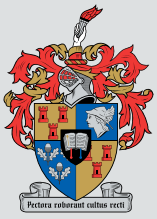


UNIVERSITEIT • iYUNIVESITHI • STELLENBOSCH • UNIVERSITY

100
1918 · 2018

INAUGURAL LECTURE

Prof. JH (Hannelie) Nel
Professor in the Department of Logistics
at Stellenbosch University

Bringing dietary data closer to the truth with statistical adjustment: The 2018 Provincial Dietary Intake Survey as an example

Bringing dietary data closer to the truth with statistical adjustment: The 2018 Provincial Dietary Intake Survey as an example

Inaugural lecture delivered on 01 October 2020

Prof. JH (Hannelie) Nel
Department of Logistics

Prof. JH (Hannelie) Nel
Professor in the Department of Logistics
at Stellenbosch University

Biography of author

Prof. JH (Hannelie) Nel is a professor in the Department of Logistics at Stellenbosch University, with a DSc in Statistics and Operations Research. Her academic career started as a junior lecturer and a lecturer at the Vaal Triangle Campus of the PU for CHE from 1979 to 1984 in Operations Research and lecturer in the Department of Computer Science at the University of the North (now Limpopo) (1984–1995). From 1995 to 2003, she did consultation work (statistics). She also obtained an MBA (*cum laude*) at Stellenbosch University in 2002. She has lectured Operations Research, Quantitative Management and Forecasting for under-graduate and postgraduate students since 2004, and supervised several postgraduate students. She was the chair of the Department of Logistics from 2011 to 2017. She has also been intricately involved as analyst in research studies in human nutrition, anthropometric studies, socio-demographic studies and transport-related studies. She is author and co-author of more than 60 accredited publications.

Bringing dietary data closer to the truth with statistical adjustment: The 2018 Provincial Dietary Intake Survey as an example

ABSTRACT

A single 24-hour recall is often used as the primary instrument for measuring dietary intake in large dietary studies. A common concern with a single 24-hour recall is the day-to-day variation in the diet of free-living populations. The magnitude of the mostly random within-person variance varies by nutrients and is largely dependent on cultural and ecological factors. These errors result in large standard deviations in populations. A result of exaggerated variation is that the percentage of respondents below or above specified cut points will be distorted. A method to distinguish within-person from between-person variation, account for extreme intakes and allow for adjustment for covariates was applied.

Results indicated that the overall unadjusted and adjusted means of the nutrient intakes were mostly similar. However, large differences between unadjusted and the adjusted values of the percentiles of the intake of some nutrients were observed. The method compressed the distribution of nutrient intake towards the mean. This compression was most apparent when examining the percentage of the population below or above AMDR/DRI limits, especially for nutrients with skew distributions. In these cases, the adjusted percentage below or above the cut points were substantially smaller, demonstrating that the prevalence of dietary insufficiency or excess could be overestimated if intakes are not adjusted. Importantly, the method allows for estimation of population exposure without conducting repeated 24-hour recalls on the total sample. This study shows that a small subsample (11%) may be sufficient to conduct the described adjustments.

1. INTRODUCTION AND BACKGROUND

It is recognised that diet is among the most important influences on health in modern societies (Katz & Meller, 2014). Unwise eating patterns can lead to premature death and chronic disease. On the other hand, healthy eating is associated with increased life expectancy and a reduction in the risk of chronic disease (Afshin *et al*., 2019). It is therefore important to monitor usual food consumption and to measure diet, especially at population level, in order to identify groups at risk of nutrition deficiencies or excess and to formulate food and nutrition policies for disease reduction and health promotion (Lee & Nieman, 2013).

Usual dietary intake is the long-term average daily intake of a nutrient or food. The concept of usual intake is important, because dietary recommendations involving populations are intended to be met over time and diet–health hypotheses are based on dietary intakes over the long term (NCI, 2020). The usual intake of selected nutrients is critical when relationships between diet and biological parameters or chronic diseases are assessed. The closer dietary intake estimations are to the truth (actual usual intake), the more robust disease–diet association research is. The estimates of usual intakes can also be used to assess inadequate intake of populations or groups within a population.

