantity per capita and monetary value

The cleaning of extreme values is performed for each product on the overall distribution of the quantities in grams expressed in per capita using the Inter-Quartile Range (IQR) approach. It is recommended to perform the outlier detection for each combination food product-area of residence or even better food product-region should the sample size allow for such level of disaggregation.

The current approach is based on univariate approach only and looks at the distribution of quantities in grams for each food item reported (and each area of residence or region if the size of the sample allows). Outlier quantities are corrected using the median of the quantity in grams for that food product at the geographic level at which the outlier detection was performed (i.e., area or region). The monetary value corresponding to the

FoodEx2 is a comprehensive food classification and description system aimed at covering the need to describe food in data collections across different food safety domains.
https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/sp.efsa.2015.EN-804.

quantity flagged as an outlier is also corrected to ensure consistency between quantity and monetary value, by using the median quantity and the median unit price in grams.

It is worth noting that data is edited at *per capita* level; accordingly, corrected food monetary value per capita and quantities per capita need to be multiplied by the number of partakers (or the household size as a proxy) to correct the consumption variables at the household level.

Step 6: Convert all quantities of in-house food consumption into dietary energy (kcal)

At this stage of the data cleaning process, the dataset includes cleaned values for both quantities in grams and monetary values, for all households and each food item. Imputations of originally missing or flagged variables have also been done to the extent possible. Step 6 describes how to convert the quantities in grams into dietary energy[33] (expressed in kcal) for all food items for which the PNDB provide the nutrient content[34].

1. First of all, we merge the Food file (the HIES dataset containing the cleaned food consumption estimates in grams) with the information contained in the Pacific Nutrient Data Base (PNDB) using the 2018 COICOP code of the food item reported in the HIES as merging variable, so that the match between PNDB and food dataset be unambiguous.
2. For each food item, we remove the non-edible portion. Indeed, the quantity of food in the dataset is reported the way it was bought, taken from own production, or received as a gift. However, not the whole food acquired is necessarily edible. Some parts of it, such as bones, peels, and seeds, need to be removed prior to consumption. To obtain the amount of dietary energy available for consumption (DEC) by the household, the part of food item *i* that is not edible (i.e., the refuse factor) needs to be removed from the total quantity consumed by the household *h*.

$$Edible\ Q\ (g)_{hi} = Reported\ Q\ (g)_{hi} * \left(1 - \frac{Refuse_i}{100}\right)$$

3. We then calculate the dietary energy available for consumption. The PNDB provides the kcal per 100 edible grams of food item, therefore the kcal consumed by household h from product i is estimated as follows:

$$Dietary\ Energy\ (kcal)_{hi} = \ Edible\ Q\ (g)_{hi} * \frac{Kcal_i}{100}$$

For analysts interested in deeper analyses of macro- (proteins, fats, carbohydrates…) and micro-nutrients (vitamins, calcium, zinc, etc.), a similar formula applies to all nutrients.

Step 7. Convert dietary energy all the remaining transactions without dietary energy (kcal)

The remaining records without dietary energy estimates are food items for which quantities could not be converted into grams (see in Step 4) and for which no nutrient value can be found in available Food Composition Tables.

For these cases, the dietary energy (in kcal) will be imputed using the monetary value and an estimate of the price per calorie. As mentioned in Step 4, these cases mainly refer to 'take away foods consumed in the house' and 'prepared meals procured and consumed away from home', which will be further referred to 'unspecified' foods for ease of reading.

---

[33] Food is the main source of dietary energy needed by the human body to function. Energy is coming from the macro nutrients (proteins, fats, carbohydrates, and fiber) that are contained in the food. Energy can be expressed in kilojoules or in kilocalories (kcal).

[34] PNDB has been developed to cover, to a large extent, most food products consumed in the PICTs. However, in case some food products cannot be found in the PNDB, then their nutrient content needs to be completed using other Food Composition Tables, starting from the Australian or the New Zealand ones.

For these foods, the dietary energy is estimated by applying the cost of one calorie (referred to as 'dietary energy unit cost') to the monetary value of the 'unspecified' foods. The dietary energy unit cost of unspecified foods is proxied by a dietary energy unit cost of products consumed in-house (referred to as 'well-defined foods') and for which a dietary energy value is available. This approach assumes that the cost of one calorie from unspecified foods consumed away from home is similar to the cost of one calorie from a basket of well-defined foods consumed at home.

Estimation of the dietary energy unit cost

The dietary energy unit cost is calculated dividing the total monetary values from well-defined foods by the total calories consumed from the same foods. In other words, it is the weighted average of the dietary energy cost of each food consumed at home, weighted by its respective share of dietary energy in the total dietary energy consumed.

$$Dietary\ energy\ unit\ cost\ (LCU\ per\ one\ kcal)_h = \frac{\sum_i Value(LCU)_{hi}}{\sum_i DEC(kcal)_{hi}}$$

Where h refers to the household, while i refers to all the food products consumed in the house and reported by household *h* that were converted into kcal and DEC refers to the dietary energy from food i, expressed in kcal.

When estimating the dietary energy cost, and if the sample allows, it may be advisable to consider only purchased food products. This is because purchased food may be more representative of the composition of meals consumed in food establishments than food consumed from own production. It is recommended as well not to include entries that have been imputed at some stage of the process, like observations imputed after being detected as outliers or missing.

The following points describes the dietary energy cost imputation process; additional explanations are provided in annex 4.

