SOFTWARE TOOL ARTICLE

Software tools for practical application of human nutrient requirements in food-based social science research [version 1; peer review: 1 approved with reservations, 1 not approved]

Kate Schneider¹, Anna Herforth²

¹Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, 02111, USA
²Independent Researcher, Brookline, MA, USA

Abstract
This article focuses on the use of human nutrition requirements for applied social science research by non-nutrition experts. Our motivation is to provide the data and default nutrient requirements to calculate the Cost of Nutrient Adequacy (CoNA) least-cost diet meeting specified nutrient requirements. The final format of the resulting datafile is ready for use in the CoNA protocol. However, we expect this article and associated software tools to be generally useful to social science researchers interested in nutrition-related research questions and the potential implications of food systems’ changes for nutrition. We provide readily usable data files (Supplement 1) containing the Dietary Reference Intakes (DRIs) and guide non-nutrition experts through appropriate use to establish the nutrient needs and assess the adequacy of diets for populations and groups. We complement the DRIs with companion data files (Supplement 2) containing the WHO Child Growth Standards and WHO Growth References for School-Aged Children and Adolescents percentiles tables of anthropometric measures, extract the median heights and weights, and calculate median reference values for the age-sex groups consistent with the DRIs nutrient requirements. We provide calculations of energy requirements using the DRIs Estimated Energy Requirement (EER) equation and WHO growth references for all age-sex groups and physical activity levels. We also calculate the protein Estimated Average Requirement (EAR) per kilogram body weight according to the WHO growth references. For children under two, we provide nutrient needs required from food in Supplement 3. We provide Stata code and R syntax (Supplements 4 and 5) to compile the single data files into usable datasets for statistical analysis. Finally, we also provide data files (Supplement 6) with the recently proposed harmonized average values and upper levels and briefly discuss their potential application.

Open Peer Review

Approval Status

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>version 1</td>
<td>✗</td>
<td>?</td>
</tr>
</tbody>
</table>

1. Olivier Ecker, International Food Policy Research Institute, Washington, USA
2. Saskia De Pee, World Food Programme, Rome, Italy

Any reports and responses or comments on the article can be found at the end of the article.
Introduction
Agricultural, development, and applied economists and other social scientists are increasingly concerned with incorporating nutrition into their research. In some cases, researchers have information about foods from survey data and are interested in the implications of changes – such as in food availability (e.g. via production, market changes, etc.), food consumption and demand (consumer behavior), and food prices – on nutrient intake and sufficiency among specific populations. Such analysis requires knowing both the human nutrient requirements for the population of interest as well as the nutrient composition of foods, provided in food composition tables. This article focuses on the use of human nutrient requirements for applied social science research by non-nutrition experts. It is explicitly motivated to calculate the Cost of Nutrient Adequacy (CoNA). The use case described below applies the default nutrient requirements we recommend for use when calculating the least-cost nutritionally adequate diet in accordance with the CoNA procedures.

We expect this paper to have broad utility for numerous social science researchers wishing to incorporate nutrition into agricultural, development, and food systems analyses. In addition to documenting the nutrient requirements protocol used in CoNA studies, it is intended to guide non-experts in applying human nutrient requirements in food-based social science studies. It may further be useful to nutrition experts as a simple compilation of requirements and growth charts in a data format readily usable by statistical software.

Methods
This article uses three main standard references: the Dietary Reference Intakes (DRIs), developed as a joint effort between the US and Canada and published by the Institutes of Medicine of the National Academies (IOM); WHO Child Growth Standards; and WHO Growth References for School-Aged Children and Adolescents (“WHO growth standards and references”). It also incorporates the recommendations for weight gain during pregnancy also from the IOM. We chose to use the 2006 “Dietary Reference Intakes: Essential guide to nutrient requirements” instead of the WHO/FAO/UNU requirements because it made a number of important advances to establishing human nutrient requirements, notably the formalization of a requirement appropriate to use in the context of populations and groups, the Estimated Average Requirement (EAR); and the use of doubly labeled water (gold-standard) methods to calculate energy requirements. The DRIs were subsequently reviewed and updated in 2011 for calcium and vitamin D and in 2019 for sodium and potassium.

Nutrient requirements
The IOM DRIs contain nutrient requirements for micronutrients (vitamins and minerals), energy, and macronutrients (carbohydrates, protein, and fats) for 25 distinct population groups that differ by age, sex, and life stage (e.g. pregnancy and lactation). The 2006 DRIs marked an important evolution in defining human nutrient requirements. In particular, the DRIs comprehensively included all requirement categories and all essential nutrients for the first time. Furthermore, the most important advancement of the 2006 DRIs was to establish an Estimated Average Requirement (EAR) appropriate for use in analyses of populations and groups. Using the EAR with survey data of individuals applies the median nutrient requirement to all individuals within the population (age, sex, and lifecycle group) for which it is defined. The UN organizations (and others) had described this as the appropriate requirement to use for populations and groups, but it had not been formally established in any prior publications. It was also the first time safe upper levels were established.

Some researchers may question the use of the DRIs for studies globally, given that they were developed for use in North America. Recommendations have been developed by WHO and compiled in collaboration with FAO, UNICEF and UNU contained across multiple documents. These global recommendations, however, do not contain EARs, and were not as rigorously developed as the IOM DRIs. The other ongoing international effort to set human nutrient requirements based on the most recent scientific evidence is carried out by the European Food Safety Authority (EFSA). There are current ongoing efforts to harmonize the nutrient requirements released by these and other scientific groups, detailed by Allen et al., who recently proposed an initial set of harmonized requirements integrating the IOM DRIs and EFSA dietary reference values for 26 macro- and micronutrients. We provide data files to utilize these harmonized values (Extended data, Supplement 6) but detail the IOM DRIs values in the following sections, as our use case (the primary motivation for this software tools paper and what is contained in our source code) employs only the DRIs values.