Diet variation is a result of the variation within and between persons. Within-person variation occurs due to daily fluctuations, weekends and seasonality. Between-person variation increases according to the heterogeneity of the population and their usual food patterns (Costa *et al*., 2008). The number of days of diet on a record is required to obtain the most correct intake evaluation, and this calculation is dependent on the ratio between the within-person variation and the between-person variation for each nutrient. The lower the ratio, the fewer repeated measurements are needed (Costa *et al*., 2008).

In the seventies and eighties, the golden standard for usual dietary intake assessment was a seven-day weighed food record (WFR) (International Dietary Data Expansion Project, 2020). The WFR requires the respondent or enumerator to weigh all beverages and foods at the time of consumption and record the weight and a detailed description of the food, including the preparation method and brand names. Any waste on the plate must also be weighed and recorded. This method was considered the most precise method of quantifying food intake and is consequently also used as a method in validation studies.

The seven-day WFR offers a high degree of accuracy in assessing food and nutrient intake, but the method can be expensive, intrusive and time-consuming and might distort behaviour of the respondents, especially if an enumerator is present throughout the day. At some point it was realised that respondent fatigue resulted in changes in the diet (simplifying) or leaving items out to reduce the burden, resulting in underestimation of intake. These factors and the resulting cost of this method preclude it from use in large samples representative of populations (Gibson, 2005).

Because of these constraints, methods such as 24-recalls and food frequencies are now typically used to investigate the dietary intake of large groups. Of these, a single 24-hour recall has been the most frequently used because of its ease of administration and the short time over which the client has to recall food intake. This quantitative daily consumption method consists of records designed to measure the quantity of all foods consumed over a 24-hour period.

The single 24-hour recall is an imprecise measure of an individual's usual intake due to the day-to-day variation in the diet of an individual. According to Gibson (2005), a single 24-hour recall, or a single WFR, with a large number of subjects and adequate representation of all days of the week is sufficient to estimate the mean nutrient intake of a group. If the purpose of the study is to determine the proportion of the population 'at risk', to calculate usual intakes of foods or to calculate rank statistics (such as median values), multiple replicate observations on each individual are further required. If the number of measurement days for the 24-hour recall is increased, the estimates of usual intakes of individuals can be obtained.

The repeated application of 24-hour recalls may not always be possible in nutritional studies of large populations due to time, staff, equipment or financial constraints (Knuppel *et al*., 2019).

Basiotis *et al*. (1987) investigated the number of days of food intake data needed to estimate usual intake for selected nutrients. They recorded the dietary intake of 13 men and 16 women for a year. It was then used to estimate the number of days of food intake records needed to estimate the usual intake of an individual, and for a group, with a given degree of statistical confidence.

In Table 1, a 'precise' estimate was defined as an X-day average intake being within 10% of the 'true average' intake for the individual or the group, 95% of the time. The true average intake was defined as the 365-day average for individuals or groups.

*Table 1: Ranges and averages of number of days required to estimate true average intake for an individual and for a group, with given statistical confidence**

Component	Men (n = 13) Mean (minimum–maximum)		Women (n = 16) Mean (minimum–maximum)	
	Individual	Group	Individual	Group
Energy	27 (14–84)	3	35 (14–60)	3
Iron	68 (18–130)	7	66 (28–142)	6
Vitamin A	390 (115–1 724)	39	474 (152–1 372)	44
Thiamine	138 (46–405)	13	198 (41–728)	16
Riboflavin	57 (13–135)	7	90 (31–231)	7
Niacin	53 (27–89)	5	78 (48–126)	6
Vitamin C	249 (90–900)	33	222 (83–328)	19

*Estimated with intake data from one-year dietary intake study by the Beltsville Human Nutrition Research Center of the US Department of Agriculture (Source: Basiotis *et al*., 1987)

Basiotis *et al*. (1987) concluded that if a study examined 100 women, a food intake record of one day would be sufficient to estimate true average food energy and iron intakes, and seven days of food intake records would be needed for vitamin A. If the study examined a three-day food intake record from each woman, a sample size of 15 individuals would be needed to estimate food energy intake with the defined level of statistical precision, a sample size of 32 would be needed for iron and one of 231 to estimate vitamin A intake. Therefore, equally precise intakes for groups may be obtained by increasing the number of food intake records per individual or the number of individuals in the group.