- Calculate the household dietary energy unit cost by dividing the total monetary value by total calories of all the foods consumed in the house and for which an estimate of kcal is available
- Calculate the median dietary energy unit cost for that food category at the most appropriate level of disaggregation. In the case of PICTs and based on the size of the sample, two levels of aggregations are used: one that corresponds to each combination expenditure quintile and area of residence, and the other that corresponds to only area of residence.
- Impute the missing dietary energy by dividing the food monetary values for the unspecified food by the median dietary energy unit cost

$$Dietary\ Energy\ (kcal)_{hj} = \frac{Value\ (LCU)_{hj}}{Median\ dietary\ Energy\ unit\ cost\ (LCU\ per\ one\ kcal)_k}$$

Where j corresponds to the 'unspecified food', h refers to household and k refers to the level of aggregation at which the median dietary energy cost was estimated (quintile/area).

Adjustments

The dietary energy cost approach to estimate the dietary energy consumption available from composite meals and food consumed away from home assumes that the cost of one calorie consumed in the house is equal to the cost of one calorie consumed away from home. To account for the additional costs to run a food business, a multiplier can be used to adjust the median dietary energy unit cost. In the case of the PICTs, an adjustment factor ranging from 1.1 to 1.25[35] is added to the dietary energy cost to account for the difference in the dietary energy cost between in-house consumption and away from home consumption.

[35] The multiplier of 1.25 was estimated by the World Bank in an old study based on the 2009 Papua New Guinea HIES.

Meals consumed away from home by meal event or type of meal

In the harmonized HIES, the module on meals consumed away from home is organized by meal events, namely breakfast, lunch, dinner, snacks, hot beverages, non-alcoholic drinks, and water. In this case, the median dietary energy unit cost estimated on all food products consumed in-house is used only to estimate the dietary energy from breakfast, lunch, and dinner. For snacks and non-alcoholic beverages consumed away from home, a dietary energy cost is estimated separately, using respectively only foods corresponding to snack or non-alcoholic beverages in the country. For hot drinks, it is assumed that a hot drink consumed away from home weighs on average 250 grams (one medium cup) and that the amount of kcal in 100 grams of hot drinks is, on average, 32.5 kcal (estimated as the average kcal content in 100 grams of most common hot drinks consumed away from home such as coffee with milk, tea, tea with milk, tea with sugar, etc). Water is calorie free, therefore the amount of dietary energy of water consumed away from home is 0.

Inclusion of drinks in meals consumed away from home

It is not always known if drinks, like alcoholic beverages, are included by respondents in the meals consumed away from home, particularly in restaurants. In the case of PICTs, alcoholic beverages are usually included in the estimation of the dietary energy cost of one kcal consumed in-house. Alcohol, in fact, is an important contributor of calories.[36]

School meals

In most PICTs, there is no system of free school meals provide to pupils: households give money to children to buy their meals at school. When this information is collected in the "education" section of the HIES, it is important to cross-check this information with what is reported in the module on meals consumed away from home to avoid duplication (both modules collect information at the individual level). If, for the same individual, a meal consumed at school is reported in both modules, then this information needs to be dropped either from the FAFH or the Education module.

Limitations

There is no standard approach to estimate the amount of dietary energy consumed away from home: the approach used mainly depends on the quality of the information collected. To date, the use of the dietary energy cost remains the most widely used approach in the absence of a better option[37]. It is widely acknowledged that further research in this area is needed and it is hoped that the recent experiment conducted in Samoa by SPC in collaboration with the Samoa Bureau of Statistics will contribute to improve current methods to collect food consumed away from home and estimate the relative dietary energy.

Step 8. Aggregation and macro editing

At this stage, the dataset contains the dietary energy and the corresponding food monetary value for each household, food item and source. These values can be aggregated to obtain the total amount of dietary energy consumed and the corresponding total food monetary value for each household.

Derive per capita estimates

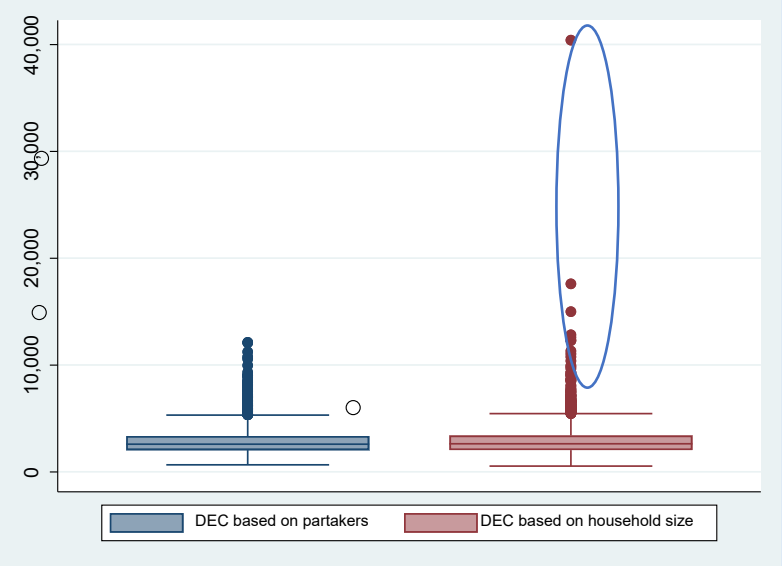
To derive estimates of dietary energy consumption per capita, the total amount of calories consumed by the household should be divided by the number of partakers. As discussed in section 1, the harmonized HIES collects a module on the number of meals consumed by guests; this number is added to the number of household members present in the household the last 7 days and collected in the household member roster.

[36] One gram of alcohol brings 7 kcal. For instance, 10 cl of beer lager with 3.6% alcohol brings 34 kcal and a shot of vodka (5 cl) brings around 111 kcal.

[37] When information on the dietary energy of a meal is known, the kcal per meal approach can be used rather than the kcal approach.

Box 4 shows an example of the impact of not accounting for guests and absent members of the household. The two boxplots show the distribution of daily per capita DEC estimated using, respectively, the household size (on the right) and the number of partakers (on the left). When the household size is used, the distribution of average DEC per capita is more skewed to the right and the average DEC is higher compared to the distribution obtained using the number of partakers.