Micronutrients. The DRIs contain four different types of requirements: the EAR (the daily amount necessary to meet the needs of 50% of the healthy population); the Average Intake (AI) (the amount estimated to be sufficient to meet the needs of nearly all individuals based on the average intake of healthy

1 Recently renamed and moved into the National Academies of Science, Engineering and Medicine, but we continue to refer to it as the IOM. For further discussion on the institutional history and establishment of nutrient requirements for the US and Canada, see Beaton 1999 and Yaktine and Ross 2019
2 See also FAO, WHO and UNU 2001 for discussion of the limitations of other approaches.
3 Beaton (1999) provides a comprehensive and detailed discussion of the terminology, history, debates, disagreements, and uses of nutrient requirements worldwide since the earliest discovery and awareness of nutrients and human nutrient needs. Beaton’s analysis focuses on the differences between publications from the IOM and UN. Other discussions of the advancement the 2006 DRIs represent can be found in Kennedy and Meyers (2005) and Yaktine and Ross (2019).
populations); the Recommended Daily Allowance (RDA) (the amount estimated to meet the needs of 97.5% of the healthy population, defined as two standard deviations above the EAR); and the tolerable upper intake level (UL) (the maximum daily intake which does not produce any signs of toxicity).

For each micronutrient per age-sex group, the IOM defines either 1) an RDA and EAR or 2) only an AI level. The RDA is a high level of intake that meets the needs of nearly all healthy people. Where an RDA can be set, beginning with the 2006 DRIs, the EAR is also established. The EAR, which defines the average nutrient requirement in a population, is the appropriate metric for social science researchers interested in questions about populations and groups.10,11. When there is insufficient evidence for an RDA and EAR, the IOM establishes an AI, which is the recommended average daily intake based on observed or experimentally determined approximations of nutrient intake in healthy individuals that are assumed to be adequate. AIs are also expected to meet the needs of most individuals. RDAs and AIs are used in clinical practice to ensure nutrient needs for any given individual are met, while EARs are used to estimate needs of populations and plan population-level interventions.

Though the EAR is the appropriate nutrient requirement to use when defining the needs or assessing the diets of populations and groups, setting an EAR requires a high degree of scientific evidence and therefore has not been established for all nutrients or all ages and sexes. Therefore, for analyses of populations and groups, we recommend including only those nutrients that have a defined EAR. Among the micronutrients with an EAR, we further recommend excluding three nutrients from food-based analyses: vitamin D, iodine, and molybdenum. Vitamin D should be excluded given high variation within populations in the quantity produced by the body and variation due to sun exposure and race/ethnicity. Iodine is often fortified in salt and typically not measured in food composition tables. Molybdenum is similarly not contained in food composition tables, is present in many foods, and diseases of deficiency have never been observed in healthy humans.

Some micronutrients can also produce harmful effects if too much is consumed. Tolerable Upper Intake Levels (ULs) are set at the level of intake that produces symptoms of toxicity. The documentation for each nutrient specifies whether the UL is possible to reach from food sources or only supplements. A new requirement termed the Chronic Disease Risk Reduction (CDRR) level was established for sodium in 2019. The reason for establishing a new type of nutrient requirement was that the committee found sufficient evidence to establish upper bound limits on sodium intake to reduce chronic disease risk, but these are not the same as symptoms of toxicity, the criteria required to set a UL.

When defining healthy diets, such as in the least-cost diet application described below, we recommend including the UL for all nutrients for which it is defined and can be reached with food sources. Additional considerations exist for vitamin A, however: the UL for vitamin A pertains only to retinol, the preformed form of vitamin A found in animal source foods as well as some medications. The EAR for vitamin A includes both preformed and provitamin A carotenoids. Carotenoids (e.g. beta carotene) are found in many foods and are converted by the body to an active form only as needed, and therefore pose no risks of toxicity. Excessive intake of retinol, on the other hand, can cause severe consequences; for example if consumed in excess by pregnant women it causes birth defects. For this reason, we treat vitamin A and retinol as separate nutrients, with the former having only an EAR and the latter having only a UL value. Finally, although not formally defined as a UL, we recommend including the CDRR for sodium as an upper bound on sodium intake which takes the same form in practice as a UL when defining healthy diets. The 2019 DRIs defined the CDRR given the high degree of scientific evidence that higher consumption of sodium substantially increases the risk of chronic disease, though it does not produce symptoms of toxicity.

Energy. The Estimated Energy Requirement (EER) is an equation to determine energy needs based on age, sex, weight, height, and physical activity level. An active level of physical activity is recommended for long-term health and therefore we recommend estimating nutrient requirements for this level of activity. For populations and groups, reference values for weight and height are taken from child growth charts, discussed in the next section.

Macronutrients. Humans need carbohydrates, protein, and lipids in larger amounts to provide energy to the body and carry out numerous metabolic functions. There is an EAR for carbohydrates and protein that sets an absolute minimum amount required. The protein EAR is calculated per kilogram of body weight and is therefore sensitive to weight in addition to age, sex, and life stage. Long-term health requires more than these minimum values. The Acceptable Macronutrient Distribution Range (AMDR) specifies a range of acceptable proportions of calories from each macronutrient. These are then converted into gram quantities based on energy needs. Carbohydrates and protein supply four calories per gram and lipids provide nine calories per gram. Therefore, the AMDR moves with the energy needs, which are sensitive to the height and weight reference parameters used and the selected physical activity level.

Reference heights and weights

Energy requirements are sensitive to the reference heights and weights, and level of physical activity selected. The 2006 DRIs reference heights and weights come from the CDC/NHCS growth references. After the release of the WHO child growth standards, the CDC recommended the use of the WHO standards for all children 0–59 months (through 5 years) in the United States.

