Estimating national and subnational nutrient intake distributions of global diets

Simone Passarelli,1 Christopher M Free,2,3 Lindsay H Allen,4 Carolina Batis,5 Ty Beal,6,7 Anja Pia Biltoft-Jensen,8 Sabri Bromage,4 Ling Cao,9 Analí Castellanos-Gutiérrez,5 Tue Christensen,5 Sandra P Crispim,10 Arnold Dekkers,11 Karin De Ridder,12 Selma Kronsteiner-Gicevic,1,13 Christopher Lee,14 Yanping Li,1 Mourad Moursi,15 Isabelle Moyerseon,12 Josef Schmidhuber,16 Alon Shepon,4,17 Daniel F Viana,1,18 and Christopher D Golden1,19,20

1Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA, USA; 2Bren School of Environmental Science and Management, University of California, Santa Barbara, Santa Barbara, CA, USA; 3Marine Science Institute, University of California, Santa Barbara, Santa Barbara, CA, USA; 4ARS Western Human Nutrition Research Center, USDA, Davis, CA, USA; 5Nutrition and Health Research Center, National Institute of Public Health, Cuernavaca, Morelos, Mexico; 6Department of Environmental Science and Policy, University of California, Davis, Davis, CA, USA; 7Global Alliance for Improved Nutrition, Washington, DC, USA; 8Division of Food Technology, National Food Institute, Technical University of Denmark, Lyngby, Denmark; 9School of Oceanography, Shanghai Jiao Tong University, Shanghai, China; 10Department of Nutrition, Federal University of Paraná, Curitiba, Brazil; 11Centre for Nutrition, Prevention and Health Services, National Institute for Public Health and the Environment, Bilthoven, The Netherlands; 12Department of Epidemiology and Public Health, Sciensano, Brussels, Belgium; 13Institute for Statistics of the Federation of Bosnia and Herzegovina, Sarajevo, Bosnia and Herzegovina; 14Harvard College, Cambridge, MA, USA; 15Intake, Center for Dietary Assessment, FHI Solutions, Washington, DC, USA; 16Trade and Markets Division, UN’s Food and Agricultural Organization, Rome, Italy; 17Department of Environmental Studies, The Porter School of the Environment and Earth Sciences, Tel Aviv University, Tel Aviv, Israel; 18Betty and Gordon Moore Center for Science, Conservation International, Arlington, VA, USA; 19Department of Environmental Health, Harvard T. H. Chan School of Public Health, Boston, MA, USA; and 20Department of Global Health and Population, Harvard T. H. Chan School of Public Health, Boston, MA, USA

ABSTRACT

Background: Access to high-quality dietary intake data is central to many nutrition, epidemiology, economic, environmental, and policy applications. When data on individual nutrient intakes are available, they have not been consistently disaggregated by sex and age groups, and their parameters and full distributions are often not publicly available.

Objectives: We sought to derive usual intake distributions for as many nutrients and population subgroups as possible, use these distributions to estimate nutrient intake inadequacy, compare these distributions and evaluate the implications of their shapes on the estimation of inadequacy, and make these distributions publicly available.

Methods: We compiled dietary data sets from 31 geographically diverse countries, modeled usual intake distributions for 32 micronutrients and 21 macronutrients, and disaggregated these distributions by sex and age groups. We compared the variability and skewness of the distributions and evaluated their similarity across countries, sex, and age groups. We estimated intake inadequacy for 16 nutrients based on a harmonized set of nutrient requirements and bioavailability estimates. Last, we created an R package—nutriR—to make these distributions freely available for users to apply in their own analyses.

Results: Usual intake distributions were rarely symmetric and differed widely in variability and skewness across nutrients and countries. Vitamin intake distributions were more variable and skewed and exhibited less similarity among countries than other nutrients. Inadequate intakes were high and geographically concentrated, as well as generally higher for females than males. We found that the shape of usual intake distributions strongly affects estimates of the prevalence of inadequate intakes.

Conclusions: The shape of nutrient intake distributions differs based on nutrient and subgroup and strongly influences estimates of nutrient intake inadequacy. This research represents an important contribution to the availability and application of dietary intake data for diverse subpopulations around the world. Am J Clin Nutr 2022;00:1–10.