Box 4. Accounting for partakers

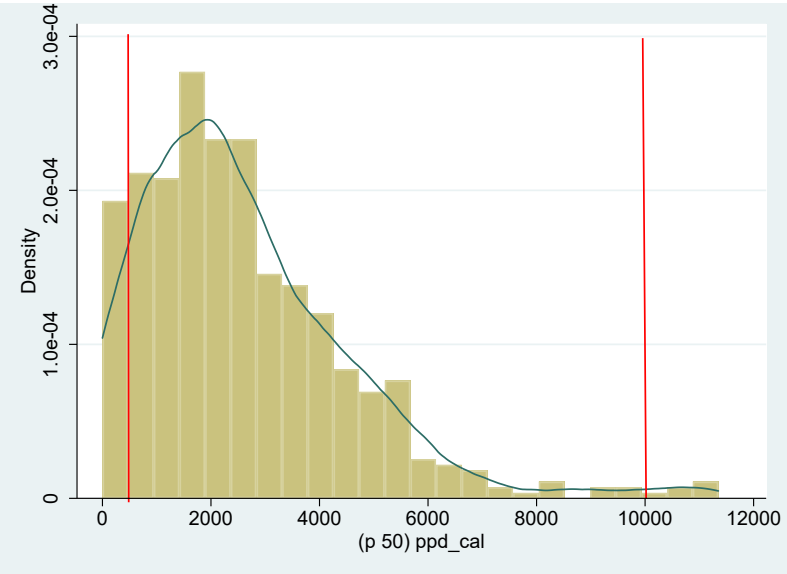


Source: author's elaboration from 2019/20 Kiribati HIES.

Check aggregates at the household level

At this level, the per capita data is aggregated and then divided by 7 (corresponding to the number of days of the reference period of food consumption), so that for each household there is only one record corresponding to the average daily per capita dietary energy consumption (DEC) and food monetary value.

Even if HIES conducted in the PICTs collect consumption (amount of food available to be consumed by the household during the food reference period) and not acquisition (amount of food acquired to be consumed during a period that can go beyond the food reference period), and data refers to apparent consumption and not individual intake. Nonetheless, too high and too low DEC need to be flagged. We recommend using threshold values of DEC below 500 kcal/capita/day and above 10,000 kcal/capita/day. These values are subject to expert judgement and are used only to flag potential issues in the data not identified earlier. Figure 8 plots the distribution of average DEC per person per day with clear issues of under- and over- reporting, corresponding to the observations external to the 2 red lines.



Source: Authors' elaborations using food data collected in 2023 Tuvalu LFC.

Figure 8: Example of distribution of average dietary energy consumption per capita

There are many potential reasons why too high values of DEC are observed at this stage without being discovered earlier in the process, as described below:

- A household reported many food items consumed in quantities slightly lower than the upper fence of the box plot. Therefore, they are not individually detected as outliers but, once aggregated, this leads to an implausibly high value of DEC. For example, one household consumed 1,500 kcal of rice per day per capita, and 1,500 kcal of taro per day per capita, which aggregated sum up to 3,000 kcal per capita per day. When taken individually, these values were not detected as outliers because the upper fence of the box plot was 1,550 kcal per capita per day.
- The estimation of dietary energy consumed away from home, unless checked in the previous step, might have been over-estimated because of the use of a too low cost for one kcal consumed in the household, or too high expenditures reported when consuming in high-end restaurants.
- None of the outlier detection procedures could be applied because the number of observations was too low. For example, if two households in the entire sample reported consumption of caviar, and one of them reported 10 kg.
- Systematic errors in the data, like wrong coding, wrong unit of measurement, or wrong conversion factors in grams of a unit. If Step 2 of the process was duly performed, then such errors should not appear at this stage. If, however, these errors occurred, then the prior editing of the data was probably not efficient and it is important to go back to the raw data.
- A wrong price was used when converting some quantities into grams. For example, a listed food item was a very expensive imported food, while the unit value used to convert the quantity into grams was a local price.
- Some households might have reported quantities of food items that were not meant to be consumed by the household. For example, a proportion of the quantities of coconuts reported was meant to be fed to pigs, or a proportion of quantities of rice was used to feed livestock. If this is known as a common practice in the country, the over reporting of the quantities can be corrected after consulting experts from the country (for example, an average 25% of the acquired rice is given to livestock).
- A significant quantity of food was wasted during storage or cooking.
- Some visitors were not accounted for in the total number of partakers.

Low DEC values, on the other hand, might be due to any of the following reasons:

- Households did not fill in the in-house and/or away-from-home food consumption sections of the questionnaire. These households should have been identified at the first stage of data cleaning and dropped.

- Households only reported the consumption of the first few food items in the recall list and stopped reporting on the consumption of the other food items after realising that saying "yes" to the filter question was leading to too many follow-up questions. These households in theory should have been dropped from the raw data from the beginning.
- Enumerator's fatigue: if the survey was conducted over a period of one year and enumerators were the same throughout the survey period.
- Wrong number of partakers. The total amount of food was divided by a number of people higher than the number of people who actually consumed the food. In such cases, you may need to go back to the questionnaire to check if all household members were truly present in the household during the reference period.
- Wrong food coding. A widely consumed food was associated to a code referring to a food with a low amount of calories while the food is highly energetic (for instance 'noodle' coded as boiled instead of raw)[38].

As a general rule, adjustments should not be performed at the aggregated level, and any gross outlier need to be traced back to determine the cause and adjusted at the item level. It is not recommended to drop values, but rather to go back to the data and identify the food quantities or monetary values, that led to excessive consumption and correct them if the earlier outlier detection procedure did not prove sufficiently efficient.