---

4 Not all food composition tables, however, will contain retinol content separate from carotenoids and/or total retinol activity equivalents (RAEs).
5 Fats do not have a minimum EAR.
The publication of the WHO child growth standards in 2006 represented an important step forward in the assessment of child growth and the definition of healthy growth trajectories. The standards were developed through a multi-center study carried out in six sites around the world, covering all continents and communities at different levels of economic development. Though it developed growth *standards* only for children 0–59 months, mathematical modeling permitted the extrapolation of growth *references* for school aged children and adolescents (ages 5–19 years) based on the child growth standards. The development of the WHO child growth standards is discussed in depth in numerous other publications, but it is important for non-nutritionists to know that the rigorous multi-center global study was designed to move beyond growth references for a single population (describing how children *do* grow), to a set of standards for how children *should* grow in the condition of optimal health and growth. They are now considered the globally recognized source for children 0–59 months and therefore are the appropriate resource from which to draw reference heights and weights to estimate nutrient requirements.

To be consistent in the source of references for all ages and sexes, and in light of the global studies underlying the references, we recommend the use of the WHO references for all ages and sexes. The end-growth (19 year old) median heights and weights are carried forward into adulthood, which is the methodology used by the *DRIs* with respect to the CDC/NCHS growth charts. The WHO growth reference study explicitly studied children age 5 years and under to determine that child growth curves are the same across cultures and populations; it did not determine the similarities or differences in observed attained adult heights across populations. We use the WHO growth standards end-growth values as the best estimate of median adult heights and weights.

The appropriate reference weight for pregnant women can be estimated as the median adult weight for her age group plus a recommended amount of weight gain during pregnancy. The IOM reference heights and weights used in the *DRIs* incorporate CDC guidelines regarding weight gain during pregnancy. To the best of our knowledge, the WHO does not have any guidance regarding weight gain during pregnancy. To develop the reference weights for pregnant women included in these software tools, we therefore take the median adult weight as defined by the WHO growth references and add the midpoint of the CDC recommended weight gain range (13.25 kg) over the total pregnancy for a woman of healthy weight at the onset of pregnancy. Researchers wishing to include this life stage in analyses may utilize the median weight gain recommended over the entire pregnancy included in these software tools; however, those particularly concerned with nutrient requirements during pregnancy are encouraged to work with a nutritionist and consider the differing needs in each trimester.

**Considerations in application**
We outline four likely ways in which researchers may need to modify the nutrient requirements for certain analyses.

 Probability of iron deficiency in adequacy analysis. Many researchers may be interested in estimating the probability of inadequacy by comparing observed nutrient intakes to requirements. At the population level, for most nutrients, the EAR cut-point method can be used because the distribution of actual needs in the population is understood to be normally distributed. This means that any individuals consuming below the EAR are estimated to be inadequate and those above to be adequate and at the population level this yields an accurate estimate of the percentage of the population likely to be inadequate. However, iron requirements are not normally distributed in the population and therefore a probability estimate must be used. The probability of iron inadequacy varies by age, sex, reproductive status, and oral contraceptive use. The *DRIs* provide the probability of inadequacy for intakes within specified ranges by age, sex, and oral contraceptive use. For population level analyses, researchers may know that oral contraception is not available and assume an entire non-user population, or, if oral contraceptives are available and likely to be used, the researcher should estimate the probability based on a mixed-user population.

**Children 6–23 months.** Children who are still breastfeeding have some of their nutrient needs met from breastmilk and the WHO recommends continued breastfeeding for the first two years of life and as long thereafter as desired by mother and child. The definitive resource on nutrient needs during this period is Dewey 2005 which contains comprehensive requirements for all nutrients including the quantities provided by breastmilk and needed from food. When considering this age group as a member of a family or otherwise heterogeneous groups, such as in establishing household needs, we recommend applying consistent criteria for the choice and inclusion of nutrient requirements across all groups and therefore including only those contained in the *DRIs* for this age range, and adjusting for the contribution of breastmilk according to Dewey 2005. This is the method we have adopted in these software tools. Given the complexity of estimating nutrient needs met through breastmilk vs. complementary feeding, we advise any social scientists carrying out studies of nutrient needs that are explicitly focused on children under two to collaborate with a nutritionist.

**Iron and zinc bioavailability.** Many populations in low- and middle-income countries consume diets high in phytates and oxalates. These compounds, present in unrefined grains and some other plant foods, bind to iron and zinc and reduce the body’s ability to absorb those nutrients. The body will only absorb some overall fraction of each nutrient, and the specific value of this fraction is dependent on the nature of the whole diet and

---

6 The resulting values using the WHO growth charts are nearly identical to the adult CDC/NHANES end-growth values. For men, the difference in weight is 1.2 kg and 0.5 cm in height (CDC is the higher value). For adult women, the difference is 0.2 cm in height (WHO is the higher value) and no difference in weight.
not yet fully understood\textsuperscript{15,36}. When evaluating the adequacy of observed diets where the dietary pattern can be characterized by low bioavailability, researchers may choose to revise the iron and zinc requirements to reflect an assumption of lower bioavailability\textsuperscript{18}. We stress that understanding bioavailability is an active area of nutrition research and therefore do not make an explicit recommendation regarding whether or not to adjust the iron and zinc requirements for low-bioavailability diets. If choosing to do so, we advise researchers to collaborate with a nutritionist.

**Physical activity in rural agricultural populations.** Researchers interested in certain populations whose daily lives are physically demanding may wish to consider energy needs at a higher level of physical activity. Where women’s daily lives include energy-demanding activities such as collecting water and firewood and pounding grain, the UN characterizes these energy needs as consistent with the “active” physical activity level in the DRIs. We therefore do not recommend any adjustments for adult women. But for men such as farmers in unmechanized settings and construction workers carrying out demanding manual tasks, the UN describes energy needs consistent with the DRIs’ “very active” physical activity level\textsuperscript{16}. Therefore, in such settings, we recommend using energy requirements for the very active level of physical activity for males aged 14–59 as they more accurately reflect needs of physically demanding lifestyles\textsuperscript{18}.

