



OPEN ACCESS

EDITED BY

Alessandra Durazzo,
Council for Agricultural Research and
Economics, Italy

REVIEWED BY

Tânia Silva-Santos,
Polytechnical Institute of Coimbra, Portugal
Kefita Kashala Kayola,
Arba Minch University, Ethiopia

*CORRESPONDENCE

Tinku Thomas
✉ tinku.sarah@sjri.res.in

RECEIVED 19 May 2025

ACCEPTED 20 October 2025

PUBLISHED 10 November 2025

CITATION

Joseph S, Ghosh S, Swaminathan S and
Thomas T (2025) Estimation of habitual intake
of infrequently consumed nutrients using the
mixture distribution method.
Front. Nutr. 12:1631495.
doi: 10.3389/fnut.2025.1631495

COPYRIGHT

© 2025 Joseph, Ghosh, Swaminathan and
Thomas. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Estimation of habitual intake of infrequently consumed nutrients using the mixture distribution method

Smitha Joseph^{1,2}, Santu Ghosh³, Sumathi Swaminathan⁴ and
Tinku Thomas^{3*}

¹Research Scholar, Manipal Academy of Higher Education (MAHE), Manipal, India, ²Division of Epidemiology and Biostatistics, St. John's Research Institute, St. John's National Academy of Health Sciences (SJNAHS), Bangalore, India, ³Department of Biostatistics, St. John's Medical College, and Hospital, SJNAHS, Bangalore, India, ⁴Division of Nutrition, St. John's Research Institute, SJNAHS, Bangalore, India

Background: The habitual intake of infrequently consumed nutrients typically exhibits a highly skewed distribution, primarily driven by the reported consumption and non-consumption of nutrients in repeated 24-h dietary recalls. The current methods for estimating this distribution are often computationally intense.

Methods: A mixture distribution method (MDM) was proposed to estimate habitual intake distribution of infrequently consumed nutrients, in which the frequency of consumption of a nutrient was modeled using a beta-binomial distribution and the amount consumed using a gamma distribution. The habitual intake using this method was compared to the Iowa State University Foods (ISUF) method using sample data consisting of four non-consecutive 24-h diet recalls collected from 120 children aged 6–59 months in Bihar, India. To assess the impact of zero inflation on the estimation of habitual intake, nutrient intakes were simulated with varying percentages of positive intakes, and habitual intakes were calculated using both methods.

Results: The median (IQR) habitual intakes estimated from the MDM and ISUF methods were 0.47 mg (0.29, 0.65) and 0.46 mg (0.29, 0.62) for vitamin B₆ and 0.38 mcg (0.14, 0.68) and 0.40 mcg (0.18, 0.69) for vitamin B₁₂, respectively. Similarly, comparable results were found for other nutrients such as vitamins B₃, B₅, B₁₂, and A and iodine. The simulated data showed that the estimated habitual intake by the MDM increased with the proportion of positive intakes considering the higher probability of consumption. When the proportion of positive intakes was below 60%, the estimates using the MDM, which considers the probability of consumption, were higher than the arithmetic mean calculated from 15 recalls.

Discussion: The proposed MDM offers a computationally simpler approach to estimate habitual intake distribution by modeling the probability distribution of non-consumption and the distribution of positive intakes. The procedure can be easily implemented using standard statistical software and estimates habitual intake for infrequently consumed nutrients from multiple 24-h dietary recalls.

KEYWORDS

habitual intake, measurement error, 24-h recall, gamma regression, beta-binomial distribution

1 Introduction

Accurate dietary assessment plays a crucial role in public health by providing the evidence base needed to design, implement, and evaluate nutrition-related policies and interventions. Reliable information on dietary intake helps identify nutrient deficiencies, excesses, and dietary patterns associated with chronic diseases such as obesity, diabetes, and cardiovascular disorders. It also supports the monitoring of population-level dietary trends, enabling timely action to address emerging nutritional challenges. In public health practice, well-conducted dietary assessment informs food fortification programs, dietary guidelines, and health promotion strategies tailored to specific population groups. Ultimately, improving the precision of dietary assessment enhances the effectiveness of nutrition policies and contributes to better health outcomes at the population level.

The public health policies and nutritional recommendations are based on the relationship between long-term nutrient intake and health outcomes but not short-term consumption. The usual dietary intake, or habitual intake, provides the average amount of food or nutrient consumed by an individual over a long period (1). The accurate estimation of habitual intake at the population level is crucial for understanding the diet–health relationships and variability of food and nutrient intake, which requires multiple 24-h dietary recalls (24HR). However, the food and nutrients consumed on a fraction of the sample days of recall for a portion of the sample are considered infrequently consumed (2). Therefore, nutrients that are not consumed daily—such as vitamins B₁₂ and E—can be considered as infrequently consumed nutrients (3). Capturing the intake of these nutrients requires a greater number of recalls to differentiate between true consumers and non-consumers (4).

The measurement error model used for habitual intake estimation assumes an approximately symmetric intake distribution (5). While commonly consumed nutrients can be transformed to meet this assumption by simple power or log transformation, infrequently consumed nutrients are typically positively skewed (6, 7). These nutrients present additional challenges, including a high proportion of non-consumption during recall and a skewed intake among consumers. The variability is also influenced by age, sex, ethnicity, and seasonality (7–9).

The Iowa State University Foods (ISUF) method is used to estimate the habitual intake distribution of infrequently consumed foods and nutrients. It uses a two-part model with person-specific effects: the first part models the probability of consuming a certain food or nutrient using a mixture of binomial probabilities and the second part models the intake amount of the food or nutrient when consumed (2). This approach modified the measurement error model to account for the mixture of the consumers' and non-consumers' intake distributions. Similar methods including the National Cancer Institute (NCI) method (7), the Statistical Program to Assess Dietary Exposure (SPADE) method (10), and the Multiple Source Method (MSM) (11) differ in their two-part model implementation of estimating frequency and the amount of consumption.

