Cost of a Healthy Diet (CoHD): Protocol for food price data collection and analysis

# Purpose

The purpose of this protocol is to provide guidance on how to calculate CoHD and related indicators from your own food price data, and also to provide guidance on how to select and/or collect food price data.

# Background

The Cost and Affordability of a Healthy Diet (CoAHD) is used to evaluate the food environment dimensions of food price and affordability (Herforth et al. 2020; 2022). The CoHD was established as a food price index to determine the minimum cost of meeting food-based dietary guidelines (FBDG). Affordability is measured as sufficient to afford that minimum cost, based on the proportion of income that can be realistically set aside for food (vs. non-food needs).

The CoAHD has been adopted by the Food and Agriculture Organization of the United Nations (FAO) for global monitoring. It is reported annually in *The State of Food Security and Nutrition in the World Report* published by the UN food agencies, and in databases curated by FAO[[1]](#footnote-1) and the World Bank[[2]](#footnote-2). The indicators have been adopted by some national governments as well (e.g. Ethiopia, Nigeria, Ghana) for monitoring food access and design and targeting of food system interventions.

The CoHD, and the cost of each food group that makes up a healthy diet, are important aspects of the food environment. While many factors influence food choices at an individual level, such as income, knowledge, time and labor availability, tastes and preferences, and habits, prices are one of the largest food environment influences that bound the choices people are able to make.

A healthy diet is defined by food-based dietary guidelines (FBDG). At global level, the Healthy Diet Basket (HDB) is used as a cost standard that reflects the commonalities among FBDG around the world (Herforth et al., 2022). In countries with quantified national FBDG (such as Ethiopia), the national FBDG are used as a cost standard, to identify the minimum cost of purchasing a diet in the market that meets FBDG.

For global and national monitoring, food prices collected by National Statistics Organizations (NSOs) are used to calculate CoHD. At global scale, food item prices are compiled by the International Comparison Program housed at the World Bank, under the auspices of the UN Statistical Commission.[[3]](#footnote-3) At national scale, NSOs typically collect retail prices at many markets around the country on a frequent basis, usually every month at a minimum, within their routine monitoring of the Consumer Price Index (CPI) for the purpose of tracking inflation.

For some research and programmatic applications, such as those that operate at large zones of influence, CPI item prices collected by the NSO may be the best data source to characterize the local food environment. CPI data have the advantages of being official statistics of the country, and being collected in a consistent and representative way. For other applications, researchers or programmers may wish to collect their own food prices, for example for understanding the food environment in a specific small or rural geographic area of interest.

# Software tools for calculating the Cost of a Healthy Diet

To facilitate the calculation and monitoring of the Cost of a Healthy Diet by national governments, researchers, and practitioners, the Food Prices for Nutrition project developed software tools designed to help convert any set of food price data into the cost per day of a healthy diet, using Excel or Stata. The most recent versions of the software tools are publicly available from Food Prices for Nutrition at: <https://sites.tufts.edu/foodpricesfornutrition/tools/>.

The software tools enable users to enter their own retail price data and automatically calculate the CoHD and the cost of each food group. Users enter price data in the units as sold in markets (e.g. cost per bunch), convert it to standard units (e.g. cost per kg), and match each item to a set of items in a curated database to (a) convert the price to price per kilocalorie and (b) classify each item into a FBDG food group (e.g. starchy staples, animal-source foods, vegetables, fruits, pulses, nuts and seeds). Then CoHD is automatically calculated in either Excel or Stata.

Users will input their data in the Excel Workbook, and proceed with the calculations in either Excel or using the Stata code provided. If users wish to understand the methodology and gain practice using the tools before they have price data, both the Excel Workbook and the Stata code are equipped with instructions and pseudodata. A step-by-step video tutorial is available.[[4]](#footnote-4)

For compatibility with the Excel Workbook or Stata code, each row of data should represent a single price observation for a food item. If users are collecting their own data, the simplest process to collect the data in exactly the format used in the toolkit. A Food Price Data Collection sheet is available in the Excel file to facilitate smooth data entry. If users are using data that has already been collected, they may need to reformat their data to input it into the Excel file.