Keywords: nutrient, dietary data, intake, global health, methods, subgroup, distribution, epidemiology, nutrient intake, nutrition

Supported by National Institutes of Health (NIH) Training Grant 2T32DK007703-26 in Academic Nutrition as well as National Institutes of Health (NIH) grant D43 TW010543.

CDG is an Associate Editor of the American Journal of Clinical Nutrition and played no role in the Journal’s evaluation.

Supplemental Tables 1–4 and Supplemental Figures 1–8 are available from the “Supplementary data” link in the online posting of the article and from the same link in the online table of contents at https://academic.oup.com/ajcn/. SP and CMF are co-first authors and contributed equally to this work. Address correspondence to SP (e-mail: spassarelli@hsph.harvard.edu). Abbreviations used: AR, Average Requirement; EAR, Estimated Average Requirement; GDD, Global Dietary Database; SPADE, Statistical Program to Assess usual Dietary Exposure.

Received February 8, 2022. Accepted for publication April 19, 2022.

First published online 0, 2022; doi: https://doi.org/10.1093/ajcn/nqac108.

1

Am J Clin Nutr 2022;00:01–10. Printed in USA. © The Author(s) 2022. Published by Oxford University Press on behalf of the American Society for Nutrition. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Introduction

Access to robust dietary intake data has been a persistent limitation in the field of nutrition research (1). Individual-level dietary data sets are scarce because of the large effort and expense involved in data collection. It is especially difficult to find data sets describing an individual’s usual intake, which requires > 1 d of dietary recall to calculate. Even when these data are collected, the resulting data sets are often proprietary. Data availability may also be limited to a mean or median level of intake of a given nutrient in a country rather than a full population distribution. Altogether, these barriers inhibit our understanding of nutritional vulnerabilities of countries and their subpopulations (2).

Consequently, researchers must often rely on coarser food supply or household expenditure data (1). Although these sources provide invaluable information about trends in the average supply of nutrients, they mask important differences in food allocation patterns among sex and age subgroups (3, 4) and are insufficient for estimating the burden of disease risk and dietary inadequacies or excesses. For this, researchers need information on the distribution of usual intakes within sex and age subgroups to assess the proportion of subpopulations at risk (5).

In recent years, there have been significant advances in access to individual-level dietary intake data (6–9). For example, the Global Dietary Database (GDD) project has compiled nutrition data from >1600 surveys (10), representing one of the greatest contributions to democratizing access to nutrition data to date. Still, ~14% of these data sets are not representative at the national or subnational level, and about a quarter are not publicly available.

Dietary data from 24-h recalls or food diaries are the best available data for population-level dietary surveillance because they estimate absolute intake with greater precision (11). However, 1 d of recall does not represent usual intake (long-term intake free from day-to-day variation) (12, 13). Collecting a second day of intake in at least a subsample makes it possible to differentiate within- and between-person variation and use statistical procedures to estimate usual intake distributions that exhibit only the between-person variation. Although other methods have been developed for estimating usual intake without multiple days (14, 15), these approaches have limitations for accurate global estimation, especially in low-income populations in which many foods are not consumed daily (16). Furthermore, these data are not consistently disaggregated by sex and age groups, although there is ample evidence showing that consumption patterns differ across dietary factors, subgroups, and countries (17–19).

The objective of this study was to fill these gaps by harnessing data from geographically and demographically diverse surveys. Using individual dietary intake data from food consumption surveys incorporating ≥ 2 recall days, we calculated usual nutrient intake distributions for age and sex subpopulations in 31 countries for 32 micronutrients and 21 macronutrients. We then found the best-fitting distribution shape to describe the intake patterns for each sex–age group within a country. We compared these distributions to highlight differences in intake patterns across country, sex, and age groups based on their mean, variance, and skewness and to show how shape is associated with inadequate intakes. To make these data and distributions accessible and applicable for future research, we developed an R package called “nutriR” as an accompaniment to this article.

Methods

Diet recall data

Most data were obtained from publicly available sources, including the GDD (10) and the FAO/Global Individual Food consumption data Tool website (20) (Figure 1; Supplemental Table 1; Supplemental Figure 1). For some countries, including the Netherlands, Brazil, Belgium, Denmark, China, and Mexico, data were made available upon request. We reviewed data sets to determine their eligibility for inclusion in our analysis, based on whether they had 1) individual-level dietary data, 2) calculated nutrient-level data, 3) ≥ 2 d of dietary intake (for at least some participants), 4) data based on a 24-h recall or diet record/food diary, and 5) a sample size ≥ 200. If there was ≥ 1 nationally or regionally representative survey for a given country, we selected the most recent data set. To our knowledge, we included all publicly available data sets meeting our parameters at the time of this analysis.