However, to use these methods, the intake data must be appropriately transformed to align with the measurement error

model to estimate habitual intake. The second part of the ISUF method involves a two-step transformation based on the Iowa State University (ISU) method to handle highly skewed intake distributions. Together, these steps make estimating habitual intake for infrequently consumed nutrients a complex process.

This study proposes a computationally simpler approach built on the mixture model framework of the ISUF method to estimate the habitual intake distributions for infrequently consumed nutrients.

2 Methods

Nutrients consumed on fewer than 90%–95% of the recorded days were classified as infrequently consumed nutrients (4). To assess the intake distribution, a histogram was used to distinguish between regularly and infrequently consumed nutrients. For infrequently consumed nutrients, a substantial portion of the sample reported no intake during recall days.

2.1 Estimation of habitual intake for infrequently consumed nutrients

The habitual intake distribution of infrequently consumed nutrients consists of zero-inflated data from non-consumers and a skewed intake data from consumers on recall days. The ISUF method assumes that the habitual intake of an individual on consumption days is independent of the probability of consumption of the nutrient under study. Thus, habitual intake of the nutrient on all days of recall can be modeled as the individual's habitual amount of intake on consumption days (the conditional distribution of positive intakes) multiplied by the individual's probability of consuming the nutrient on any recall day (2).

Let Y_{ij} be the observed intake for individual i on day j of recall, y_i represent the habitual intake of individual i , and p_i be the probability that an individual i consumes the nutrient on any given day. Let Y_{ij}^* be the observed positive intake and y_i^* be the corresponding habitual intake. Then, the ISUF model is

$$y_i = y_i^* p_i ; p_i \sim D(p; \theta) \quad (1)$$

where $D(\cdot)$ is a suitable probability distribution of p_i . In the ISUF method, the habitual intake of the amount consumed on consumption days was estimated using the ISU method as explained by Nusser et al. (6) which requires a two-step transformation of nutrient intake to normal distribution. The consumption probability distribution was modeled as a discrete set of equally spaced probabilities (ranging from 0.0 to 1.0), with specific probability masses for the number of days of recalls. The proportion of individuals consuming the nutrient on l out of r days (where r is the number of days of recalls and l ranges from 0 to r) was derived from the combination of those binomial probabilities with the weighted probability mass estimated using the modified minimum chi-square estimator (2, 12).

2.2 Mixture distribution method (MDM) of infrequently consumed nutrients

This study suggests two modifications to the ISUF method. First, the conditional distribution of habitual intake on consumption days, $y_i^* = E(Y_{ij}|i, Y_{ij} > 0)$, is modeled using a gamma distribution to account for skewness in observed intake. Second, the distribution of probability of consumption is estimated by modeling the proportion of consumption days from multiple recalls by beta-binomial probability distribution (13) to account for potential overdispersion and varying probability of consumption.

The practical application of gamma distribution to model positive intakes have been evaluated against lognormal and mixture normal distribution by comparing their Akaike information criteria (AIC) values. The data on frequency of consumption were examined for the best-fitting distribution among binomial, Poisson, negative binomial, and beta-binomial distributions. The AIC values are given in [Supplementary Table 2](#). The merit of gamma distribution for modeling skewed nutrient intake is explained elsewhere (14, 15).

As an alternative to transforming individual non-zero or positive nutrient intake data to normal variate, we modeled it using non-normal distribution specifically gamma probability distribution. If Y_{ij} is distributed as gamma with pdf:

$$f_y(y) = \frac{\lambda}{\Gamma(v)} (\lambda y)^{v-1} e^{-\lambda y}, y > 0, v > 0, \lambda > 0, \quad (2)$$

where mean, $E(Y) = \frac{v}{\lambda}$, and $Var(Y) = \frac{v}{\lambda^2} = \frac{mean^2}{v}$. Here, $\Gamma(v)$ was the gamma function, and λ and v were scale and shape parameters, respectively.

Let $Y_{ij}^*: \{i = 1, 2, \dots, n, j = 1, 2, \dots, r_i\}$ denote the set of unadjusted positive observed intakes for a dietary nutrient, where n is the number of individuals with at least one positive intake and r_i is the number of positive intake days for individual i .

The unobserved positive habitual intakes were modeled using gamma distribution with a log link within a measurement error framework as follows:

$$\log(E\{Y_{ij}^*\}) = y_i^* + u_{ij} \quad (3)$$

where y_i^* was the unobserved positive intake of individual i with mean μ_y and variance σ_y^2 , and u_{ij} was the unobserved measurement error with mean 0 and variance σ_u^2 . The variance σ_u^2 represented within-individual variance, and σ_y^2 represented the between-individual variance in intake or the variance of habitual intakes.

The estimates of $\{\mu_y, \sigma_y, \sigma_u\}$ were obtained by the gamma random effect model, and the habitual positive intake was obtained as follows:

$$\hat{z}_i = \log(\hat{y}_i) = \hat{\alpha} + \frac{\hat{\sigma}_y}{\sqrt{\hat{\sigma}_y^2 + \frac{\hat{\sigma}_u^2}{r}}}(z_i - \hat{\alpha}) \quad (4)$$

where $z_i = \log(y_i)$, $\log(\hat{\mu}_y) = \hat{\alpha}$, intercept of gamma random effects model, and $\hat{\sigma}_y$ the estimate of between-individual variability and $\hat{\sigma}_u$ the estimate of within-individual variability. Finally, $\hat{y}_i$ could be estimated by $\exp(\hat{z}_i)$.

The probability of the positive intake (p_i) was estimated by the beta-binomial probability distribution fitted to the frequency of positive intake for r repeated 24-h recalls. The maximum likelihood estimation technique was used to estimate the parameters of the distribution.