However, in the case of too low values of DEC, it may be needed to drop some households and to re-calculate the entire household sampling weights. A decision to drop households needs to be part of a consultation process and cannot be taken unilaterally.

4. Concluding remarks

This guidance note aims to address the needs of PICTs to better understand the process behind the measurement of food consumption used to estimate poverty and food security, or to inform national food consumption patterns. It shows the main steps to follow to prepare the food data from their raw form to a form suitable for the analysis. It assumes that the food data are collected using the harmonized food modules developed by SPC for PICTs and which are aligned to the WB/FAO guidelines on food data collection. Some additional checks or steps not included in this note might be needed; in these cases, users interested to know more about the process are invited to consult the UNCEAG guidelines on food data processing (https://unstats.un.org/UNSDWebsite/statcom/session_55/documents/BG-3i-ProcessingFoodComsuptionData-E.pdf).

Finally, this process implies outlier detection methods and approaches that are not discussed in this document, but they have a big impact on survey results. For this reason, they will be part of a separate guidance note.[39]

[38] 100 grams of 'Noodles instant (Maggi type), dry' brings 448 kcal, while 100 grams of 'Noodles instant (Maggi type) boiled, drained' brings 187 kcal.

[39] The expected timeline for the guidance note on food data cleaning is to present its first draft at the 15th PSMB Meeting in May 2025.

References

COICOP (Classification of Individual Consumption According to Purpose) 2018. United Nations Department of Economic and Social Affairs, Statistics Division. Available at: https://unstats.un.org/unsd/classifications/unsdclassifications/COICOP_2018__pre-edited_white_cover_version_-_2018-12-26.pdf

Conforti, P., Troubat, N., Grünberger, G., 2017. The impact of household budget surveys' design on the measurement of food consumption. *FAO Statistics Division Working Paper Series. Forthcoming.*

FAO International Network on Food Data Systems (INFOODS) 2022. Available at: https://www.fao.org/infoods/infoods/en/ (accessed 5 November 2023).

FAO / INFOODS 2012. Databases: Density Database (version 2.0). Available at: https://www.fao.org/3/ap815e/ap815e.pdf

FAO/World Bank 2018. Food Data Collection in Household Consumption and Expenditure Surveys: Guidelines for Low- and Middle-Income Countries. [Abbreviated in this publication to the 'IAEG-AG 2018 guidelines']. Available at: https://www.worldbank.org/en/programs/lsms/publication/Food-data-collection-in-household-consumption-and-expenditure-surveys-guidelines-for-low-and-middle-income-countries

Filzmoser P., Gussenbauer J. and Templ, M. 2016. Detecting outliers in household consumption survey. Vienna University of Technology/World Bank. Available at: http://www.ihsn.org/node/695

Oseni, G., Palacios-Lopez, A., Mugera, H.K. and Durazo, J. (2021). Capturing What Matters: Essential Guidelines for Designing Household Surveys. Washington DC: World Bank. Available at: https://documents1.worldbank.org/curated/en/381751639456530686/pdf/Capturing-What-Matters-Essential-Guidelines-for-Designing-Household-Surveys.pdf

Oseni, G., Durazo, J., & McGee, K. 2017. The Use of Non-Standard Units for the Collection of Food Quantity: A Guidebook for Improving the Measurement of Food Consumption and Agricultural Production in Living Standards Surveys. Washington DC: World Bank. Available at: https://documents1.worldbank.org/curated/en/981711587040434398/pdf/The-Use-of-Non-Standard-Units-for-the-Collection-of-Food-Quantity-A-Guidebook-for-Improving-the-Measurement-of-Food-Consumption-and-Agricultural-Production-in-Living-Standards-Surveys.pdf

Sharp M., Buffière B., Himelein K., Troubat., N. and Gibson J. 2022. Effects of Data Collection Methods on Estimated Household Consumption and Survey Costs: Evidence from an Experiment in the Marshall Islands (English). Policy Research working paper; no. WPS 10029. Washington, DC: World Bank Group. Available at: http://documents.worldbank.org/curated/en/099834004282216747/IDU014657f420d00e042a30a68903563423f5b37

Smith L.C., Dupriez O. and Troubat N. 2014. Assessment of the Reliability and Relevance of the Food Data Collected in National Household Consumption and Expenditure Surveys. IHSN working paper no. 008. Available at: http://www.ihsn.org/sites/default/files/resources/IHSN_WP008_EN.pdf

Smith L.C. and Subandoro A. 2007. Measuring food security using household expenditure surveys. International Food Policy Research Institute (IFPRI): Food security in practice technical guide series 3. Available at: https://ideas.repec.org/p/fpr/fsprac/3.html

Annex 1. Harmonised food consumption modules (PAPI version)

Table A1.a: PAPI version of the in-house food consumption module

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
		TOTAL CONSUMPTION		PURCHASED IN CASH				HOME PRODUCTION				RECEIVED AS GIFT				EXCHANGE (BARTER)				
	In the last 7 days, did any member of this household consume any of the following in the household?	In the last 7 days, how much (quantity) of item] did your household consume?	What is the unit of quantity CONSUMED?	Was any of the quantity of [item] consumed PURCHASED IN CASH? Yes...5 No.....8	How much (quantity) of [item] did you PURSCHASE IN CASH?	What is the unit of quantity PURCHASED in CASH?	How much did your household SPEND for the CASH PURCHASE)	Was any of the quantity of [item] comes from HOME PRODUCTION? Yes .....9 No.....12	How much (quantity) of [item] HOME PRODUCED?	What is the unit of quantity HOME PRODUCED?	How much would your household SPEND (estimated cost) for the quantity HOME PRODUCED	Was any of the quantity of [item] consumed RECEIVED AS GIFT? Yes ..... No....	How much (quantity) of [item] RECEIVED AS GIFT?	What is the unit of quantity RECEIVED AS GIFT?	How much would your household SPEND (estimated cost) for the quantity RECEIVED AS GIFT?	Was any of the quantity consumed of {item] consumed purchased from EXCHANGE of items (barter)?	How did your household pay in EXCHANGE (barter)	How much (quantity) of [item] you PURCHASE through exchange of items?	What is the unit of quantity PURCHASED trhough ECHANGE of items?	How much would your household SPEND (estimated cost) for the quantity of [item] PURCHASED THROUGH
CEREALS & GRAINS	Rice																			
	Flour, wheat, maize																			
	Bread (sliced, loaf, french bread)																			
	Local bread																			
	Crackers																			
	Biscuits cracker (cabin...)																			