Ideally, dietary intake methods should be validated for application in specific target populations. Apart from using the seven-day WFRs for validating 24-hour recalls, biomarkers have also been used for these purposes. Biomarkers are variables measured in body fluids (urine and blood samples, for example) that independently reflect intake of a food (Johnson, 2002). A commonly used biomarker is urinary nitrogen excretion, which is used to estimate protein, potassium and sodium intake (Wark *et al*., 2018). Urinary fructose and sucrose concentrations are quantified and used to predict total sugar intake. Blood samples can be used to validate vitamin C and vitamin E intake, for example. The challenge here is that experts showed that one would need to collect eight days of urine collections to estimate urinary nitrogen output to within 5% to be used to validate dietary nitrogen intake (Bingham, 2003).

To evaluate nutrient adequacy, measures of usual dietary intake to account for day-to-day variation in food consumption are now preferred. To date, few international studies have investigated usual dietary intakes that also account for the effect of within-person variability, and none in South Africa. The Provincial Dietary Intake Study (PDIS) is the first of this kind undertaken in South Africa (Senekal *et al*., 2019). The purpose of this paper is to describe the National Cancer Institute (NCI) method, which implements a non-linear mixed regression model to account for within- and between-person variation (NCI, 2019). The advantage of this method is that it is possible to estimate usual intake distributions even if repeated dietary data are available only for a relatively small subsample of the total population (Piernas *et al*., 2015).

2. THE PROVINCIAL DIETARY INTAKE STUDY

2.1 Introduction

The National Food Consumption Survey (NFCS) was undertaken in one- to nine-year-old children in South Africa in 1999. This study provided the Department of Health with vital information on the nutritional status of children and the foods eaten and purchased by households in South Africa. Based on the results of this study, the Department of Health was able to develop and implement strategies to address malnutrition in children. Dietary findings indicated that most children consumed a diet deficient in energy and of poor nutrient density to meet their nutrient requirements. For South African children as a whole, the dietary intake of the following nutrients were deficient: energy; calcium; iron; zinc; selenium; vitamins A, D, C, E and B6; riboflavin; and niacin folic acid (Labadarios *et al*., 2005).

The results of the NFCS formed the basis on which decisions on food fortification were made. Regulations pertaining to the mandatory fortification of all maize meal and wheat bread flour were printed in a Government Gazette, published under the Foodstuffs, Cosmetics and Disinfectants Act (No. 54 of 1972) on 7 April 2003 (Department of Health, 2003). These regulations became legally applicable and implementable on 7 October 2003. Since then, it is mandatory for manufacturers to add iron, zinc, vitamin A, thiamine, riboflavin, niacin, folic acid and vitamin B6 to maize and wheat bread flour.

The NFCS was conducted approximately 20 years ago and no national follow-up of children of the same age has been done. For this reason, the PDIS was launched in two provinces as a follow-up (other provinces will be surveyed when funding is secured).

2.2 Sampling

In 2018, a PDIS of one- to nine-year-old children was launched in Gauteng and the Western Cape (Senekal *et al*., 2019). These are regarded as the two most economically active and rapidly urbanising provinces in South Africa, with extensive migration from rural areas to cities in these two provinces in search of jobs and a better quality of life (Statistics South Africa, 2020).

The sampling strategy of the PDIS incorporates a multistage stratified cluster random sampling design, using the methodology applied in demographic and health surveys as described in the USAID *Sampling and household listing manual* (ICF International, 2012). Six strata were identified during the design phase, namely two provinces (Gauteng and Western Cape), with each having three areas of residence: urban formal, urban informal and rural areas. A stratified two-stage sample design was used with a probability proportional to size sampling of enumerator areas (EAs) at the first stage and systematic sampling of households within the EAs at the second stage. A total of 84 EAs were selected from the six strata: 25 formal residential, 10 informal resi-dential and 11 rural EAs in Gauteng, and 18 formal residential, 10 informal residential and 10 rural EAs in the Western Cape, resulting in 1 326 children – 733 in Gauteng and 593 in the Western Cape (Senekal *et al*., 2019).