Thus, habitual intake was obtained by the [Equation 1](#) as follows:

$$y_i^* = \hat{y}_i \times \hat{p}_i \forall i = 1 \dots n \quad (5)$$

Both regression methods—gamma regression and beta-binomial regression—can easily be implemented in standard statistical software. R package “lme4” was used for the calculation of within- and between-individual variability for the estimation of habitual positive intakes using the gamma regression method. A package named “VGAM” was used for the estimation of probability of consumption using beta-binomial distribution. The R-program code for executing the MDM method is provided in [Supplementary material 1](#).

2.3 Data used for application of the methods

Two surveys were conducted in a cohort of households in Gaya and Nalanda districts of Bihar state during two seasons—the first season between July and August 2019 and the second season between December 2019 and January 2020, to examine the production, distribution, and consumption of nutrient-rich foods. The primary variable of interest was the anthropometric growth of children aged 6 to < 60 months. Two 24-h dietary recalls, including breastmilk intake, were performed for children aged 6 to < 60 months in the sampled households during the two seasons. The details of the sampling procedure, sample size, and other nitty-gritty of this study are described elsewhere (16). A sample of 120 participants with intake data on all four recalls available ([Supplementary Figure 1](#)) was considered for the current analysis. The trained interviewers conducted face-to-face interviews with mothers to collect the 24-h dietary recalls from their children. The second recall was captured on a non-consecutive day. First, the participant listed the foods and beverages consumed during the previous day, including vitamin and mineral supplements from when the child woke up, and for the next 24 h, using food portion size aids (utensils commonly used by the community to eat food). Following this, interviewers assisting in their recall asked queries about breastfeeding habits and foods that they may have forgotten to report, such as snacks, foods consumed during special occasions, and the timing of food consumption. Nutrient data were analyzed using an MS Excel calculator created using the food composition database developed specifically for this purpose (17). The intake of infrequently consumed nutrients (vitamin B₆, vitamin B₁₂, vitamin A retinol activity equivalent (RAE), vitamin B₃, vitamin B₅, and iodine) for each of the four recalls was considered for this study. The prevalence of inadequate intake of these nutrients was calculated using the probability approach (18) based on the dietary recommendations for Indian children (19).

2.4 Impact of zero inflation on estimation of habitual intake of infrequently consumed nutrient

A simulation study was carried out to assess the impact of varying proportions of zero inflation in observed intake on an estimated habitual intake of infrequently consumed nutrient. We assumed a scenario of $n = 15$ repeated recalls and generated a random sample (Z_{ij}) of size 2,000 from a multivariate normal distribution with $\mu = (0.77, 0.74, 0.71, 0.61, 0.68, 0.78, 0.64, 0.67, 0.52, 0.61, 0.54, 0.61, 0.69, 0.78, 0.72)$ and $\text{Diag}(\Sigma) = (1.70, 1.35, 1.41, 1.13, 0.91, 1.31, 1.29, 1.02, 1.50, 1.43, 1.34, 1.21, 1.13, 1.54, 1.27)$; $\sigma_{ij} = \rho\sigma_i\sigma_j$. $\rho = 0.6$ was the within-individual correlation, Σ was the variance-covariance matrix, and $\text{Diag}(\Sigma)$ was the diagonal element in the matrix, which corresponded to the variance. The variance-covariance matrix $\text{Diag}(\Sigma)$ and the correlation coefficient ρ were obtained from the sample intake data of vitamin B₁₂ for children as mentioned above, and the means were a range of values within the Recommended Dietary Allowance (RDA) for vitamin B₁₂ in children and adolescents, which provided a reasonable range as the mean intake in the sample data was very low (less than the estimated average requirement (EAR) for the age group). Then, the actual intake was defined as $Y_{ij} = \exp(Z_{ij})$, which was a skewed distribution. Another random number from the binomial probability distribution with $n = 15$, $p = \{0.2, 0.3, \dots, 0.8\}$ was generated for the positive intakes, where p was the proportion of positive intakes. To simulate zero inflation, a proportion $(1-p)$ of the values of the series of 2,000 nutrient intakes Y_{ij} was replaced by 0. For each value of p , 2,000 samples of intakes were similarly generated. Then, both the ISUF model and MDM were fitted for each simulated dataset. The geometric mean and 95% confidence interval of estimated habitual intakes were computed (Supplementary Table 1). For comparison, habitual intake was also calculated as the arithmetic mean of 15 recalls, with results summarized as the geometric mean with a 95% confidence interval. Smoothed distribution curves of the habitual intakes estimated using the MDM, the ISUF method, and individual mean were plotted to visually assess the impact of varying proportions of zero intake.

3 Results

3.1 Estimation of habitual intake distribution of infrequently consumed nutrients

The data used for analysis consisted of 120 individuals with 4 recalls each. Vitamin B₆, vitamin B₁₂, vitamin B₃, vitamin B₅, vitamin A RAE, and iodine were considered for the application of this method, as they were consumed infrequently, with 17% non-consumers for vitamin B₁₂, 9% non-consumers for vitamin B₆ and B₅, 2% non-consumers for vitamin A, 1% of non-consumers for vitamin B₃, and 1% of non-consumers for iodine. Vitamins A and B₅ and iodine were selected for the demonstration of the method, as they were positively skewed even though they were not inflated by

zero intakes. The description of the observed nutrient intake from the example data is given in Table 1.

For the estimation of habitual intake of vitamins B₆ and B₁₂ from this sample, a two-step transformation was applied as suggested by the ISU method, where the first step was a power transformation, and then, a piecewise cubic estimate was performed as the second step. The transformed data was examined for the normality assumption. However, the density plot (Figure 1) of vitamins B₆ and B₁₂ showed that the transformed data were skewed and not meeting the requirement for the measurement error model used in the ISU method. Therefore, there is a need for a specialized method for estimating habitual intake distribution of infrequently consumed nutrients.