Deriving usual intake distributions

In general, ≥ 2 d of dietary recall data are required to estimate the proportion of total variance that is attributable to within-individual variation (16). Although a recently developed National Cancer Institute methodology uses only 1 d of dietary information to estimate usual intake distributions (15), this approach faces several limitations in the context of global assessments. First, it requires that foods are consumed almost daily, which is often not the case in low-income countries for foods such as fruit, meat, fish, and eggs. Second, it is highly sensitive to the variance ratio selected, and large variations in these ratios have been observed across populations (16). At least some members of a population should have ≥ 2 d of dietary recall data to generate an unbiased estimate of usual intake, especially when populations have different demographic characteristics (16). Because one of the purposes of our analysis was to estimate usual intake distributions among diverse populations and subgroups, we chose to only include data sets with a minimum of 2 d of dietary data for at least some members of the population.

We used the Statistical Program to Assess usual Dietary Exposure (SPADE, v4.1.0) to estimate usual intakes of 21 macronutrients and 32 micronutrients from the available survey data (Supplemental Table 2) (20). SPADE is an R-based software package that models usual intake distributions based on repeat 24-h dietary recall data by 1) transforming data to a normal distribution, 2) removing within-person variability, 3) estimating usual nutrient intake distributions as a function of age, and 4) back-transforming the data onto its original scale (21, 22). We chose to use this statistical model instead of the raw dietary data (or the mean intakes across multiple days) because without removing within-person variation, the variance of the intake distribution would be higher than the true variance. As a result, this can lead to highly biased estimates of the proportion of the population above or below a reference value, especially when a reference is closer to the tails of a distribution (12).
Inclusion criteria

Datasets had:
• Individual-level dietary data
• Calculated nutrient-level data
• Two or more days of intake (for at least some respondents)
• A 24-hour recall or diet record/food diary format
• Sample size >200
• Variables for age and sex

31 total datasets included

GDD
Canada, n = 20,483
Ecuador, n = 650
Estonia, n = 3049
Ethiopia, n = 323
Germany, n = 921
Guatemala, n = 226
Kenya, n = 718
Mozambique, n = 543
Pakistan, n = 270
Portugal, n = 6424
Sweden, n = 4288

FAO/WHO GIFT
Bangladesh, n = 475
Bulgaria, n = 1723
Burkina Faso, n = 960
Democratic Republic of Congo, n = 214
India, n = 242
Italy, n = 3323
Lao PDR, n = 2045
Malaysia, n = 573
Philippines, n = 1203
Romania, n = 1382
Uganda, n = 551
Zambia, n = 771

NHOES
United States, n = 7483

Individual Research Partners
Belgium, n = 3146
Bosnia and Herzegovina, n = 853
Brazil, n = 46,164
China, n = 10,197
Denmark, n = 3806
Mexico, n = 4,343
The Netherlands, n = 6,331

FIGURE 1 Flowchart of data sets included in this analysis. For the full references and survey details, please see Supplemental Table 1. The n listed refers to the sample size of participants in each survey who had at least 1 d of recall. GDD, Global Dietary Database; GIFT, Global Individual Food consumption data Tool.

The calculation of inadequate intakes using the probability method (see details below) requires that intake distributions be described as probability density functions. We therefore used the fitdistrplus R package v1.1-6 (23) (in R v4.1.1) to fit gamma and log-normal distributions to the usual intakes of all available sex–age groups (Supplemental Figures 2 and 3). We defined age groups in 5-y intervals from 0–5 y to 95–99 y to align with the GDD (2) and other common demographic data sets. We selected the distribution with the best Kolmogorov–Smirnov goodness-of-fit statistic as the final distribution for each group (Supplemental Figure 4). We considered gamma and log-normal distributions because they are continuous probability distributions bounded between zero and positive infinity that accommodate the skewed (asymmetric) shapes common among these types of data.