Table A1.b: PAPI version of the FAFH module

	In the last 7 days, did [member] consume any [meal event] meals outside the house that were purchased or gifted? **Yes…..** **No….**	**In the last 7 days, how many [meal event] meals did [member] consume away from home?**	**How many of the [MEAL EVENT] meals were pruchased?**	**In total how much (in local currency) was paid for the [meal event] meals**	**How many of the [meal event] meals were provided for free?**	**If [member] were to pay for the free [meal event] meals, what is the amount (local currency) [member] could have paid?**
BREAKFAST						
LUNCH						
DINER						
SNACKS						
HOT DRINKS						
BOTTLED WATER						
NON ALCOHOLIC DRINKS						

Table A1.c: PAPI version of the market survey module (example).

Are there any bread, grains and cereals product present? Yes/No	**Select all the bread, grains and cereal products that are available today?**	**For each bread, grain and cereal products, how many different units are available today?**	**Record the price of one unit of product**	**Record the weight (number) of one unit of product**	**Record the unit in which the weight is reported** **1- gram** **2- kg** **3- pounds**
	Rice	Pounds			
		Once			
		Galion			
		Quarts			
		Pints			
		Cups			
		etc.			
	Flour	Pounds			
		Once			
		etc.			

Annex 2. From wide to long form of the food dataset

Table A2.a. Food dataset structured in wide form

	interview_~y	food_recal~d	h1501ab	h1501ac	h1501ad	h1501b	h1501c	h1501cn	h1501d	h1501e	h1501g	h1501l	h1501m	h1501o	h1501p	h1501q	h15(
1	00-01-...	Rice	Short ...	Island...	011100102	1.00	Other...	201b	Yes	1.00	24.00	.	.	.	.	.	
2	00-01-...	Bread...	Falaoa...	Loaf	011100301	10.00	Loaf		Yes	10.00	20.00	.	.	.	.	.	
3	00-01-...	Cakes	keke s...	keke s...	011100603	1.00	Medium		No	.	.	Yes	1.00	80.00	.	.	
4	00-01-...	Noodl...	Noddle	Chicke...	011100901	10.00	Other...	65g	Yes	10.00	7.00	.	.	.	.	.	
5	00-01-...	Break...	cornfl...	Skippy	011101003	1.00	Other...	200g	Yes	1.00	8.90	.	.	.	.	.	
6	00-01-...	Chick...	chicke...	chicke...	011201306	6.00	Kilo-...		Yes	6.00	25.00	.	.	.	.	.	
7	00-01-...	Sausa...	Sosisi...	chef c...	011201909	5.00	Packe...		Yes	5.00	38.00	.	.	.	.	.	
8	00-01-...	Reef ...	Fresh ...	fresh ...	011302297	1.00	Strin...		Yes	1.00	40.00	.	.	.	.	.	
9	00-01-...	Tinne...	Apaele...	Master	011302597	6.00	Other...	425g	Yes	6.00	27.00	.	.	.	.	.	
10	00-01-...	Dairy...	full c...	Anchor	011402603	15.00	Litres		Yes	15.00	87.00	.	.	.	.	.	
11	00-01-...	Eggs	chicke...	Chicke...	011403501	3.00	Dozen		Yes	3.00	23.40	.	.	.	.	.	
12	00-01-...	Marga...	Margarine	sun ch...	011503899	1.00	Other...	500g	Yes	1.00	7.50	.	.	.	.	.	
13	00-01-...	Cooki...	cookin...	Fielde...	011504004	1.00	Litres		Yes	1.00	8.90	.	.	.	.	.	
14	00-01-...	Brown...	popo l...	popo l...	011604210	10.00	Each ...		No	.	.	Yes	10.00	20.00	.	.	
15	00-01-...	Other...	Vi	Green ...	011604221	15.00	Each ...		No	.	.	Yes	15.00	15.00	.	.	
16	00-01-...	Pumpk...	maukegi	Maukeg...	011704528	1.00	Each ...		No	.	.	Yes	1.00	10.00	.	.	
17	00-01-...	Chine...	kapisi...	kapsis...	111131121	4.00	Bundl...		Yes	4.00	12.00	.	.	.	.	.	
18	00-01-...	Onion	Brown ...	Brown ...	111131159	1.00	Kilo-...		Yes	1.00	3.50	.	.	.	.	.	
19	00-01-...	Cocoa...	kokopa...	kokppa...	012106901	7.00	Packe...		Yes	7.00	10.50	.	.	.	.	.	
20	00-01-...	Lemon...	Laumoli	Laumol...	012106899	10.00	Cups ...		No	.	.	Yes	10.00	20.00	.	.	
21	00-01-...	Sugar...	Brown ...	Brown ...	011804901	2.00	Kilo-...		Yes	2.00	3.80	.	.	.	.	.	
22	00-01-...	Chine...	Lole s...	Lole s...	011805503	1.00	Packe...		Yes	1.00	7.90	.	.	.	.	.	
23	00-01-...	Ice c...	ice cr...	Vanilla	011805805	5.00	Cones		Yes	5.00	12.50	.	.	.	.	.	
24	00-01-...	Snack...	Samgo	Samgo	011704706	5.00	Other...	20g	Yes	5.00	4.00	.	.	.	.	.	
25	00-01-...	Salt	Sea salt	sea salt	011906001	1.00	Pound...		Yes	1.00	1.00	.	.	.	.	.	
26	00-01-...	Curry...	Rice m...	mix sa...	111126086	3.00	Plate...		Yes	3.00	30.00	.	.	.	.	.	
27	00-01-...	Vaisa...	kokoal...	kokoal...	111131724	5.00	Cups ...		Yes	5.00	12.50	.	.	.	.	.	
28	00-01-...	Meaai...	Faalif...	Faalif...	111126005	2.00	Plate...		Yes	2.00	20.00	.	.	.	.	.	
29	00-01-...	BBQ (...	mix BBQ	Mix BB...	111126065	2.00	Plate...		Yes	2.00	20.00	.	.	.	.	.	
30	00-01-...	Burge...	Hot dog	Hot dogs	111126091	6.00	Each ...		Yes	6.00	21.00	.	.	.	.	.	
31	00-01-...	Bun (...	kekepuaa	kekepu...	111126010	20.00	Each ...		Yes	20.00	40.00	.	.	.	.	.	
32	00-12-...	Rice	short ...	island...	011100102	1.00	Kilo-...		Yes	1.00	3.80	.	.	.	.	.	