It is important to clean the food price data and address outlier price observations as a preliminary step before calculating the Cost of a Healthy diet. The tools do not provide detailed guidance on data cleaning or outlier identification, which should be done by researchers.

## Selecting a healthy diet standard

A healthy diet is defined by FBDG. Dietary guidelines define a healthy diet, specifying criteria for the Cost of a Healthy Diet calculation, using food group classification (i.e. fruits, vegetables, starchy staples) and recommended quantities for each food group. Users can choose to calculate CoHD using a global and/or a national standard. The dietary standard selected is inputted into the Cost of Healthy Diet\_Excel Workbook (see “Target Quantities” sheet).

In countries that have quantified national FBDG, these are typically used as a cost standard, to identify the minimum cost of purchasing a diet in the market that meets the country’s FBDG. If quantified FBDGs are not available or if comparisons across countries are being made, the Healthy Diet Basket (HDB) can be used. The HDB is used as a cost standard globally that reflects the commonalities among FBDG around the world and is used for global monitoring (Table 1).

Reasons to use the Healthy Diet Basket could include any of the following:

* It would be useful to compare your local results with the results reported for the country in FAOSTAT.
* You are working in several countries in a region, and need a comparable standard across countries.
* The country where you are working has no FBDG, or has FBDG which are not quantified.
* The country where you are working has FBDG, but it has not already been quantified in the Food Prices for Nutrition Excel Workbook (see “FBDG Recommendations” sheet), and you or your team do not have the expertise required to quantify it.

**Table 1. Healthy Diet Basket food group quantities**

Source: Herforth et al. 2022. More information at: <https://www.fao.org/faostat/en/#data/CAHD/> and <https://sites.tufts.edu/foodpricesfornutrition/methods/>

# Data

Retail food prices are used to calculate the Cost of a Healthy Diet because they reflect the actual costs that consumers pay when purchasing food items from markets or other outlets. Wholesale or farmgate prices do not capture the additional costs and considerations associated with purchasing food at the retail level – such as with transportation, storage, handling, packaging, and other retail markups – so they are less relevant for assessing the cost and affordability of a healthy diet for consumers.

When considering the use of retail food prices, it is important to explore existing data sources before initiating a new data collection effort. Existing data collection efforts often benefit from well-established and consistent data collection protocols, long-term relationships with markets and vendors that improve ease of price data collection, and local linguistic and cultural fluency, all of which can increase accuracy of price data observations.

## Existing data sources

In many countries, retail food price data is typically collected by national statistical organizations (NSOs) for ~50-150 commonly consumed food items on a monthly basis in districts, states, or regions throughout the country in order to monitor inflation using the Consumer Price Index (CPI). These data typically have sufficient diversity to calculate CoHD, and are the data source for existing national and global monitoring of the CoHD indicator (Herforth et al., 2022). If the CPI food price data is not publicly available at item level, it is possible to request it from the NSO.

For some research and programmatic applications, such as those that operate at large zones of influence, CPI data may the best data source to characterize the local food environment. CPI data have the advantages of being official statistics of the country, and being collected in a consistent and representative way. For other applications, researchers or programmers may wish to understand the food environment in a specific small or rural geographic area of interest.

In some contexts, there may be existing sources of retail food price data for specific small or rural geographic areas of interest. These include data from Market Information Systems (MIS) and Early Warning Systems (EWS) – two generic terms which refer to retail price data collected by a wide variety of entities.

* Market Information Systems are public or private systems for the collection, analysis and dissemination of comprehensive price data related to agricultural and food markets — they may collect data for commercial purposes, and may charge researchers for access. MIS may have a different mandate (and a larger or smaller scope of commodities) than the NSO CPI data collection efforts.
* Early Warning Systems provide data on the availability and price of agricultural and food products with the goal of alerting governments and stakeholders on potential food security issues such as price volatility, commodity shortage, or other market shocks. Their primary objective is to detect potential crises and allow for timely mitigation efforts. EWS may be national or regional (like WFP-FAO) and usually monitor a more limited set of food items.
* If the food price list for an MIS or EWS has at least the minimum number of items in each of the six food groups of the HDB, it can be used for price monitoring (see p6).