The ISUF method and the proposed MDM, which have been developed for infrequently consumed nutrients, were applied to the dietary intake of the nutrients under consideration. In the ISUF method, the habitual intake distribution of positive intakes was estimated using the ISU method. In the ISU method, the positive intakes of vitamins B₆, B₃, and A RAE were transformed to normality using power transformation. Powers of 0.3, 0.5, and 0.4 were sufficient to transform vitamins B₆, B₃, and vitamin A RAE to normal distribution, respectively. However, a two-step transformation was required for the positive intake of vitamin B₁₂, vitamin B₅, and iodine, which was highly skewed. The power used in the first step was 0.2 for vitamins B₁₂ and B₅ and 0.3 for iodine, and a piecewise cubic transformation, as explained in Nusser et al.'s (6) study, was used in the second step for these nutrients. The habitual intake of positive intakes in the transformed scale was estimated using the shrinkage estimator of measurement error model (5, 20).

Since the positive intake data of vitamins B₆, B₃, and A RAE were normally distributed after the power transformation, an inverse power transformation was used for converting positive habitual intake in the transformed scale to the original scale. For the more skewed distribution of positive intake of vitamin B₁₂, vitamin B₅, and iodine, which require two-step transformation, the relation between nutrient intake in the transformed scale and the original scale was developed using a polynomial curve fitting. The following relation was used for the back-transforming habitual intake of vitamin B₁₂ in the transformed scale to the original scale.

$$\begin{aligned} \text{Habitual vitamin B12 intake} &= 0.59 - 3.64 \\ &+ \text{Habitual vitamin B12 intake in normal scale} + 8.52 \\ &+ \text{Habitual vitamin B12 intake in normal scale}^2 - 9.47 \\ &+ \text{Habitual vitamin B12 intake in normal scale}^3 + 5.0 \\ &+ \text{Habitual vitamin B12 intake in normal scale}^4 \end{aligned}$$

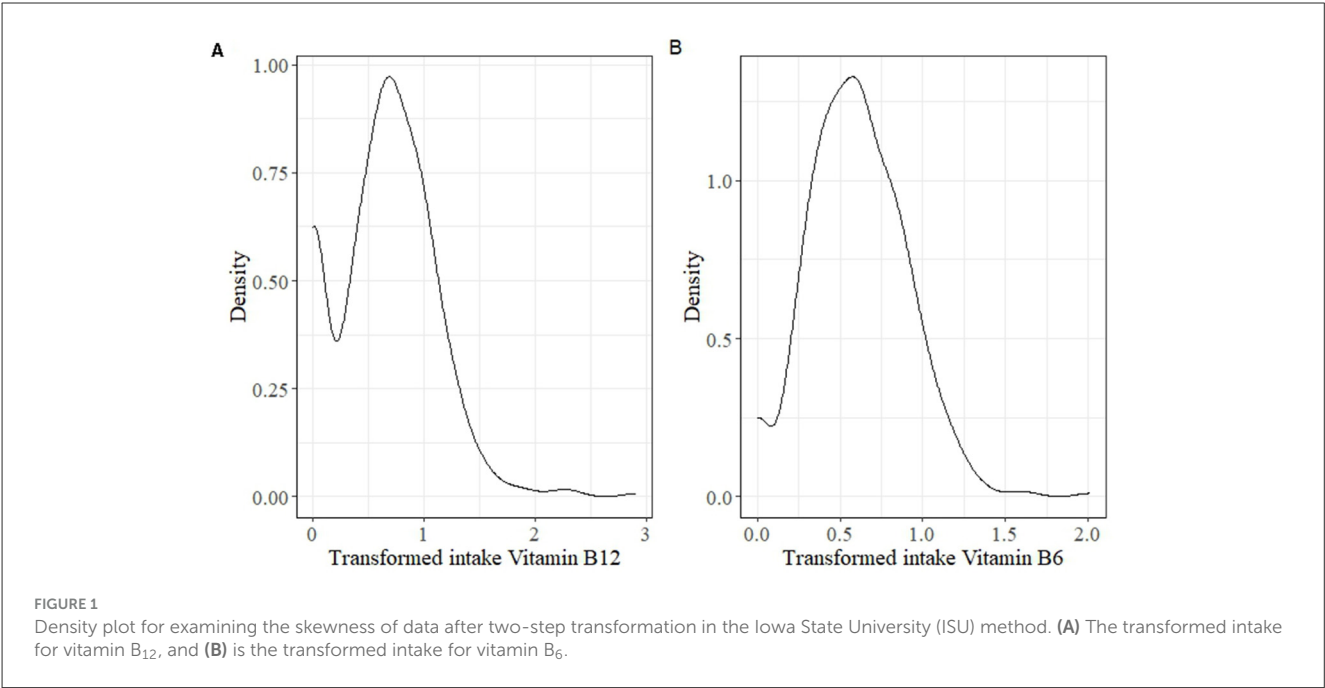
Thus, the estimation of habitual positive intake of nutrients was computationally intense in the ISUF method.

The goodness of fit of lognormal, gamma, and mixture normal distributions for the positive intakes of the nutrients was tested using the AIC (Supplementary Table 2). The AIC was lowest for gamma distribution for all nutrients except for vitamin B₆ and closer to that of the lowest AIC of mixture normal distribution for vitamin B₆. Thus, the gamma regression method was considered suitable for estimating habitual intake of positive intakes for all the nutrients under consideration.

TABLE 1 Summary of observed nutrient intake in each recall (*n* = 120).