Table A2.b. Food dataset reshaped to long form

	interview_~y	section	PNDBCode	fooddescriptioninthehies	h1501c	h1501cn	qty	value	qty_new	source
1	00-01-34-11	Food_recall	01.1.1.1.2_02	Rice, white, uncooked	Other...	20lb	1.00	24.00	9080	in-house cash purchase
2	00-01-34-11	Food_recall	01.1.1.3.1_04	Bread, loaf, all others	Loaf		10.00	20.00	.	in-house cash purchase
3	00-01-34-11	Food_recall	01.1.1.3.9_07	Cake, chocolate, commercial	Medium		1.00	80.00	.	in house home produced
4	00-01-34-11	Food_recall	01.1.1.4.0_03	Breakfast cereal, fruit added e.g. sultana b...	Other...	200g	1.00	8.90	200	in-house cash purchase
5	00-01-34-11	Food_recall	01.1.1.5.0_01	Noodles, instant (Maggi-type), dry	Other...	65g	10.00	7.00	650	in-house cash purchase
6	00-01-34-11	Food_recall	01.1.2.2.4_02	Chicken, quarters	Kilo-...		6.00	25.00	.	in-house cash purchase
7	00-01-34-11	Food_recall	01.1.2.5.1_03	Sausage, chicken, raw	Packe...		5.00	38.00	.	in-house cash purchase
8	00-01-34-11	Food_recall	01.1.3.1.9_98	Fish, reef, not further specified	Strin...		1.00	40.00	.	in-house cash purchase
9	00-01-34-11	Food_recall	01.1.3.3.2_02	Mackerel, canned, not further specified	Other...	425g	6.00	27.00	2550	in-house cash purchase
10	00-01-34-11	Food_recall	01.1.4.2.0_99	Milk, fresh, not further specified	Litres		15.00	87.00	.	in-house cash purchase
11	00-01-34-11	Food_recall	01.1.4.8.1_01	Egg, chicken, fresh	Dozen		3.00	23.40	.	in-house cash purchase
12	00-01-34-11	Food_recall	01.1.5.1.9_02	Oil, vegetable	Litres		1.00	8.90	.	in-house cash purchase
13	00-01-34-11	Food_recall	01.1.5.3.0_99	Margarine, not further specified	Other...	500g	1.00	7.50	500	in-house cash purchase
14	00-01-34-11	Food_recall	01.1.6.1.8_03	Coconut, brown	Each ...		10.00	20.00	.	in house home produced
15	00-01-34-11	Food_recall	01.1.6.1.9_04	Ivi	Each ...		15.00	15.00	.	in house home produced
16	00-01-34-11	Food_recall	01.1.6.7.9_99	Mixed dried fruit, not further specified	Packe...		1.00	7.90	.	in-house cash purchase
17	00-01-34-11	Food_recall	01.1.7.1.2_02	Cabbage, Chinese	Bundl...		4.00	12.00	.	in-house cash purchase
18	00-01-34-11	Food_recall	01.1.7.2.5_01	Pumpkin	Each ...		1.00	10.00	.	in house home produced
19	00-01-34-11	Food_recall	01.1.7.4.3_01	Onion, brown	Kilo-...		1.00	3.50	.	in-house cash purchase
20	00-01-34-11	Food_recall	01.1.7.9.9_06	Chips, taro, commercial	Other...	20g	5.00	4.00	100	in-house cash purchase
21	00-01-34-11	Food_recall	01.1.8.2.0_99	Sugar, not further specified	Kilo-...		2.00	3.80	.	in-house cash purchase
22	00-01-34-11	Food_recall	01.1.8.5.3_01	Cocoa, cocoa powder	Packe...		7.00	10.50	.	in-house cash purchase
23	00-01-34-11	Food_recall	01.1.8.6.0_03	Ice cream, cone or bar	Cones		5.00	12.50	.	in-house cash purchase
24	00-01-34-11	Food_recall	01.1.9.1.2_09	Chicken, grilled/bbq	Plate...		2.00	20.00	.	in-house cash purchase
25	00-01-34-11	Food_recall	01.1.9.3.1_99	Salt, not further specified	Pound...		1.00	1.00	.	in-house cash purchase
26	00-01-34-11	Food_recall	01.2.3.0.9_99	Tea, not further specified	Cups ...		10.00	20.00	.	in house home produced
27	00-01-34-11	Food_recall	11.1.1.1.1_05	Banana, cooked	Plate...		2.00	20.00	.	in-house cash purchase
28	00-01-34-11	Food_recall	11.1.1.1.1_08	Bun, Chinese steam/ keke puaa	Each ...		20.00	40.00	.	in-house cash purchase
29	00-01-34-11	Food_recall	11.1.1.1.1_48	Takeaway, curry, chicken, with rice	Plate...		3.00	30.00	.	in-house cash purchase
30	00-01-34-11	Food_recall	11.1.1.1.1_51	Takeaway, hot dog, bread roll, frankfurt fil...	Each ...		6.00	21.00	.	in-house cash purchase
31	00-01-34-11	Food_recall	11.1.1.1.1_64	Vaisalo/Kokoesi	Cups ...		5.00	12.50	.	in-house cash purchase
32	00-01-34-11	6 FAFH	110.1.1.1.2	Breakfast away from home	.		7.00	10.00	.	FAFH-cash purchase