**Operation**

The software tools in this article are developed to run using open source R statistical software. Code is also provided in Stata format. The extended data files can be opened with any spreadsheet software or freely available browser application. It is not required to open the extended data files separately from compiling the larger datasets with statistical software, however it is helpful to become familiar with the shape of the data to understand how they must be reshaped and transformed in order to compile a single long form dataset. The workflow in brief starts with compiling the DRIs data files, resulting in a dataset that is long form by age-sex group, nutrient, and physical activity level. Then, optionally, the energy and protein requirements can be recalculated using WHO reference heights and weights. Then, also optionally, the requirements for children 6–23 months can be reduced to only the percentage required from food. A single age-sex group at a single level of physical activity is extracted in the final step as an example for use in calculating a least-cost diet.

**Implementation**

The software tools accompanying this analysis provide four extended data supplements and two software files in Stata code and R syntax to compile a nutrient requirements dataset in a format amenable to calculating a least-cost diet meeting these nutrient requirements. Six supplements are provided, each described in further detail in the subsequent sections:

- **Extended data, Supplement 1\textsuperscript{17}:** The DRIs (EARS, ULs, CDRRs, AMDRs and energy) for all age and sex groups. Supplement 1 contains 12 data files including a notes document\).
- **Extended data, Supplement 2\textsuperscript{19}:** The WHO growth references with recalculated energy and protein requirements (per the DRIs) using WHO median weights and heights. Supplement 2 contains 14 data files including a notes document\textsuperscript{18}.
- **Extended data, Supplement 3\textsuperscript{19}:** Requirements for ages 6–23 months during continued breastfeeding. Supplement 3 contains two data files including a notes document\textsuperscript{17}.
- **Extended data, Supplement 4\textsuperscript{19}:** The Stata code (.do file) producing a single dataset (in .dta format) with the DRIs nutrient requirements at all levels of physical activity, and with energy and protein calculations based on the WHO growth references. Optional code is also shown to adjust the needs for infants under 2. Finally, it also provides example code to extract a single age-sex group at a specific activity level in the necessary shape for use in a least-cost diet application.
- **Extended data, Supplement 5\textsuperscript{19}:** R syntax (.r file) identical to the Stata code in Supplement 4 to produce an identical dataset in .Rdata format.
- **Extended data, Supplement 6\textsuperscript{19}:** Harmonized nutrient requirements proposed by Allen et al. (2020)\textsuperscript{18}. Supplement 6 contains four data files and a notes document.

**Supplement 1: DRIs nutrient requirements\textsuperscript{7}**

The nutrients and requirements contained in the Extended data, Supplement 1\textsuperscript{19} data files follow the selection criteria denoted above and are taken directly from the nutrient requirement tables in the DRIs documents\textsuperscript{5,4}. Extended data, Supplement 1\textsuperscript{19} also includes a file with the iron probability of inadequacy by age, sex, consumption range, and oral contraceptive use\textsuperscript{17}. Any further adjustments, important details for application, or data management procedures are denoted here.

**Units of measurement.** Each micronutrient is required in different amounts by the body and therefore the micronutrient requirements are expressed appropriately in either micrograms or milligrams. The Extended data, Supplement 1\textsuperscript{19} files make a few unit conversions to facilitate analysis. In the DRIs, the units of measurement for calcium and phosphorus differ between the EAR and the UL, so we have converted the ULs for both nutrients to milligrams to be consistent with the unit in which

---


\textsuperscript{8} Reproduced with permission of WHO.

\textsuperscript{9} Reproduced with permission of the author.
the EAR (and RDA) are expressed as well as the unit of measurement commonly used in food composition tables. Similarly, copper is measured in micrograms in the DRIs, but we have converted the requirement to milligrams since this is the common unit of measurement in food composition tables.

Researchers combining nutrient requirements with food composition are advised to confirm that the unit of measurement for the requirement is the same, per nutrient, between the requirement and the particular food composition table used in the analysis, as they do vary in some food composition tables. The AMDRs are expressed as a percentage of calories from each food, but the quantities are expressed in food composition tables in grams. Therefore, the percent of calories must be converted to the total grams of each nutrient, which depends on individual energy needs. Since energy needs are between units to all the aforementioned parameters (age, weight, height, physical activity), this conversion is done in the dataset compilation (Supplement 4). The other nutrient where units may not agree is vitamin A, which has been measured in three different units over time: retinol equivalents (RE), international units (IU), and retinol activity equivalents (RAE). The RAE is the currently accepted standard unit and is used in the USDA food composition table and the DRIs; however, other tables, especially if older, may use other units and may also separate retinol and carotenoids or provide multiple units. Conversion between units is not straightforward and non-expert researchers should not attempt to do so and could instead consider using alternate food composition data (such as USDA) where RAEs are not available.

**Inclusion of the sodium CDRR.** For the purposes of defining the requirements for healthy life, the sodium CDRR is functionally equivalent to the food-based ULs for other nutrients. In other words, both requirements put an upper bound on the amount of a nutrient that is considered within the healthy range for consumption. As such, for analytical convenience, the sodium CDRR is included in the same sheet as the other micronutrient ULs in Extended data, Supplement 1. We stress that this upper bound requirement is not based on symptoms of toxicity, the criteria required to set a UL. So, while it is functionally equivalent for social science research applications, it is important for researchers to acknowledge the fundamentally different meaning of the sodium CDRR compared to other micronutrient ULs.

**Supplement 2: WHO growth references, energy & protein requirements**

Growth references are expressed with anthropometric measures that describe the relationships between weight, height and age and are sex specific. These measurements are explained in Table 1.

Measurements are presented in the growth charts as percentiles and z-score distributions, which are used to calculate population statistics for stunting, wasting, underweight and overweight. Only percentile charts are needed to calculate reference heights and weights by age and sex to determine nutrient requirements and only the percentiles charts are contained in the *Extended data*, Supplement 2 data files. The “Notes” document describes each dataset. The three files necessary to recalculate the DRIs using the WHO growth references are are based on the extraction of the median heights and weights from the subsequent sheets. The “Weights Heights by DRI Group” file calculates the median of the 50th (median) percentile heights and weights corresponding to the 25 age-sex groups contained in the DRIs. The “EnergyEER” file takes those calculated reference weights and uses the DRIs EER equation to calculate energy requirements for each age-sex group at all four levels of physical activity. Finally, the “ProteinEAR” file uses the DRIs requirement of protein per kilogram body weight and calculates a total protein EAR for every age-sex group using the WHO reference weights.