Nutrient		First season		Second season	
		First recall	Second recall	First recall	Second recall
Vitamin B ₆ mg	Mean ± SD	0.47 ± 0.37	0.49 ± 0.37	0.56 ± 0.33	0.54 ± 0.32
	Median (Q ₁ , Q ₃)	0.39 (0.19, 0.65)	0.43 (0.22, 0.70)	0.50 (0.30, 0.79)	0.49 (0.30, 0.73)
Vitamin B ₁₂ µg	Mean ± SD	0.57 ± 0.64	0.57 ± 0.68	0.70 ± 0.79	0.67 ± 0.93
	Median (Q ₁ , Q ₃)	0.43 (0.08, 0.86)	0.43 (0.05, 0.87)	0.54 (0.23, 0.91)	0.44 (0.17, 0.91)
Vitamin A mcg RAE	Mean ± SD	106 (77)	94 (68)	151 (89)	148 (93)
	Median (Q ₁ , Q ₃)	92 (55, 138)	82 (50, 126)	134 (88, 196)	130 (78, 201)
Vitamin B ₃ mcg	Mean ± SD	4.9 (2.9)	5.0 (3.5)	4.6 (3.4)	4.5 (2.8)
	Median (Q ₁ , Q ₃)	4.5 (2.7, 6.7)	4.2 (2.5, 6.7)	3.8 (2.2, 6.3)	4.2 (2.7, 5.9)
Vitamin B ₅ mg	Mean ± SD	0.99 (1.19)	0.95 (1.11)	0.81 (0.97)	0.94 (1.06)
	Median (Q ₁ , Q ₃)	0.39 (0.09, 1.69)	0.40 (0.08, 1.43)	0.40 (0.07, 1.38)	0.52 (0.12, 1.49)
Iodine mcg	Mean ± SD	32 (27)	29 (20)	30 (25)	35 (30)
	Median (Q ₁ , Q ₃)	25 (14, 43)	26 (13, 40)	26 (13, 43)	24 (16, 44)

SD, standard deviation; Q1, quartile 1; Q3, quartile 3; RAE, retinol activity equivalent.



In the MDM, the habitual intake distribution of positive intakes was directly estimated using the gamma regression method. Habitual intake was then estimated using a shrinkage estimator in the measurement error framework as given in Equation 4 (5, 20).

In the ISUF method, the probability of positive consumption was modeled as a discrete set of equally spaced probabilities (ranging from 0.0 to 1.0), and specific probability masses for 1, 2, 3, and 4 days of positive recalls were independently estimated.

The overdispersion parameter phi for the binomial distribution of the frequency of consumption was 2.2 for vitamins B₁₂ and B₆, 1.8 for vitamin A, 0.99 for vitamin B₃ and iodine, and 1.2 for vitamin B₅. Overdispersion was present in vitamins B₁₂, B₆, and A. The goodness of fit of binomial, Poisson, negative binomial, and beta-binomial distributions for the frequency of consumption of

these nutrients was tested using the AIC (Supplementary Table 2). The AIC was lowest for the beta-binomial distribution for the nutrients except for vitamin B₃ and iodine and for these two nutrients, the AIC was equal for the binomial and beta-binomial distributions. Hence, the beta-binomial regression method was considered suitable for estimating the probability of consumption for all these nutrients under consideration.

Thus, in the MDM, the frequency of consumption of the nutrients was modeled using beta-binomial distribution. Each individual's probability of consumption was estimated from the observed frequency of intake using the parameters of the beta-binomial distribution, as presented in Supplementary Table 3.

The distribution of habitual intake for each individual was then estimated as the product of estimated habitual intake on

TABLE 2 Estimated habitual intake using different methods of estimation.

Nutrient	Estimate using the ISUF method		Estimate using the MDM	
	Mean ($\pm$ SD)	Median (Q ₁ , Q ₃)	Mean ($\pm$ SD)	Median (Q ₁ , Q ₃)
Vitamin B ₆ mg	0.48 ($\pm$ 0.26)	0.47 (0.29, 0.65)	0.46 ($\pm$ 0.26)	0.46 (0.29, 0.62)
Vitamin B ₁₂ μ g	0.50 ($\pm$ 0.42)	0.40 (0.18, 0.69)	0.47 ($\pm$ 0.41)	0.38 (0.14, 0.68)
Vitamin A RAE μ g	114 ($\pm$ 51)	113 (80, 141)	107 ($\pm$ 50)	107 (76, 134)
Vitamin B ₃ mg	4.21 ($\pm$ 1.08)	4.20 (3.69, 4.88)	4.02 ($\pm$ 1.26)	4.16 (3.48, 4.78)
Vitamin B ₅ mg	0.74 ($\pm$ 0.79)	0.26 (0.10, 1.23)	0.69 ($\pm$ 0.76)	0.24 (0.08, 1.21)
Iodine mg	25.56 ($\pm$ 8.36)	26.07 (20.76, 30.87)	25.0 ($\pm$ 9.96)	24.8 (19.4, 31.12)

ISUF, Iowa State University Foods; MDM, mixture distribution method; SD, standard deviation; Q1, quartile 1; Q3, quartile 3; RAE, retinol activity equivalent.

TABLE 3 Prevalence of inadequate intake of the habitual intakes obtained using various methods.

Nutrient	Prevalence of inadequacy (%)	
	ISUF method	MDM
Vitamin B ₆ mg	93.6 (89.8, 97.3)	95.0 (91.7, 98.2)
Vitamin B ₁₂ mcg	89.1 (83.8, 94.3)	89.6 (84.6, 94.7)
Vitamin A RAE μ g	55.7 (54.8, 56.7)	56.9 (55.5, 58.4)
Vitamin B ₃ mg	96.7 (94.5, 98.9)	96.7 (94.5, 98.9)
Iodine mg	99.5 (99.4, 99.7)	99.4 (99.1, 99.7)

ISUF, Iowa State University Foods; MDM, mixture distribution method, RAE, retinol activity equivalent.

consumption days and the individual’s probability of consumption. The descriptive statistics for estimated habitual intake are given in [Table 2](#).