Annex 3. Multivariate approach to detect and correct outliers – how does it work?

The table below provides an example of how inconsistencies between quantity and monetary values are detected and corrected. Each row corresponds to a different case for one product-unit combination for one household. Each column refers to:

- 1st column: number of the case
- 2nd column: description of the food item
- 3rd column: unit of measurement corresponding to the reported quantity of consumption
- 4th column: quantity of consumption of the food product corresponding to the unit of measurement
- 5th column: median quantity corresponding to the product-unit combination for the entire sample
- 6th column: monetary value corresponding to the quantity of reported consumption for that food item
- 7th column: median monetary value corresponding to the combination product-unit for the entire sample
- 8th column: unit value which is the monetary value divided by the quantity (i.e., column 6 / column 4)
- 9th column: the median unit value corresponding to the combination product-unit for the entire sample
- 10th to 12th column: dummy variables to indicate if the amount, unit value or quantity are identified as outlier
- 13th column: the quantity corrected (or not) based on decision matrix
- 14th column: monetary value corrected (or not), based on the decision matrix
- 15th column: ratio of unit value over the median unit value
- 16th column: documentation on the change (if any) to the quantity reported for that record in the dataset
- 17th column: documentation on the change (if any) to the monetary value reported for that record in the dataset

Case	Description of the food	Unit	Quantity	Median quantity	Monetary value	Median amount	Unit value	Median unit value	Outlier amount	Outlier unit value	Outlier quantity	Quantity corrected	Monetary value corrected	ratio	Quantity is changed	Monetary value is changed
1	Rice	Grams	200	500	1	1.9	0.005	0.004	0	0	1	200	1	1.25	No	No
2	Rice	Kilo-grams	15	2	20	7	1.333	3.800	0	1	1	5	20	0.35	Yes	Yes
3	Coconut, brown	Basket	1	1	3	10	3.400	10.000	1	0	0	1	3	0.34	No	No
4	Milk, powdered	Bag	1	1	30	15	30.000	15.000	1	1	0	1	15	2.00	No	Yes
5	Bread	Grams	11040	800	43	4	0.004	0.005	1	0	1	11040	43	0.78	No	No
6	Flour	Kilo-grams	5	2	15	4	3.000	2.000	0	1	0	5	15	1.50	No	No
7	Cabbage	Unit/ each	12	3	1	1.5	0.058	0.500	0	1	0	1	1	0.12	Yes	No
8	Milk, powdered	Grams	400	300	42	6.3	0.104	0.021	0	1	0	400	8	4.95	No	Yes
9	Oil, cooking	Milliliters	5	400	0.2	2.5	0.04	0.007	1	1	1	359	2.5	5.71	Yes	Yes

Action taken in each case

1 The quantity is an outlier but neither the price nor the amount were detected as outlier - we do not change

2 The quantity and unit value are outlier within their respective distributions - we change the quantity as it is too high for the amount reported

3 The amount is an outlier but neither the quantity nor price were detected as outlier - we do not change

4 The amount and the unit value are both outlier within their own distribution - we change the amount as it is too high compared to the quantity reported

5 Quantity and amout are detected as outliers within their respective distributions but the unit value is not an outlier - we do not change as the quantity seems to be consistent with the amount reported

6 The unit value is an outlier but neither the quantity nor the amount are detected as outlier - the unit value does not depart too much from the median unit value and therefore neither the quantity nor the amount are corrected

7 The unit value is an outlier but neither the quantity nor the amount are detected as outlier - however the unit value is ten times lower than the median unit value which means that the quantity is too high for the amount reported and it is therefore changed assuming the amount is correct

8 The unit value is an outlier but neither the quantity nor the amount are detected as outlier - however the unit value is five times higher than the median unit value which means that the amount is too high for the quantity reported and it is therefore changed assuming the quantity is correct

9 The unit value is an outlier as well as the quantity and the amount - in such a case quantity and amount are changed - the amount is changed using the median amount and the quantity is changed using the corrected amount and the median unit value

Annex 4. Estimation of the dietary energy consumed away from home (DEC-AFH)

Figure below shows an example of how the dietary energy corresponding to foods consumed away from home is estimated.