While only the “Weights Heights by DRI Group”, “EnergyEER”, and “ProteinEAR” files are required to compile the nutrient needs dataset using the software code in Supplements 4 and 5, the additional files provide the calculations and full data. For children 0–59 months, only the weight-for-age (WFA) and the

---

**Table 1. Anthropometric measures definitions.**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
<th>Units of measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weight-for-Age</strong></td>
<td>The standard used to determine under- or overweight in children under 5. Specific to age in months.</td>
<td>Weight in kg by age in months</td>
</tr>
<tr>
<td><strong>Length/Height-for-Age</strong></td>
<td>The standard used to measure linear growth (low LHFA defines stunting). Children 0–24 months are measured by recumbent length, and age two and up are measured by standing height.</td>
<td>Height in cm by age in months</td>
</tr>
<tr>
<td><strong>Weight-for-Length/Height</strong></td>
<td>The standard for determining wasting, a measure of acute or recent insufficient energy intakes. Length measured for children 0–24 months.</td>
<td>Weight in kg/ Height in cm, specific to sex and age in months</td>
</tr>
<tr>
<td><strong>Body Mass Index</strong></td>
<td>The measure for healthy weight status and only measured for children 5 and up. It is also the measure used for adults.</td>
<td>Weight in kg/ Height in m²</td>
</tr>
</tbody>
</table>


* Note this measure is not necessary to specify nutrient requirements and is only included in this table for completeness. It is not provided in the Extended data, Supplement 2 files.
length/height-for-age (LHFA) are needed and the 50th (median) percentiles are extracted from the full tables and provided for each month of age, by sex, in the “Median Weight Height Month 0–19” file. For children 5–19 years, the growth reference is expressed in terms of body mass index (BMI) and height-for-age, so an extra arithmetic step is required to determine the weights. This calculation is provided in the “Median Weight Height Month 0–19” file for all children over age 5 since the child and adolescent percentiles are expressed per month of age as well. The “Median Weight Height Year 5–19” file provides the resulting calculation of the median of the 50th (median) percentile weights and heights per year of age for all children 5–19 to provide a single value per year.

Supplement 3: Nutrient requirements during continued breastfeeding
From 6–23 months, during continued breastfeeding, the infant’s nutrient requirements change dramatically as does the relative proportion of breastmilk to food in the total diet. The DRIs divide infants in this period into only two groups: 6–11 months and 12–23 months. However, the proportion of nutrients needed from foods are much lower for 6–8 months than for 9–11 months of life. The variable bf_match_id matches the DRIs age and sex groups in the “Reference Values” file to the nutrient needs required from foods in the “6–23mo_FoodNeeds” file, effectively dividing each of the DRIs age-sex groups aged 6–11 months (age_sex_grp=2) and 1–2 years (age_sex_grp=3) into two groups. The bf_match_id variable makes possible the adjustment carried out in the data compilation below, provided with the important cautions noted above regarding nutritional vulnerability and nutrient needs during this life stage.

Supplements 4 and 5: Dataset compilation in Stata or R
We demonstrate the compilation of requirements and data management to shape the data into the format required for a least-cost diet application. The final data are in long form by age-sex group, nutrient, and physical activity level and wide form by nutrient requirement. The process is divided into five parts as follows:

- **Part 1**: Import and reshape the DRIs requirements and keep only upper limits that can be reached from food.
- **Part 2**: Merge DRIs nutrient requirements from Extended data, Supplement 1<sup>4</sup> into a single dataset and calculate the grams of each macronutrient corresponding to the upper and lower bounds of the AMDRs.
- **Part 3**: Import the WHO growth references, merge with the DRIs, and replace energy and protein requirements calculated based on WHO growth references (provided in the Extended data, Supplement 2<sup>9</sup> data files).
- **Part 4**: (Optional, if including this age group in analysis) Reduce the needs for children 6–23 months to only the amounts that must come from foods.
- **Part 5**: Extract single data files for an age-sex group of interest at the appropriate level of physical activity.

Supplement 6: Harmonized requirements data files
The proposition of harmonized nutrient requirements across scientific bodies marks an important milestone for global applications of human nutrient requirements. Our use case (the primary motivation for this software tools paper) uses only the IOM DRIs values, but other related work uses the harmonized values for global analysis (e.g. Herforth et al. 2013). Given that the harmonized values may be of interest for researchers to apply, we also provide a data files with the recently published harmonized requirements by Allen et al. (2020).<sup>9</sup> Extended data, Supplement 6<sup>6</sup> provides the harmonized average requirements (H-ARs) and harmonized upper levels (H-ULs) contained in Tables 2 and 6 of the aforementioned paper. We include all the nutrients covered in that paper that are commonly found in food composition tables, excluding vitamin D, iodine, pantothenic acid, biotin, and choline. The harmonized requirements do not cover energy or macronutrients beyond an average requirement for protein, and therefore researchers wishing to apply the harmonized requirement will need to choose how to define the energy and macronutrient requirements. We do not provide specific guidance and suggest working with nutritionists to make the appropriate decisions relevant to a particular study.

We would, however, like to note that the DRIs age-sex disaggregation for which the EER formulas are defined do not perfectly align with the harmonized groups. The EFSA average requirements for energy would be more straightforward to use. After selecting an energy requirement, the AMDRs do not present a conflict between the requirements and the harmonized age-sex groups (with the exception of lipids for 18-year-olds) and can be applied if desired based on percentage of calories from each macronutrient.