As shown in [Table 2](#), it can be observed that the estimated habitual intakes using the MDM for vitamins B₆ and B₁₂ were comparable to the estimates using the complex ISUF method. The median and quartiles for vitamin B₆ were 0.47 (0.29, 0.65) using the ISUF method and 0.46 (0.29, 0.62) using the MDM, although the method employed was much simpler and direct with no transformation of data. Similarly, comparable results were found for other nutrients as well. The habitual intake estimates stratified by age are given in [Supplementary Table 4](#).

The prevalence of inadequacy was comparable for the habitual intakes estimated using the ISUF method and MDM ([Table 3](#)). The prevalence of inadequacy for habitual intake using the ISUF method and MDM was 93.6% (89.8%, 97.3%) and 95% (91.7%, 98.2%) for vitamin B₆ and 89.1% (83.8%, 94.3%) and 89.6% (84.6%, 94.7%) for vitamin B₁₂, respectively. The analysis showed comparable prevalence for inadequate intakes for other nutrients also. The prevalence of inadequate intake stratified by age is given in [Supplementary Table 5](#).

3.2 Validation of the MDM using simulation

Using the simulated data, the geometric mean of the estimated habitual intake distribution was obtained and plotted for visual comparison, as given in [Figure 2](#).

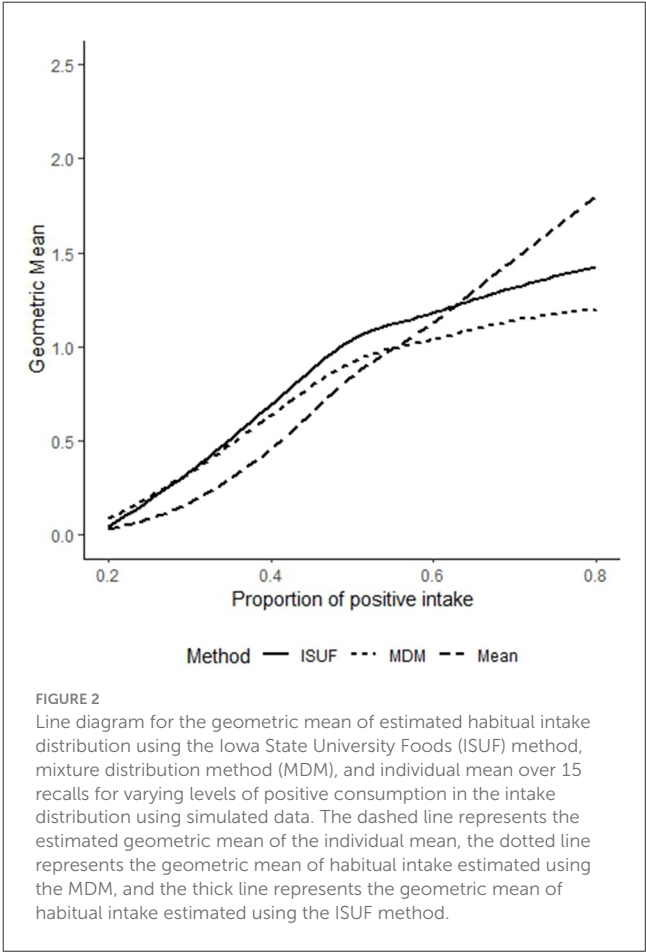


FIGURE 2 Line diagram for the geometric mean of estimated habitual intake distribution using the Iowa State University Foods (ISUF) method, mixture distribution method (MDM), and individual mean over 15 recalls for varying levels of positive consumption in the intake distribution using simulated data. The dashed line represents the estimated geometric mean of the individual mean, the dotted line represents the geometric mean of habitual intake estimated using the MDM, and the thick line represents the geometric mean of habitual intake estimated using the ISUF method.

Figure 2 illustrates that, when the proportion of positive intakes was below 60%, the habitual intakes estimated using the MDM and ISUF method were comparable and tended to exceed the individual mean. This was attributable to both methods incorporating the probability of zero intake, which helps mitigate the downward bias introduced by a high frequency of zero observations in the data. Thus, approaches that model the probability of non-consumption could yield more accurate estimates of habitual intake under zero-inflated conditions. Conversely, as the proportion of positive intakes exceeded 70%, the habitual intakes estimated by the MDM and ISUF method fell below the individual mean. This reflected the

need for using models with the probability of positive consumption, such as the MDM and ISUF method, for estimating habitual intake of infrequently consumed nutrients.

4 Discussion

Accurate estimation of habitual intake is essential for developing evidence-based nutrition policies such as setting dietary reference values and designing fortification and supplementation programs (21). Estimation of habitual intake of nutrients is challenging when the intake data are skewed and become more complex for infrequently consumed nutrients due to zero inflation from non-consumption on the recall day. The proposed method, referred to as the MDM, addresses this using a two-part model that separately estimates the probability of consumption and the amount consumed on intake days. The probability of consumption was estimated by the modeling frequency of consumption as a beta-binomial distribution while intake amounts on consumption days were modeled using the gamma regression method. The simulation study by changing the proportion of positive intakes showed that the estimates using the MDM were closer to the individual means for lower proportions of positive intakes (<60%) compared to the estimates using the ISUF method.

Studies have shown that the shape of habitual intake distributions varies by nutrient, country, gender, and age group, with vitamin intakes often displaying greater variability (22). These differences influence the estimated prevalence of inadequate intake. Understanding intake distributions and applying appropriate methods to estimate habitual intake directly impact the assessment of inadequacy and the evaluation of nutrition interventions. Studies have modeled skewed intake data using the gamma distribution to estimate inadequacy (23) or summarized intake without normal transformation or measurement error correction (15). Nutrient intake was modeled using gamma distribution in studies where the adjustment on the variability was performed using the variance ratio from external sources and applied on the parameters of the observed intake distribution (24). Then the gamma distribution, adjusted for the external variance ratio was used to represent the habitual intake distribution.