## Collecting new retail food price data

If researchers or programmers wish to collect their own food prices, careful attention is needed to price data collection methods to get accurate prices representing the market food environment of interest.

The sampling approach to obtain retail food prices will include multiple considerations: 1) geographical location; 2) outlets within locations; and 3) food list; 4) replicate price observations; and 5) timing. In order to sample at each level, probability sampling, purposive sampling, or a combination of both might be used depending on the study needs as well as resources available. Additional considerations related to sampling can be found in the IMF and UN [Consumer Price Index Manual](https://www.imf.org/-/media/Files/Data/CPI/cpi-manual-concepts-and-methods-white-cover-version.ashx).

### Geographical location

Depending on your study needs, you may purposively select specific locations for the collection of price data. Alternatively, if you wish to use probability sampling, you could randomly select geographical units within a given administrative area based on population size or other characteristics. Overall, the sampling approach for the geographical location, as well as price data collection more generally, is flexible and will always depend on the needs of the researchers or practitioners.

### Outlets within locations

The outlet selection for collecting the food prices for each of the food items included on the food list will depend on where people from the population of interest (or community) typically purchase their foods. This information can be obtained from the participatory social mapping focus group discussion, that is recommended as a first step to inform the implementation of the food environment toolbox assessments. If the population purchases foods from a variety of outlet types (e.g., wet market, mobile vendors, supermarkets, and local kiosks or small groceries), it makes sense to collect food prices at each of these types of outlets. The different market names and locations can be recorded in the data collection form.

Once the outlet types that will be included in the sample have been identified, the next step is to identify which specific outlets will be selected for the data collection. For CPI price data collection, preference is to choose larger or more stable or popular outlets that will reliably have prices for all or most items continuously. This approach is recommended.

If a variety of niche markets is desired for the research aims (e.g. markets that sell only one or a few types of food), there may be business directories or registers that include all of the food outlets in a given location, from which outlets could be randomly selected. However, in many LMIC contexts, this will not be possible given that a detailed business directory might not be available. In these cases, a purposive or convenience selection of outlets might be necessary. In a dynamic market environment, there may be regular turnover in outlets. If a vendor ceases their operation or declines to participate, they should be substituted with a similar outlet (i.e. replace a market stall with another market stall), located in the same area and offering a similar category of food items.

### Items and Specifications

The food list will be used to guide the collection of price data for specific foods.

The first step in collecting new retail food price data is to develop a list of foods to be included in the price data collection, which should be based on what is locally consumed. The list should include at least 3-10 items in each food group of the dietary guidelines used for the CoHD (e.g. starchy staples, vegetables, fruits, animal source foods, pulses nuts or seeds, oils and fats). Aim to select a set of food items that are commonly consumed and low cost (e.g. lobster is unlikely to be either the most common or the least cost animal source food, while fluid milk and dried milk powder may be more likely to be common and least cost). A suggested minimum number of food items is:

* 10 starchy staple foods
* 9-12 vegetables
* 6-10 fruits
* 6-10 animal-source foods
* 5-6 legumes, nuts or seeds
* 3-5 oils or fats

As a tool to help select common food items for each food group, a systematic identification of commonly consumed foods for over 120 countries globally is available at [www.worldfoodmap.org](http://www.worldfoodmap.org). We recommend working closely with local partners (or existing food consumption data sources) to finalize the list of foods to include in the price data collection.

The list of product types should include specifications to help enumerators to consistently select products. It is important to ensure that the specifications (e.g., brand, package size) are not too tight, leading to difficulties for the enumerator to identify a product that meets the specifications. “Loose” specifications give the enumerator more flexibility in choosing population local products and varieties and to adjust the sample to local conditions (UNECE 2009).

Food prices should be collected for the average quality in the market. It is important that the food prices collected represents the typical purchasing behavior of area consumers. If there are various varieties (e.g., imported vs local rice) of food items, collect the price for the most commonly purchased variety (which can be pre-identified ahead of data collection).