1. For each household sum up calories and respective values of foods consumed in the house

Household id	Food description	Expenditure quintile	Amount spent on the food	Quantity in gram	Calories
00-12-41-73	Rice, not further specified	5	3.5	3178	10751
00-12-41-73	Crackers, not further specified	5	21.5	5000	21635
00-12-41-73	Doughnut, not further specified	5	4	1816	7293
00-12-41-73	Beef, canned, corned	5	16.8	978	2248
00-12-41-73	Fish, pelagic/ocean, not further spe	5	30	9080	8252
00-12-41-73	Milk, powdered, not further specifie	5	14	400	1606
00-12-41-73	Egg, chicken, fresh	5	8.4	464	517
00-12-41-73	Oil, cooking	5	2.5	920	8280
00-12-41-73	Banana, common e.g. cavendish	5	10	2724	1856
00-12-41-73	Coconut, brown	5	1.4	6832.7	13270
00-12-41-73	Breadfruit	5	8	1816	1544
00-12-41-73	Onion, brown	5	2.4	454	96
00-12-41-73	Sugar, not further specified	5	4.2	3178	12521
00-12-41-73	Salt, iodised	5	1.5	1000	0
00-12-41-73	Lunch away from hom	5	25		
00-12-41-73	Lunch away from hom	5	10		
00-12-41-73	Lunch away from hom	5	25		
00-12-41-73	Dinner away from home	5	6		
00-12-41-73	Dinner away from home	5	6		

2. For each household estimate the dietary energy cost dividing total value by total calories consumed in the house

Household id	Food description	Expenditure quintile	Amount spent on the food	Quantity in gram	Calories	Household cost of one calorie
00-12-41-73	Rice, not further specified	5	3.5	3178	10751	
00-12-41-73	Crackers, not further specified	5	21.5	5000	21635	
00-12-41-73	Doughnut, not further specified	5	4	1816	7293	
00-12-41-73	Beef, canned, corned	5	16.8	978	2248	
00-12-41-73	Fish, pelagic/ocean, not further spe	5	30	9080	8252	
00-12-41-73	Milk, powdered, not further specifie	5	14	400	1606	
00-12-41-73	Egg, chicken, fresh	5	8.4	464	517	0.0016
00-12-41-73	Oil, cooking	5	2.5	920	8280	
00-12-41-73	Banana, common e.g. cavendish	5	10	2724	1856	
00-12-41-73	Coconut, brown	5	1.4	6832.7	13270	
00-12-41-73	Breadfruit	5	8	1816	1544	
00-12-41-73	Onion, brown	5	2.4	454	96	
00-12-41-73	Sugar, not further specified	5	4.2	3178	12521	
00-12-41-73	Salt, iodised	5	1.5	1000	0	
00-12-41-73	Lunch away from hom	5	25			
00-12-41-73	Lunch away from hom	5	10			
00-12-41-73	Lunch away from hom	5	25			
00-12-41-73	Dinner away from home	5	6			
00-12-41-73	Dinner away from home	5	6			

3. Take the median of the dietary energy cost

Household id	Food description	Expenditure quintile	Amount spent on the food	Quantity in gram	Calories	Household cost of one calorie	Median cost of one calorie consumed in the house by area and expenditure quintile
00-12-41-73	Rice, not further specified	5	3.5	3178	10751		
00-12-41-73	Crackers, not further specified	5	21.5	5000	21635		
00-12-41-73	Doughnut, not further specified	5	4	1816	7293		
00-12-41-73	Beef, canned, corned	5	16.8	978	2248		
00-12-41-73	Fish, pelagic/ocean, not further spe	5	30	9080	8252		
00-12-41-73	Milk, powdered, not further specifie	5	14	400	1606		
00-12-41-73	Egg, chicken, fresh	5	8.4	464	517	0.0016	
00-12-41-73	Oil, cooking	5	2.5	920	8280		
00-12-41-73	Banana, common e.g. cavendish	5	10	2724	1856		
00-12-41-73	Coconut, brown	5	1.4	6832.7	13270		0.0019
00-12-41-73	Breadfruit	5	8	1816	1544		
00-12-41-73	Onion, brown	5	2.4	454	96		
00-12-41-73	Sugar, not further specified	5	4.2	3178	12521		
00-12-41-73	Salt, iodised	5	1.5	1000	0		
00-12-41-73	Lunch away from hom	5	25	13471.51	13472		
00-12-41-73	Lunch away from hom	5	10	5388.605	5389		
00-12-41-73	Lunch away from hom	5	25	13471.51	13472		
00-12-41-73	Dinner away from home	5	6	3233.163	3233		
00-12-41-73	Dinner away from home	5	6	3233.163	3233		

4. Use that cost to estimate the dietary energy of meals consumed away from home

Household id	Food description	Expenditure quintile	Amount spent on the food	Quantity in gram	Calories	Household cost of one calorie	Median cost of one calorie consumed in the house by area and expenditure quintile
00-12-41-73	Rice, not further specified	5	3.5	3178	10751		
00-12-41-73	Crackers, not further specified	5	21.5	5000	21635		
00-12-41-73	Doughnut, not further specified	5	4	1816	7293		
00-12-41-73	Beef, canned, corned	5	16.8	978	2248		
00-12-41-73	Fish, pelagic/ocean, not further spe	5	30	9080	8252		
00-12-41-73	Milk, powdered, not further specifi	5	14	400	1606		
00-12-41-73	Egg, chicken, fresh	5	8.4	464	517	0.0016	
00-12-41-73	Oil, cooking	5	2.5	920	8280		
00-12-41-73	Banana, common e.g. cavendish	5	10	2724	1856		
00-12-41-73	Coconut, brown	5	1.4	6832.7	13270		0.0019
00-12-41-73	Breadfruit	5	8	1816	1544		
00-12-41-73	Onion, brown	5	2.4	454	96		
00-12-41-73	Sugar, not further specified	5	4.2	3178	12521		
00-12-41-73	Salt, iodised	5	1.5	1000	0		
00-12-41-73	Lunch away from hom	5	25		13471		
00-12-41-73	Lunch away from hom	5	10		5389		
00-12-41-73	Lunch away from hom	5	25		13471		
00-12-41-73	Dinner away from home	5	6		3233		
00-12-41-73	Dinner away from home	5	6		3233		