Modeling the consumption frequency using beta-binomial distribution was suggested in previous studies (13, 25). However, the estimation of habitual intake based on the amount of consumption was performed by transforming the data to normal distribution. The Statistical Program to Assess Dietary Exposure (SPADE) also uses beta-binomial distribution for estimating the probability of consumption. However, the SPADE method requires the positive intake to follow normal distribution. The amount of consumption was modeled using a linear mixed effect model after transforming to normal distribution using Box-Cox transformation and back-transforming using Gaussian quadrature (10). The estimation using MSM (11) handled zero-inflated dietary data by distinguishing between consumers and non-consumers, thereby integrating data from 24-h recall and food frequency questionnaire (FFQ). In this method, the probability of consumption was estimated using a logistic regression model for consumers and not by recall, and the habitual intake was estimated after transforming data with the help of Box-Cox transformation.

This method cannot be preferred while dealing with highly skewed data. One of the limitations of this method lies in the transformation to normal distribution and back-transformation process, which may introduce bias if not handled appropriately. The advantage of the MDM method proposed in this study is that it models intake data based on the actual distribution of positive intake using the gamma regression method.

The method developed by the National Cancer Institute, known as the NCI method, also uses the two-part model where the probability of consumption was predicted using the logistic model, and the amount consumed was transformed to normality using the Box-Cox transformation (7). It incorporated individual and recall-specific covariates into the model. This procedure of estimation of the parameters is complicated. The use of the MDM reduces the complexity in estimating habitual intake.

An ensemble approach had been proposed for estimating habitual intake from a single 24-h recall where the variance ratio could be obtained from an external source (8). However, the usage of the external variance ratio of between- and within-variability does not address the problem of skewed and infrequent intake of nutrients.

The limitation of the existing methods of estimating infrequently consumed nutrients was that the transformation to normal and back transformation was complex due to high skewness in the consumption data of infrequently consumed nutrients. Such transformations, if not handled appropriately, can introduce bias to the estimates. The advantage of the MDM method proposed in this study addresses these challenges by directly modeling the actual distribution of the positive intake using the gamma regression method and the probability of intake using the beta-binomial regression method.

Studies discussing the minimum sample size required for the estimation of habitual intake state that at least 50 individuals with at least 2 recalls are sufficient to estimate habitual intake using a measurement error model (4, 26). This holds good for nutrients consumed regularly, as the multiple recalls will be capturing the consumption of that nutrient. However, for infrequently consumed nutrients, the estimates might be biased if the probability of consumption of the nutrient on any given day is low. Differentiating between occasional consumers and true non-consumers is not possible with a limited number of recalls. Therefore, either the sample size or the number of repeated recalls needs to be higher to ensure a sufficient number of individuals with at least two positive intakes are presented in the data. Sample size can affect the estimated habitual intake, as the bias reduces with increased sample size (27). The dietary data of 120 children with 4 recalls each, used to demonstrate the proposed MDM, can be considered sufficient for the estimation of habitual intake, as the estimates were comparable with the ISUF method (4, 13).

The application of the MDM method to larger and diverse intake data is needed to examine the effects of sample size and the minimum number of recalls required for estimating habitual intake of infrequently consumed nutrients.

The MDM needs to be explored further to accommodate factors associated with positive consumption. Additionally, the incorporation of weights for varying numbers of recalls across participants in a study also needs to be investigated to improve

generalizability. While the current study focuses on nutrient intake, the MDM could be extended to estimate habitual intake of infrequently consumed foods. However, such an extension must carefully distinguish between true non-consumers and infrequent consumers to ensure accurate estimation.

5 Conclusion

There are several challenges in estimating the habitual intake of infrequently consumed nutrients when the number of repeated 24-h recalls available is low. This study proposes a computationally simpler MDM that modeled the frequency of consumption of the nutrient and the amount consumed. The suggested MDM method is straight forward and helps to estimate habitual intake for infrequently consumed nutrients precisely and accurately for zero-inflated infrequently consumed nutrient data.

Data availability statement

Data described in the manuscript will be made available upon request. Requests to access these datasets should be directed to Dr. Tinku Thomas, tinku.sarah@sjri.res.in.

Ethics statement

The studies involving humans were approved by Institutional Ethics Committee St. John's Medical College & Hospital, Bangalore, India. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

SJ: Conceptualization, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing. SG: Conceptualization, Methodology, Supervision, Writing – review & editing. SS: Data curation, Writing – review & editing. TT: Conceptualization, Data curation, Methodology, Supervision, Writing – review & editing.

References

1. Toozé JA, Kipnis V, Buckman DW, Carroll RJ, Freedman LS, Guenther PM, et al. A mixed-effects model approach for estimating the distribution of usual intake of nutrients: the NCI method. *Stat Med*. (2010) 29:2857–68. doi: 10.1002/sim.4063
2. Nusser SM, Fuller WA GP. Estimation of usual dietary intake distributions: adjusting for measurement error and nonnormality in 24-hour food intake data. In: Trewin D, editor. *Survey Measurement and Process Quality*. New York, NY: Wiley (1996). 689–709 p.
3. Rossato SL, Fuchs SC. Diet data collected using 48-h dietary recall: within—and between-person variation. *Front Nutr*. (2021) 8:667031. doi: 10.3389/fnut.2021.667031
4. Toozé JA. *Estimating Usual Intakes from Dietary Surveys : Methodologic Challenges, Analysis Approaches, and Recommendations for Low- and Middle-Income Countries*. Washington, DC: Intake – Center for Dietary Assessment/FHI Solutions (2020).
5. Carriquiry AL. Estimation of usual intake distributions of nutrients and foods. *J Nutr*. (2003) 133:601S–8S. doi: 10.1093/jn/133.2.601S
6. Nusser SM, Carriquiry AL, Dodd KW, Fuller WA. A semiparametric transformation approach to estimating usual daily intake distributions. *J Am Stat Assoc*. (1996) 91:1440–9. doi: 10.1080/01621459.1996.10476712