Describing and comparing usual intake distributions

We quantified and compared the shape of the fitted usual intake distributions using 2 metrics: 1) the CV, which measures the relative variability of a distribution, and 2) the skewness, which measures the asymmetry of a distribution. Both metrics are comparable across scales and distribution types. Although we also calculated excess kurtosis—which measures the tailedness of a distribution—we omitted these results from the article because of their correlation with skewness; however, the kurtosis measures are included in the “nutriR” data package described below. Using the equations in Supplemental Table 3, we calculated these metrics for each usual intake distribution. These quantities were derived from the fitted distributions, rather than from the distributions of the pseudo-population modeled through the SPADE analysis, to maintain consistency throughout the analysis workflow. If distributions had extremely low variability (CV <0.01; Supplemental Figure 5), they were considered unrepresentative of population-wide usual intakes and omitted from the analysis.

To compare the similarity of usual intake distributions among countries, sex, and age groups, we measured the percent overlap between pairwise combinations of distributions within groups. Percent overlap between paired distributions was assessed using the Bhattacharyya coefficient (21), in which a value of 0% indicates entirely divergent distributions and a value of 100% indicates perfectly identical distributions. We evaluated the similarity in distributions among countries as the median percent overlap of all unique pairwise combinations of countries within each sex–age group. To evaluate similarity in distributions between sexes, we considered the median percent overlap in each age group across all countries. To assess similarity in distributions among age groups, we calculated the median percent overlap of all unique pairwise combinations of age groups within each country–sex group.
Estimating prevalence of inadequate intakes

To evaluate the implications of the shape of usual intake distributions for achieving nutrient adequacy, we related metrics of distribution shape to the estimated prevalence of inadequate intakes. We estimated population-level intake inadequacy, also known as the summary exposure value (24), for 16 nutrients with estimated nutrient requirements using both the probability method (25) and the Estimated Average Requirement (EAR) cut-point method (hereafter referred to as the cut-point method) (26). The probability method compares intake distributions with a continuous relative risk curve, whereas the cut-point method is a simplification of the probability method and assumes that the prevalence of inadequacy of a certain population can be estimated by assessing the fraction whose intake is below a certain cut-point value (the EAR). The cut-point method does not require knowledge of the shape of the intake and requirement distributions, only that these distributions adhere to a set of criteria. These include that the shape of the requirement distribution be symmetrical around the EAR and that the variance of the intake distribution be larger than the variance of the requirement distribution (5, 26).

In both methods for estimating intake inadequacy, we used the Average Requirements (ARs) provided by Allen et al. (27) as the estimated nutrient requirements (Supplemental Figure 6). Allen et al. (27) reviewed alternative sources of nutrient reference values and identified which were most up-to-date and most appropriate for use in global-scale studies. We selected ARs for iron and zinc based on their bioavailability in national diets. We used the Human Development Index, a summary measure of average achievement in key dimensions of human development (28), as an indicator of likely dietary availability (Supplemental Table 4). Although average nutrient requirements are available for vitamin D, they assume no sun exposure, so we do not calculate inadequate vitamin D intakes here.

The continuous risk curves of the probabilistic approach have a value of 1 at low intakes, 0.5 at the relevant EAR, and zero at large intakes. These absolute risk curves are based on the cumulative normal distribution function of requirements (26) with a mean at the EAR and a CV of 0.25 for the AR for vitamin B-12 and a CV of 0.10 for the AR of all other nutrients (29).

The “nutriR” R package

We developed an open-access R package called “nutriR” to share our data and facilitate its use by other public health researchers. This package contains a series of functions that allow the user flexibility in retrieving and plotting distributions, shifting a distribution around a new mean (e.g., in response to an intervention or based on a user’s own data set), estimating inadequacy of nutrient intakes based on ARs, and characterizing and comparing the shapes of distributions across populations using descriptive statistical parameters. The R package, user manual, and an interactive R Shiny web application for exploring the fitted subnational nutrient intake distributions are all available on GitHub (https://github.com/cfree14/nutriR).

Results

Our data compilation resulted in individual-level dietary intake data with calculated nutrients from 31 countries and 5 continents (Supplemental Figure 1; Supplemental Table 1). The methods of recall included a mix of food diaries and 24-h recalls, generally based on the literacy of the surveyed populations. The year of data collection varied by country, from as early as 2002 in the case of the Philippines to as recently as 2018 in the United States. Most data sets (23 of 31) employed a 24-h recall methodology. The level of representativeness varied, with 18 data sets representative at the national level and the remaining representative at a lower administrative level.