Avoid collecting prices for food items that are obviously damaged or close to expiration date — which may be offered at a discounted price and could skew the data. Avoid collecting prices for premium (very high) quality food items that only wealthy consumers may purchase (unless researching price premiums).

If the researcher wishes to explore differences in price based on quality, convenience, or sustainability characteristics, items with increased levels of quality, convenience, or sustainability should be selected in addition to basic “average” versions of the item. These should be clearly marked in a separate data collection sheet so that the price premium for these items can be identified (see data collection form in the Appendix). For example:

* Quality characteristics: premium, imported, Grade A, etc.
* Convenience characteristics: pre-washed, pre-chopped, individual serving sizes, etc.
* Sustainability characteristics: organic, fair trade, plastic free, etc.

### Number of samples

Aim to collect a minimum of three prices per item included on the food list. These three prices should be obtained from different vendors or outlets, depending on the goals for food environment characterization.

### Quantity of food in non-standard units

Foods like fruits and vegetables are often sold in non-standard units (e.g., bunch, piece, piles). In these cases, record the **unit sold** as well as the **price per unit.** Spaces for this data are provided in the Food Price Data Collection sheet (Appendix). Weigh the units sold (e.g., bunches of greens) for each price sample, so that an accurate price per kg can be determined. It is important to standardize quantity standards, as larger bulk quantities may be sold at a reduced price per unit.

It may be necessary to purchase the food if the vendor does not allow the enumerator to weigh the food. It is important to ensure that there is sufficient money budgeted to make these food purchases throughout the data collection. Enumerators should also have either a portable scale to allow them to weigh foods to calculate the price per kg in the market, or plan to purchase items and weigh them after returning from the market.

### Timing of data collection

**Setting collection dates:** The selection of days of the week and month for price data collection should ideally align with when consumer purchases are most concentrated (common market days) and when prices and available goods accurately represent the entire month. However, it is important to consider retailer cooperation, as vendors may be less willing to cooperate when busy. While fixed intervals may be useful to establish ahead of time, data collection should be avoided during weeks that include public holidays. In terms of sampling strategy, also consider when people are paid. For example, the Nigeria CPI team does not collect food price data when main salaries are paid out (fortnight payments) because of a spike in buying and prices.

**Seasonality:** Food prices often fluctuate seasonally. When characterizing food environments, consider how seasonal variations will be captured. Some items disappear in markets in certain seasons. The least-cost method selects the least cost items in each food group at the time and place where data are collected; thus, the least cost items are not the same over time. This is one reason to include more than the minimum number of items in the food list.

**Time of day**: Time of day may be an important consideration, depending on contexts. In many markets, certain products tend to experience greater price volatility throughout the day, including perishable items — fruit, vegetables, meat, fish, etc. — as well as fuel. In some cases, prices may higher in the mornings (when products are fresh) and lower at the end of the day (especially where refrigeration is limited). The main overarching consideration is trying to get “average” prices for typical products. Therefore prices are often best gathered in the late morning in LMICs.

### Prices not listed

In many food outlets in LMIC, food prices are not listed. In these cases, it is often customary to negotiate the price with the vendor. To derive the normal real price paid by a typical consumer, the enumerator has a few options. One is to develop a long-term relationship with the vendors, as CPI enumerators often do – this affects vendors’ willingness to report prices. Absent this relationship, the enumerator may need to negotiate with the vendor to get the real price, behaving as an ordinary consumer trying to buy the product in that context. This should be discussed and practiced in the enumerator training. It is important to consider the gendered nature of food shopping in each context, recognizing that male and female shoppers may get different prices, and either select or train enumerators with this awareness. Techniques to secure a representative final transaction may vary by country. The [IMF CPI manual](https://www.imf.org/-/media/Files/Data/CPI/cpi-manual-concepts-and-methods.ashx) (2020) has further guidance on this issue (e.g. p81).