**Use case**
As detailed in related publications<sup>4,9</sup>, the Cost of Nutrient Adequacy (CoNA) is an index of the least-cost nutritionally adequate diets for a specific place and time based on observed market retail food prices<sup>1</sup>. The cost of a nutrient-adapted diet tracks whether a food system is capable of providing the appropriate mix of foods at an affordable total cost, necessary for an active and healthy life. This appropriate mix of foods is the one that meets all the minimum nutrient needs (EARs and AMDR lower bounds), does not exceed any upper limits (ULs, CDRR, and AMDR upper bound) and meets overall energy needs as defined herein. Since nutrient needs vary by age, sex, and life stage, even considering only the active level of physical activity recommended for health, the lowest cost diet that meets those needs will vary depending on the population of interest. The related publications take the case of non-pregnant women of reproductive age (19–30) as the benchmark population of interest. This population is of particular concern for global nutrition for a number of reasons. First, these women are nutritionally vulnerable themselves and often suffer a high prevalence of poor nutrition outcomes such as anemia, underweight and overweight. Second, the health of a mother at conception and during pregnancy and lactation is a strong determinant of child birth and early life growth outcomes<sup>24,41</sup>.  

---

<sup>1</sup>Part 2
<sup>2</sup>Part 3
To calculate a lowest-cost diet requires observed market retail food prices and food composition data in addition to nutrient requirements. Using the nutrient requirements as constraints for linear programming will identify the quantity and combination of specific foods that together meet all the requirements without exceeding any upper bounds at the lowest total cost. Supplement 4 or 5\textsuperscript{3}\textsuperscript{7} software code compiles and shapes the nutrient requirements for women 19–30 at an active level of physical activity for use in a least-cost diet application such as Bai et al.\textsuperscript{3}. The authors find the median cost of a nutritionally adequate diet across 177 countries to be $1.35 per day for women 19–30 in 2011.

**Conclusions**

We provide readily usable data files containing the Dietary Reference Intakes, the WHO Child Growth Standards and WHO Growth References for School-Aged Children and Adolescents percentiles tables of anthropometric measures, nutrient needs for children 6–23 months, and recently proposed harmonized average requirements and upper levels. We describe the methodology and considerations for applications of human nutrient requirements in social science research. We provide Stata and R software code to compile nutrient requirements based on WHO growth standards and references with the appropriate adjustments for children 6–23 months during continued breastfeeding. Finally, we demonstrate a use case where least-cost nutritionally adequate diets are calculated for non-pregnant women of reproductive age.

**Data availability**

**Underlying data**

All data underlying the results are available as part of the article and no additional source data are required.

**Extended data**

Harvard Dataverse: Data files for “Software tools for practical application of human nutrient requirements in food-based social science research”. https://doi.org/10.7910/DVN/5GIMLE\textsuperscript{39}.

This project contains the following extended data:

- Supplement 1 (contains 12 datafiles)
  - 1_NutrientRequirements_Notes.pdf
  - 1_NutrientRequirements_Definitions.csv
  - 1_NutrientRequirements_UnitsNotes.csv
  - 1_NutrientRequirements_ReferenceValues.csv
  - 1_NutrientRequirements_Energy.csv
  - 1_NutrientRequirements_MicronutrientsEAR.csv
  - 1_NutrientRequirements_MicronutrientsRDA.csv
  - 1_NutrientRequirements_MicronutrientsUL.csv
  - 1_NutrientRequirements_MacronutrientsEAR.csv
  - 1_NutrientRequirements_MacronutrientsAMDR.csv
  - 1_NutrientRequirements_ProbIronInadequacy.csv
- Supplement 2 (contains 14 data files):
  - 2_WHOGrowthCharts_Notes.pdf
  - 2_WHOGrowthCharts_WeightsHeightsbyDRI-Group.csv
  - 2_WHOGrowthCharts_EnergyEER.csv
  - 2_WHOGrowthCharts_ProteinEAR.csv
  - 2_WHOGrowthCharts_BMI5–19Boys.csv
  - 2_WHOGrowthCharts_HFA5–19Boys.csv
  - 2_WHOGrowthCharts_BMI5–19Girls.csv
  - 2_WHOGrowthCharts_CPU0–5Boys.csv
  - 2_WHOGrowthCharts_CPU0–5Girls.csv
  - 2_WHOGrowthCharts_BMI5–19Boys.csv
  - 2_WHOGrowthCharts_BMI5–19Girls.csv
- Supplement 3 (contains two data files):
  - 3_NutrientRequirements6–23months_Notes.pdf
  - 3_NutrientRequirements6–23months_6–23moFood-Needs.csv
- Supplement 6 (contains five data files):
  - 6_HarmonizedNutrientRequirementsAllenEtAl2020_Notes.pdf
  - 6_HarmonizedNutrientRequirementsAllenEtAl2020_UnitsNotes.csv
  - 6_HarmonizedNutrientRequirementsAllenEtAl2020_AgeSexGroups.csv
  - 6_HarmonizedNutrientRequirementsAllenEtAl2020_H-AR.csv
  - 6_HarmonizedNutrientRequirementsAllenEtAl2020_H-UL.csv

Extended data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

**Software availability**

Source code in Stata and R files available from: https://github.com/KateSchneider-FoodPol/Nutrient-Requirements-Software-Tools.
Author contributions
Kate Schneider and Anna Herforth prepared the manuscript, together conceptualized the intent, and determined the final set of recommendations contained herein. Kate Schneider developed the data files and software code.

Acknowledgements
We acknowledge the foundational work carried out by Will Masters, Yan Bai, Anna Herforth and others in the initial development of the Cost of Nutritious Diets (CoNA) price index. We thank Zhining Sun for translating the Stata code we prepared into R syntax. We are grateful to Lila Cardell for her early user feedback.

References


The authors provide a nice set of data and scripts in R and Stata for computations that can facilitate the integration of considerations around cost of meeting nutrient needs into nutrition-sensitive programs across sectors such as agriculture, social protection, economics.

I have the following suggestions for further improvement of the paper, the data included in the files, and the recommended use of those:

1. **Additional data are required.** The motivation to share the data and scripts is to enable calculation of CoNA, the cost of nutrient adequacy. In addition to these files on nutrient requirements, one will also need food composition data, food price data and maximum portion sizes per food (or food type or food group). It would be good to state the need for these additional data early on and to provide some recommendations on how to source those or to refer to other documentation for further guidance in that regard.