Our results show that there are large differences in the shape of usual intake distributions across nutrients, countries, sexes, and age groups and that these differences have profound impacts on the prevalence of inadequate nutrient intakes. Figure 2 provides a stylized example to illustrate the concepts of nutrient intake...
The shape of usual intake distributions varied widely across nutrients and between sexes. We compared usual intake distributions in terms of their variability, measured as their CV, and their asymmetry, measured as their skewness (Figure 3A). Although usual intakes of minerals such as phosphate, potassium, and magnesium generally exhibited low variability (CV < 0.5) and asymmetry (skewness < 1.0), usual intakes of vitamins such as vitamins K, D, B-12, A, and C generally exhibited high variability and asymmetry (Figure 3A). Vitamin intake distributions also exhibited a wider variety of shapes than mineral distributions and most fatty acid and other macronutrient intake distributions (Figure 3A). The exceptions were the intake distributions for omega-3 fatty acids and β-carotene, which were highly diverse, variable, and skewed (Figure 3A), reflecting subpopulations dominated by very low usual intakes but with rare instances of very high usual intakes. Figure 3B, C illustrates how increasing the variability and skewness, respectively, of usual intake distributions explored in our analysis. Figure 2A shows how the type of distribution (normal, gamma, or log-normally distributed) can result in different estimates of inadequate intake, even when the mean intake is identical. Figure 2B similarly shows this concept but for the underlying assumptions about the variability of a distribution. These hypothetical examples illustrate why distributions matter when estimating usual nutrient intake for a population.
intake distributions affects distribution shape and estimates of intake inadequacy.

The similarity in intake distributions varies by nutrient, country, sex, and age

The similarity of usual intake distributions varied more by country and sex than by age group. We measured the similarity of pairs of distributions as their percent overlap and the similarity of groups of distributions as the median percent overlap of all unique pairs (Figure 4A). The usual intake distributions of vitamin K, retinol, ω-3 fatty acids, and folate exhibited the least similarity across countries (see Figure 4B using vitamin K as an example). On the other hand, intakes of minerals—especially magnesium, phosphorus, and potassium—were very similar across countries (see Figure 4D using magnesium as an example). In general, intakes of the remaining micronutrients exhibited moderate similarity across countries (see Figure 4C using calcium as an example). Usual intakes of niacin; macronutrients such as protein, energy, and alcohol; and minerals such as selenium, sodium, zinc, and phosphorus were especially different between sexes within countries and age groups (Supplemental Figure 7), whereas intakes of vitamins C, D, E, and K were generally similar between sexes (Supplemental Figure 7). Usual nutrient intake distributions were, in comparison, more similar among age groups within country and sex groups (Supplemental Figure 8).

Prevalence of inadequate intake by nutrient and country

We measured and compared the prevalence of inadequate intakes for the 16 nutrients with harmonized ARs (27) using both the probability method and the cut-point method (5, 25, 26). The probability method accounts for the shape of
intake and requirement distributions by comparing actual intake distributions with a relative risk curve of requirement. The cut-point method simplifies the probability method by assuming that the fraction of the population’s intake below a cut-point can determine the prevalence of inadequacy if various criteria about the requirement and intake distributions are met.

Using the probability method (25), we generated estimates of subnational nutrient intake inadequacy for the 31 evaluated data sets. There was acute and geographically widespread intake inadequacy for calcium, folate, vitamin E, and magnesium (Figure 5A). We found moderate and geographically dependent intake inadequacies for vitamin A (retinol activity equivalents), vitamin C, thiamin, zinc, vitamin B-12, riboflavin, and vitamin B-6 (Figure 5A). Copper, iodine, niacin, and iron exhibited comparatively lower intake inadequacies. (Figure 5A). Across nutrients, we observed a high prevalence of intake inadequacy in Bangladesh, Laos, Ethiopia, the Philippines, China, and Mexico and a comparably lower prevalence of intake inadequacy in Denmark, Romania, the United States, Canada, Italy, and Sweden (Figure 5A).