### Discounted prices

There are many ways in which vendors may discount prices to incentive consumers to purchase their food. The [IMF CPI manual](https://www.imf.org/-/media/Files/Data/CPI/cpi-manual-concepts-and-methods.ashx) (2020) and Practical Guide to Producing Consumer Price Indices (2009) provides many examples of how enumerators collecting food prices should treat these scenarios. We describe several of these below, based on the guidance provided in the Practical Guide:

* *Discounts on lower quality stock*: vendors may sell foods that are close to expiration, damaged, less fresh, stale or deteriorated quality at a lower price. Prices for these foods should not be collected.
* *Special offers/temporary discounts*: If vendors offer temporary discounts for a specific food, it can be included in the price data collection unless it is being discounted based on being lower quality stock (see point above).
* *Bonus offers, extras and free gifts*: If a food is being sold with x% free (i.e., a bonus quantity of the food), the price should be adjusted to reflect the higher quantity. However, if it is thought that consumers would not want the x% free, then the adjustment of price may not be warranted. Additional information regarding this scenario can be found in the Practical Guide.
* *Loyalty program rebates, refunds, or coupons*: Loyalty rebates or coupons should be ignored.

There are many considerations when collecting food prices. We outline several considerations that often arise while collecting price data above. Enumerators should be trained on context-specific factors, such as dealing with unlisted prices, negotiating with vendors if bargaining us customary, handling discounts and premiums, addressing quantity-related challenges, and accounting for other food characteristics. Enumerators can help to inform the data collection process by documenting the circumstances they encounter, which can also enable more accurate data analysis.

# References

Herforth, A, Y Bai, A Venkat, K Mahrt, A Ebel, and WA Masters. 2020. ‘Cost and Affordability of Healthy Diets across and within Countries: Background Paper for The State of Food Security and Nutrition in the World 2020’. Rome, Italy: Food & Agriculture Org. https://doi.org/10.4060/cb2431en.

Herforth, A, A Venkat, Y Bai, L Costlow, C Holleman, and WA Masters. 2022. *Methods and Options to Monitor the Cost and Affordability of a Healthy Diet Globally: Background Paper for The State of Food Security and Nutrition in the World 2022*. FAO Agricultural Development Economics Working Papers 22. Rome, Italy: FAO. https://doi.org/10.4060/cc1169en.

International Labour Organization/International Monetary Fund/Organisation for Economic Co-operation and Development/European Union/United Nations/The World Bank. 2020. Consumer price index manual: concepts and methods. Washington, DC : International Monetary Fund. https://www.imf.org/-/media/Files/Data/CPI/cpi-manual-concepts-and-methods-white-cover-version.ashx

UNECE. 2009. ‘Practical Guide to Producing Consumer Price Indices’. ECE/CES/STAT/NONE/2009/2. New York and Geneva: United Nations Economic Commission for Europe. https://unece.org/fileadmin/DAM/stats/publications/Practical\_Guide\_to\_Producing\_CPI.pdf.

WFP. 2017. ‘Findings for IANDA / Cost of the Diet in Ghana’. World Food Programme. https://docs.wfp.org/api/documents/743ba945c9b54b668643d77786b71b0a/download/.

# Acknowledgements

This guide was written by Anna Herforth, Rachel Gilbert, Kristina Sokourenko, and Shauna Downs. It was supported by the Bill & Melinda Gates Foundation and UKAid through two mechanisms: an IMMANA Grant ([The Food Environment Toolbox: Tools to Measure Natural and Built Food Environments in Low- and Middle-Income Countries](https://www.anh-academy.org/immana/grants/grants-round-3/food-environment-toolbox-tools-measure-natural-and-built-food-environments)) and the [Food Prices for Nutrition project](https://sites.tufts.edu/foodpricesfornutrition/) (INV-016158).

**Food Price Data Collection sheet**

Instructions:

The Excel workbook titled **“CostofHealthyDiet\_ExcelWorkbook\_TEMPLATE\_withDataCollectionSheet”** is where you can enter price observations that you collect. The data entry sheet is formatted as shown here. If you wish to print a paper copy for the purpose of price data collection, use this as a template and add more rows to accommodate more items.