2. **Set the nutrient intake target at RDA/RNI rather than EAR/H-AR.** The EAR or H-AR defines the intake level that is adequate for 50% of the population and hence inadequate for the other 50%. For assessments of adequacy of a population’s intake, the proportion with inadequate intake is defined as the proportion that has an intake below the EAR/H-AR. For intake planning purposes, e.g. fortification or provision of food in institutions, population intake is planned such that no more than 2.5-10% have an intake below the EAR/H-AR and no more than e.g. 5% an intake above the UL (e.g. see Murphy et al, Adv Nutr 2020;00:1-9; doi: https://doi.org/10.1093/advances/nmaa119).¹ The name of the indicator ‘Cost of Nutrient Adequacy’ does not suggest that it would be a cost at which there is only a 50% chance that intake is adequate. The target median intake (TMI) of a distribution that sits between EAR and UL, and would hence be adequate and safe for almost everyone, depends on the population’s intake distribution and is hence population specific. In the absence of such values for a specific population of interest, the RNI/RDA, and AI where the EAR and RDA have not been defined, are a reasonable target intake to set for a population to ensure that the individuals in the population meet their own biological requirements as
this level would be adequate for 97.5% of them. When defining CoNA as the lowest cost combination of foods that meets the requirements for macro- and several micro-nutrients concurrently, and setting the lowest level of nutrient requirements at the EAR, the actual nutrient level may reach the RNI/RDA value for most nutrients, because foods contain combinations of nutrients. However, it is important to recommend that users work with the RDA value (from IOM) for micronutrients because EAR is inadequate for 50% of the population, and users may do calculations for only one or a few micronutrients rather than for a comprehensive set.

3. **Protein requirements should be adjusted for biological value.** Protein requirements as defined by WHO and expressed as g/kg/d are for high biological value protein, i.e. it assumes a good balance of essential amino acids and good bioavailability, equivalent to a Protein Digestibility Corrected Amino Acid Score (PDCAAS) of 1.0 (WHO technical report 935: https://apps.who.int/iris/bitstream/handle/10665/43411/WHO_TRS_935_eng.pdf?ua=1). The recommendation to consume approximately 12 percent of energy in the form of protein takes lower bioavailability and lower quality of protein into account. See also p219 of the WHO report where a PDCAAS of 0.5-0.8 is mentioned for mixed diets in developing countries. Hence, protein intake recommendations need to be adjusted, either to reflect 12 energy percent, which is a rough measure or by a score that reflects the likely PDCAAS of the diet, where a value between 0.5-0.8 may be recommended. Recommending the involvement of a nutritionist, which the authors have also recommended for other nutrients and target groups, is also good in this case.

4. **Mineral bioavailability needs to be adjusted based on type of diet.** While bioavailability of iron and zinc is indeed an area of active research, there are also recommendations for intake depending on whether the diet provides for low, medium or high bioavailability. These should be applied as appropriate, and nutritionists can guide decisions in this regard.

5. **Vitamin A: splitting retinol from total VA content and compatibility of RE (requirement) and RAE (content).** The authors’ recommendation to use vitamin A intake requirements in combination with the UL for retinol is an elegant way to get around the problem that provitamin A carotenoids will not be converted to retinol unless the body needs it and should therefore not be counted towards the intake level that should remain below the UL. However, as most food composition tables do not make this distinction, it would be good to recommend that the user counts only vitamin A from animal source foods and foods fortified with provitamin A towards retinol intake. Also, while RAE has been introduced to take the lower bioavailability from provitamin A carotenoids from vegetables and fruits into account, the fact that intake requirement is still expressed as RE has led to confusion among some people. Therefore, it would be helpful to state that 1 RE of vitamin A intake requirement equals 1 RAE of vitamin A content from foods.

**Questions for clarification:**

1. Do I understand correctly that the recommended weight and height achieved at the age of 19 is recommended to be used as the reference value for dietary energy intake? If so, what would the authors recommend to users who would like to use actual average population weights to derive those values?

2. Under use case, last line, it would be good to also reference the cost of a nutrient adequate
diet as published in the SOFI 2020 report, from Herforth et al. (your ref 38).

References

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Partly

*Competing Interests:* No competing interests were disclosed.

*Reviewer Expertise:* Nutrition, including modelling of diets for meeting nutritional needs at the lowest cost and assessment of affordability of diets to inform situation analysis for improving nutrition through multi-sectorial pathways

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.
The article presents a software tool to compile human nutrient requirements for application in population-based dietary analyses. The tool is specifically designed for application in the computation of Cost of Nutrient Adequacy (CoNA) least-cost diets, but it is expected to have "broad utility for numerous social science researchers wishing to incorporate nutrition into agricultural, development, and food systems analyses" (p.3). The article focuses on describing the methods and data used in the tool.

The article is concise and well-written overall. Yet it requires major revisions and needs to clearly answer the following questions or address them in greater detail and specificity: What is new about the software tool, and how does it differ from and complement similar tools/databases? Is the provided “tool” indeed a software tool, or is it rather a database? Why is it relevant/useful, for whom, and for which type of analysis and in which context? What are the limitations of the tool, and which data (and level of data quality) are needed for practical application in dietary analyses?

My main comments are detailed below. I provide additional, specific comments and suggestions in an annotated version of the article. I hope that my comments and suggestions are helpful for the revision of the article, and I encourage the authors to reach out to me for questions of clarification if needed.