The shape of intake distributions influences estimates of intake inadequacy

The shape of usual intake distributions strongly determines the prevalence of inadequate intakes estimated using the probability method (Figure 5B). Subpopulations with usual intake distributions that have identical means but different levels of variability experienced different levels of intake inadequacy (Figure 5B). When mean intakes were below the average requirement, high variability was associated with a lower prevalence of intake inadequacy, but when intakes were above the average requirement, higher variability was associated with dramatically higher intake inadequacies (Figure 5B).

The cut-point method, which makes some simplifying assumptions regarding the intake and requirement distributions,
estimated levels of inadequate intake similar to the probability method when properly accounting for the asymmetry of usual intake distributions (Figure 5D). However, when assuming symmetric usual intake distributions, the cut-point method estimated different levels of intake inadequacy relative to the probability method (Figure 5C). Differences were mostly between 0% and 25%, but in some extreme cases, differences of 50% were observed. When mean intakes were below the average requirement, the cut-point method (assuming normally distributed usual intakes) produced higher estimates of intake inadequacy compared with the probability method and painted a gloomier picture of population adequacy (Figure 5B,C). On the other hand, when mean intakes were above the average requirement, the cut-point method (assuming normally distributed usual intakes) estimated lower levels of intake inadequacy relative to the probability method and painted a rosier picture of population adequacy (Figure 5B,C).

Discussion

Although previous work has assumed similar distribution shapes across nutrients, sexes, ages, and countries, our findings show that there is notable variation in shapes across these subgroups. The degree of variation differs across nutrient types, with nutrients such as long-chain ω-3 fatty acids, vitamin K, retinol, β-carotene, and selenium exhibiting the least similarity across countries. We find evidence of high levels of inadequate intake across a number of nutrients, but more so for females compared with males.

Our analysis unpacks the relation between distribution shapes and the prevalence of inadequacy. When mean intake is below the average requirement, higher variability is associated with a lower prevalence of intake inadequacy, but when mean intake is above the average requirement, higher variability is associated with dramatically higher intake inadequacy. This variability could mean different things in different contexts; where nutrients are scarce, variability may indicate that only a few fortunate individuals achieve adequate diets, but in a context of plenty, large deviations from the mean could result from highly unbalanced and insufficient diets.

When meeting certain criteria, the cut-point method, which has been widely used in the literature due to its lower data requirements compared with the probability method, can calculate the prevalence of inadequate intake relatively accurately (5, 30–32). Yet in our analysis, the cut-point method gave higher or lower estimates of the prevalence of inadequate intakes compared with the probability approach when failing to account for distribution skewness. This trend is supported by other studies (31), but a solution has been limited by the lack of detailed information on nutrient-specific distribution shapes (26). Oftentimes, the true shape of the intake distribution is unknown and a symmetrical distribution is assumed for ease. Moreover, the cut-point method’s assumption that the nutrient intake and requirement distributions are independent of each other is likely violated in many populations, especially for macronutrients (5, 26, 33), which could lead to biased estimates of inadequate intake. Calculations using the cut-point method in conjunction with real (skewed) distributions may result in more precise estimates of nutrient inadequacy.

Our findings related to the high variation in ω-3 fatty acids and selenium are consistent with previous research showing the geographical dependency of these nutrients (34, 35). The concentrations of selenium in soil and water are sensitive to environmental and anthropogenic conditions; thus, populations living in highly localized food systems are more vulnerable to deficiency and toxicity (36). Vitamin K deficiency receives little focus globally because it is thought to be rare (37); the divergent patterns of inadequate intakes observed in our findings should be further explored. A better understanding of the distributions of retinol and carotenoids has direct implications for estimating inadequacy and determining the effects of fortification and supplementation strategies. Vitamin A deficiency is widespread and affects nearly 30% of children under 5 y (38).

Inadequate iron intake was particularly prevalent among females in South and Southeast Asia, sub-Saharan Africa, and Ecuador. In some cases, our estimates of inadequate iron intake were lower than estimates of the prevalence of iron deficiency based on biomarkers, perhaps because iron deficiency biomarkers like serum ferritin may better account for iron bioavailability (39–41). But in populations with a high prevalence of inflammation and infection, estimates of iron deficiency based on anemia prevalence are overestimates, due to the multiple determinants of anemia in these settings (42). Depending on the context, estimates of inadequate iron intake from diet may either underestimate or overestimate deficiency. For example, in Bangladesh, iron deficiency is lower than intakes suggest due to the high amounts of iron in groundwater (43), but in the United States, iron deficiency is higher than what intake estimates show (44).