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Data used for this study was taken from a primary study funded by the Bill & Melinda Gates Foundation, Seattle, WA [Grant Number: OPP1194964].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. During the preparation of this work, the author(s) used online version of ChatGPT to improve language. After utilizing this service, the author(s) thoroughly reviewed and edited the content as necessary and take full responsibility for the content of the published article.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnut.2025.1631495/full#supplementary-material>

7. Toozee JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, et al. A New statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *J Am Diet Assoc.* (2006) 106:1575–87. doi: 10.1016/j.jada.2006.07.003
8. Chi SA, Lee H, Lee JE, Lee HS, Kim K, Yeo IK. An ensemble method based on marginal-effect models (EMM) for estimating usual food intake from single-day dietary data and internal/external two-day dietary data. *Eur J Clin Nutr.* (2022) 77:325–34. doi: 10.1038/s41430-022-01231-1
9. Rossato SL, Olinto MTA, Henn RL, Moreira LB, Camey SA, Anjos LA, et al. Seasonal variation in food intake and the interaction effects of sex and age among adults in southern Brazil. *Eur J Clin Nutr.* (2015) 69:1015–22. doi: 10.1038/ejcn.2015.22
10. Dekkers ALM, Verkaik-Kloosterman J, van Rossum CTM, Ocké MC. SPADE, a new statistical program to estimate habitual dietary intake from multiple food sources and dietary supplements. *J Nutr.* (2014) 144:2083–91. doi: 10.3945/jn.114.191288
11. Haubrock J, Nöthlings U, Volatier JL, Dekkers A, Ocké M, Harttig U, et al. Estimating usual food intake distributions by using the multiple source method in the EPIC-Potsdam calibration study. *J Nutr.* (2011) 141:914–20. doi: 10.3945/jn.109.120394
12. Dodd K. *Technical Guide to C-SIDE (Software for Intake Distribution Estimation). Dietary Assessment Research Series Report 9, A.* Technical report 96-TR 32. Center for Agricultural and Rural Development; Iowa State University (1996).
13. de Boer WJ, van der Voet H, Bokkers BGH, Bakker MI, Boon PE. Comparison of two models for the estimation of usual intake addressing zero consumption and non-normality. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess.* (2009) 26:1433–49. doi: 10.1080/02652030903161606
14. Joseph S, Swaminathan S, Ghosh S, Thomas T. A new approach to estimate habitual intake of nutrients with skewed distribution. *J Nutr.* (2025) 155:3066–74. doi: 10.1016/j.tjn.2025.05.050
15. Corrente JE, Fumes G, Fontanelli MM, Fisberg RM, Marchioni DLM. Use of asymmetric models to estimate the distribution of usual nutrient intakes. *J Nutr Heal.* (2016) 2:1–6. doi: 10.13188/2469-4185.1000020
16. Makkar S, Manivannan JR, Swaminathan S, Travasso SM, John AT, Webb P, et al. Role of cash transfers in mitigating food insecurity in India during the COVID-19 pandemic: a longitudinal study in the Bihar state. *BMJ Open.* (2022) 12:1–9. doi: 10.1136/bmjopen-2021-060624
17. Bharathi AV, Kurpad AV, Thomas T, Yusuf S, Saraswathi G, Vaz M. Development of food frequency questionnaires and a nutrient database for the Prospective Urban and Rural Epidemiological (PURE) pilot study in South India: methodological issues. *Asia Pac J Clin Nutr.* (2008) 17:178–85. doi: 10.6133/APJCN.2008.17.1.25
18. Carriquiry AL. Assessing the prevalence of nutrient inadequacy. *Public Health Nutr.* (1999) 2:23–33. doi: 10.1017/S1368980099000038
19. ICMR-NIN. *Revised Short Summary Report-2023, ICMR-NIN Expert Group on Nutrient Requirement for Indians, Recommended Dietary Allowances (RDA) and Estimated Average Requirements (EAR)-2020. [Report].* ICMR-NIN Expert Group on Nutrient Requirement for Indians (2023).
20. Evaluation NRC. (US) *Report on C for D. Nutrient Adequacy: Assessment Using Food Consumption Surveys.* Washington, DC: National Academies Press (US) (1986).
21. Institute of Medicine. *Dietary Reference Intakes: Applications in Dietary Assessment.* Washington, DC: The National Academies Press (2000).
22. Passarelli S, Free CM, Allen LH, Batis C, Beal T, Biltoft-Jensen AP, et al. Estimating national and subnational nutrient intake distributions of global diets. *Am J Clin Nutr.* (2022) 116:551–60. doi: https://doi.org/10.1093/ajcn/nqac10810.1093/ajcn/nqac108
23. Yokoi K. Simplified population data analysis using gamma distribution for nutritional requirements and its application to the estimation of iron requirements for women of child-bearing age. *J Trace Elem Med Biol.* (2020) 62:126597. doi: 10.1016/j.jtemb.2020.126597
24. Chang HY, Suchindran CM, Pan WH. Using the overdispersed exponential family to estimate the distribution of usual daily intakes of people aged between 18 and 28 in Taiwan. *Stat Med.* (2001) 20:2337–50. doi: 10.1002/sim.838.abs
25. Slob W. Probabilistic dietary exposure assessment taking into account variability in both amount and frequency of consumption. *Food Chem Toxicol.* (2006) 44:933–51. doi: 10.1016/j.fct.2005.11.001
26. Kirkpatrick SI, Subar AF, Toozee JA. Statistical approaches to mitigate measurement error in dietary intake data collected using 24-hour recalls and food records/diaries. *Adv Assess Diet Intake.* (2017) 30:19–43. doi: 10.1201/9781315152288-2
27. Souverein OW, Dekkers AL, Geelen A, Haubrock J, de Vries JH, Ocké MC, et al. Comparing four methods to estimate usual intake distributions. *Eur J Clin Nutr.* (2011) 65:S92–101. doi: 10.1038/ejcn.2011.93