Main comments:

1. Novelty: There are several software tools and databases available for download or forthcoming (often for free) that provide human nutrient requirements. Many of them go beyond providing nutrient requirement levels and include tools for e.g., assessing nutritional needs, analyzing dietary survey data, defining nutrient gaps, and determining “optimal” individual diets and healthy and environmentally sustainable menus/recipes. Examples of such tools and databases include the Nutrition Surveys and Calculations ("NutriSurvey") program, the Optifood tool developed by the FANTA program, and the Optimeal tool developed by Blonk Consultants. Why is this software tool novel; how does it differ from previous tools/databases; and which features present technical advances? How does this software tool complement other tools and databases such as for assessing nutrient intakes and dietary adequacy? These tools/databases include the Global Nutrient Database and Global Dietary Database (that are both supported by the Bill and Melinda Gates Foundation as well). The Introduction section should be expanded in this regard.

2. Terminology: The authors use a few terms that are likely to be unclear to a broad audience or appear to be imprecise and therefore may be misleading. The authors should consider using different terms or clarify the used terms. The most critical cases (occurring in the article title) include:
   a. “Software tools”: In my understanding, the authors compile a database of nutrient requirements, saved in several data files, rather than developing a software tool. The files provided by the authors do not allow to produce analytical results. Hence, the authors do provide a resource that serves as input into an external analysis. The codes that compile the different data files and convert them to a format amenable to using in e.g. the CoNA least-cost diet analysis are provided in form of Stata and R execution files (editable by users). They are not incorporated into a (closed) program/function that independently produces results by execution. Apart from that, it appears that only one “tool” is provided.
   b. “Food-based social science research”: I have not seen the term “food-based social science”
before, and a Google search suggests that this term has been used only by the authors.

3. Relevance/usefulness and scope of application: Statements in the article regarding the relevance/usefulness of the software tool and the scope of its application are little specific and insufficiently comprehensive, beyond its original core purpose for use in the CoNA least-cost diet analysis. The “broad utility for numerous social science researchers wishing to incorporate nutrition into […] [their] analyses” needs to be demonstrated, and the potential user groups should be better defined. This point should be addressed in a revised Introduction section. It may be helpful especially to the (targeted) non-expert readership to provide (prominent) examples of the stylized research questions, types of analyses, and study contexts for which the tool is useful. Note that it may be faster (and, to some degree, more flexible) to directly extract nutrient requirements from the original sources in analyses that examine intake/consumption of a single nutrient or a small set of nutrients. So, when is the use of the software tool of advantage? Furthermore, in the Introduction, the authors may want to consider to first define the broader scope of the tool's application before turning to the specific use case—i.e. the CoNA least-cost diet analysis, and not vice versa.

4. Article purpose: The purpose of the article should be better specified in the Introduction, and its execution can be optimized in this respect. Beyond documenting the methods and data used in the tool, the article seems to primarily serve as a guide for practitioners that are looking for an easy way to incorporate nutrient requirements in their analysis. Hence, the article would benefit from a clearer, more intuitive structure (see next point); more practical recommendations (based on the authors' experience in this field of work); references to additional resources that may be relevant during the analysis; detailed instructions of how to install and operate the tool and potential computational requirements; and a discussion of limitations (see below).

5. Article structure: The structure of the article can be improved in several instances. This will likely increase the accessibility of its content to non-experts. I suggest including a section before the Methods section (or as an introductory Methods subsection) that provides an overview of the tool's methodological approach, which could be nicely summarized in a flowchart. This outline could then be used to structure the (remainder of) the Methods section. The structure of the Methods section is a bit confusing to me, as it first lists the sources of several nutrient requirement levels before explaining what they mean and how they are used. This first paragraph is also unclear about which macro- and micronutrient and dietary energy requirement levels are obtained from which source. The first two paragraphs of the following description of nutrient requirements provide justifications for the used sources, before turning to the explanations of nutrient requirement levels by micro- and macronutrients and dietary energy. In the latter parts, the sources of nutrient requirements are not always apparent immediately. Moreover, I am wondering why the authors first discuss micronutrient requirements before turning to dietary energy and then macronutrient requirements. Starting with energy and macronutrients before turning to micronutrients seems to be more logical.

6. Limitations: The article misses to specify what the tool cannot do or is unsuited for. The authors may also want to address caveats, which can help to shield the tool from potential (incorrect) attributions of implausible results of dietary analyses that used the tool for computing nutrient requirements. Thus, a section describing the methodological limitations
of the tool and specifying the data requirements for analyses using the tool (including data quality/suitability issues) should be included. Issues that should be addressed pertain to these questions: How are impaired nutrient absorption and loss due to e.g. infectious diseases considered? How is nutrient loss due to food preparation methods considered? What are potential problems of using the tool in combination with nutrient consumption amounts derived from food consumption data in common household surveys (such as LSMS-type surveys) to estimate nutrient consumption gaps, prevalence rates of nutrient deficiencies, etc.?

7. Maintenance and further development: More broadly, I am wondering about the authors'/developers' plans to maintain and perhaps further advance the tool. Several similar tools (including FAO's WorldFood Dietary Assessment System) are not functioning anymore (possibly because of a lack of maintenance). Can the authors provide such an outlook (possibly in the Conclusions section)? Additionally, it may be useful to establish user feedback channels.

Other issues:
There are several other passages in the article that should be revised. These include:
1. Abstract: The abstract is too long and too detailed in its overview of the used sources of nutrient requirements. It should also be revised to attract a broader audience (including researchers that may not have initially considered to incorporate dietary issues in their analysis).

2. Conclusions: The last section is a short summary of the article. I wish to see some concluding remarks by the authors.

3. There are several (minor) typos that should be corrected. Some of them are highlighted in the annotated manuscript.

Is the rationale for developing the new software tool clearly explained?
No

Is the description of the software tool technically sound?
Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
No
**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Agricultural economics, conflict, demand analysis, food systems, nutrition, developing countries

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

---

**Comments on this article**

**Version 1**

Reader Comment 06 Dec 2021

**Stacia Nordin**, NeverEndingFood, Lilongwe, Malawi

A limiting factor in food composition tables - especially using USDA tables - are indigenous foods from around the world that are not yet in the tables. This should be listed in both the limitations and the recommendations to expand food composition tables to include all indigenous foods.

**Competing Interests:** Passionate about indigenous resources.