Inaccuracies in the estimation of intake distributions could have substantial implications for public health; underestimation of adequate intake could result in a failure to identify and intervene on important nutrient needs, whereas overestimation could lead to an inefficient allocation of resources. Accurate methods for determining inadequate intake are critical to inform policies related to agriculture, fortification, nutrition, and dietary guidelines—especially in low- and middle-income countries, where inadequacies are most highly concentrated (45).

There are numerous applications of these distributions for nutritional epidemiology and public health. These include, for example, estimating the impacts of an intervention on nutrient intakes, assessing the proportion of populations and subpopulations at risk of nutrient deficiency or overload, understanding the consequences of distributions for nutrition-related disease risk, targeting nutrition interventions to specific sex or age groups, population-level modeling with microsimulations, forecasting future nutrient gaps, and planning fortification and supplementation policies and programs. Researchers could also use these data to further characterize patterns of covariance, such as by examining patterns of covariance between nutrients to see if the presence of one deficiency might predict the presence of another, and by understanding covariance across age–sex groups and populations. A better understanding of these trends can help to shape recommendations of dietary guidelines and patterns, to determine the need for supplementation and fortification, and to improve intervention targeting.

These findings also influence our understanding of important questions in nutrition-related research. First, these methods could help identify subgroups or nutrients that are most in need of...
context-specific data or intake distributions. Because multiple days of recall data are still rare, future research could help to understand the extent to which other data sources (e.g., food supply data or consumption and expenditure surveys) can provide valid estimates of usual intake distribution shapes. Researchers could apply the methods presented here to assess the implications of assumptions about distribution shapes for the Global Burden of Disease study estimates (46). Last, future analyses that link these distributions to micronutrient biomarkers could improve our understanding of the links between nutrient intake and risk of deficiency with greater precision.

This analysis has several limitations. There are substantial differences in dietary nutrient intakes within countries—for example, due to region, crop composition, or soil factors (47)—that could result in distributions being unrepresentative of some populations. The representativeness of our underlying data sets also varies. Users should refer to Supplemental Table 1 when drawing conclusions about representativeness. Our geographical coverage was subject to data availability. The dietary recall methods used varied across surveys, which could influence estimations of intake (48, 49). Differences in the underlying food composition databases and methodologies could reduce the external validity for comparisons across countries (50). The number of repeated measures available for each country could differentially affect precision, particularly for nutrients that have a large ratio of within- to between-person variability (32, 51). Although multiday dietary recalls measure “usual intakes,” they insufficiently capture natural fluctuations in diets due to holidays, seasons, and so on (45). The inclusion of fortification in this analysis may vary based on the underlying data used for each country. Last, because we focus on food consumption, micronutrient intake from supplements is excluded. Based on these limitations, we ask users of these data and findings to take caution when extrapolating our results to other populations or across countries.

In conclusion, our article highlights how distribution shape matters when assessing a population’s usual nutrient intake. To our knowledge, this is the largest compilation of nutrient intake distributions generated for public use. Our inclusion of the results in the accompanying R package ensures that these methods can be directly applied by researchers. Future work should incorporate greater nuance into assumptions about distribution shapes to more accurately describe population intakes so that research, policy, and programmatic approaches can better meet nutritional needs.

The authors’ responsibilities were as follows—SP and CMF: wrote the paper and conducted the analyses; CDG, AS, SP, and CMF: conceived the idea and analysis plan; AD: designed the SPADE software and advised on and assisted with its use; LHA, SB, JS, and TB: advised on the study design and approach; AS, CL, and DFV: contributed to the analysis; CB, APB-J, LC, AC-G, TC, SPC, AD, KDR, SG, YL, MM, and IM: contributed data and analysis; and all authors: reviewed and approved the final version of the manuscript. The authors report no conflicts of interest.

Data Availability

All code analyzing raw data will be available on a public GitHub page upon publication at https://github.com/cg0lden/subnational_distributions-extended. All raw data sets that we are allowed to share will also be available for download on that page. Code and data pertaining to the modeled distributions are available on GitHub here: https://github.com/cfree14/subnational_nutrient_distributions. Last, the distributions are publicly available through the R package that has been produced as an output of this research, which is free and available for download at https://github.com/cfree14/nutriR/.

References