* Collect three prices for each of the items included on the price list below.
	+ Note: If you are unable to find 3 markets/vendors selling any of the foods on the list and therefore have fewer than 3 prices for an item, you will need to adjust the formula for computing the average price in column U of the “Food Price Data Collection” sheet.
* Food prices will be documented in the local currency per unit sold.
* Include the units that the food is sold in. If the units are not grams, kg, L or ml, you will need to weigh the unit sold (e.g., bunch) to ascertain the weight of the item in grams or kilograms.

Example:

**Market Name:** Pike Place Market

**Market Location:** Seattle

**Market District or Region:** WA

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date** |  **Item Code**  | **Item Name** | **Unit Sold - sample 1** | **Price per Unit - sample 1** | **Item Weight** (kg) **- sample 1** | **Unit Sold - sample 2** | **Price per Unit - sample 2** | **Item Weight** (kg) **- sample 2** | **Unit Sold - sample 3** | **Price per Unit - sample 3** | **Item Weight** (kg) **- sample 3** |
| 11/1/2023 | 1 | apples | bag | 4.00 | 5.00 | bag | 4.50 | 6.10 | kg | 1.00 | 1.00 |
| 11/1/2023 | 2 | salmon | lb | 8.00 | 2.20 | lb | 12.00 | 2.20 | fish | 50.00 | 10.00 |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

**Instructions for Exploring price variation by quality, convenience, or sustainability characteristics**

**Step 1: Preparation for data collection / entry**

Copy the entire Excel workbook. Label it “v2\_convenience”. If you are recording price observations on paper, make a new copy of the data sheet, where you will record price observations for items that have increased convenience characteristics (e.g. pre-washed, pre-chopped, individually packaged serving sizes).

**Step 2: Data collection / entry**

In the original workbook, record prices as usual, i.e. prices of average versions of each item.

In version “v2\_convenience”, record prices of the same items (e.g. apples), but with increased convenience characteristics (e.g. pre-chopped).

**Step 3: Analysis of CoHD and the price premium for convenience**

Calculate CoHD and related indicators in V1, and V2. Compare the cost and note the cost premium for convenience.

**Ensuring Data Quality through Electronic Data Collection**

Data Collection on Tablets:

Collecting food price data on digital tablets can enhance the accuracy, speed, and efficiency of data collection, and streamline the process for calculating Cost of a Healthy Diet and related indicators. For best results, enumerators or data collection officers should be equipped with tablets that have a pre-installed data collection functionality — there are several robust data collection applications available, many of which are open-source and customizable.[[5]](#footnote-5) The mobile data collection application should be user-friendly and easy to use. It is important that data collectors are thoroughly trained not only on the technical use of the tablets but also on performing data checks, to identify genuine data variations versus errors, and validate or correct data as needed.

Programming Data Checks:

To ensure high data quality, the data collection software you use should be programmed to automatically flag outliers to identify when a price falls outside of its expected range. Many applications use pre-set thresholds to prompt enumerators to confirm anomalous entries that are significantly higher or lower than seen in historical data. If outliers are detected, enumerators may be reminded to document the unit of sale or item quality, or could document the displayed price of the item in question. If data is missing for certain items, applications can prompt enumerator to confirm unavailability — nudging them to double check that certain items were not overlooked. Other consistency checks can be installed to query entries for potential errors.

1. https://www.fao.org/faostat/en/#data/CAHD/ [↑](#footnote-ref-1)
2. https://databank.worldbank.org/source/food-prices-for-nutrition [↑](#footnote-ref-2)
3. <https://www.worldbank.org/en/programs/icp> [↑](#footnote-ref-3)
4. This 18-minute tutorial is available at <https://sites.tufts.edu/foodpricesfornutrition/tools/>. The methods are covered more in depth in a six-hour self-paced World Bank [eLearning course](https://www.worldbank.org/en/olc/course/62176). [↑](#footnote-ref-4)
5. e.g. ODK Collect (Open Data Kit), KoBoToolbox (suite of tools developed by Harvard humanitarian initiative, known for ease of use and ability to work offline) [↑](#footnote-ref-5)