2.3 Methods: Dietary intake

A single 24-hour recall was done for each participant. For one- to six-year-old children, the mother/caregiver reported on the intake of the child on the previous day, with no input from the child. For seven- to nine-year-old children, the mother/caregiver and child were interviewed together to record the dietary intake during the prior 24 hours. If the child had been at a daycare centre the previous day, it was visited by the fieldworker and the meals and portion sizes determined for the 24 hours in question. All weekdays and Sundays were covered proportionally. The multiple pass method of the 24-hour recall was used to administer the 24-hour recall (Moshfegh *et al*., 2008; Subar *et al*., 2003). The interviewer and respondent reviewed eating episodes of the previous day several times to obtain accurate information about food intake. Commonly, the first step or pass (quick list) is to compile a quick list of foods eaten in the previous 24 hours. In the second pass (forgotten foods), the interviewer reviews the data collected and probes for food that may have been forgotten. In the third pass (time and occasion), the mealtime and eating location are collected. During the fourth pass (detail), detailed descriptions and portion sizes are collected and the time interval between meals is reviewed to again check for additional foods. Clarification is obtained regarding the food portion sizes using household dishes and measures, geometric shapes and food labels (Gibson, 2005). The fifth pass (final) is a last opportunity to remember foods consumed.

The 24-hour recall data were analysed using the South African Food Composition Tables (SAFCT) (Van Graan & Chetty, 2017). As mentioned, a common concern with a single 24-hour recall is the day-to-day variation in the diet of free-living populations. The magnitude of the mostly random within-person variance varies by nutrient and is largely dependent on cultural and ecological factors. Methodological challenges in the estimation of dietary intake may also contribute to within-person error. These errors result in large standard deviations in population groups and insignificant regression coefficients. Another important result of exaggerated variation is that the percentage of subjects below or above specified cut points will be distorted. The NCI method (Tooze *et al*., 2010) that was developed to distinguish within-person from between-person variation accounts for extreme intakes, including zero intake, and allows for adjustment for covariates. Association analysis was applied in this study to estimate the usual dietary intake from repeated 24-hour dietary recall assessments. This study focused on the estimation of single dietary components consumed daily (e.g. calcium) and a ratio of two dietary components consumed daily (e.g. percentage of energy from fat, also referred to as a 'bivariate model'). This application of the NCI method is also referred to as the 'amount-only' method (NCI, 2019).

Two additional 24-hour dietary recalls were completed on a subsample of 148 (second recall) and 146 (third recall) children in the PDIS sample. For logistic reasons, this subsample was recruited from the last five EAs visited in each province. The same houses were revisited and the same children and/or their caregivers were interviewed. Comparison of socio-demographic variables between those who completed one 24-hour recall and those who completed repeated recalls showed only two significant differences, namely marital status and ethnic group. Whether the 24-hour recall was less, the same or more than the child's usual intake was also recorded for the total group as well as for the two additional recalls completed for the subgroup.

The data obtained from the three 24-hour recalls of the subsample were used to adjust the observed distributions of the single 24-hour recall completed by the larger sample for the effects of random within-person variation to establish usual intakes. The steps followed using the amount-only method (Luo *et al*., 2019) were as follows:

Step 1: Input Day 1, Day 2 and Day 3 24-hour recall intakes. A subset of individuals has three repeated recalls on non-consecutive days to ensure that observations within an individual are independent. Preliminary data adjustments are made by setting zero values to half of the minimum amount values (data must contain <5% zero values).

Step 2: Fit the model and Box-Cox transform to near normality. Recalls are modelled, incorporating a Box-Cox transformation, to account for the skewed distributions and the within-person error of the single 24-hour recall measurements. The usual intake calculations in the NCI method require a normal or near-normal distribution. However, nutrient intake data are usually positively skewed, therefore the models of the NCI method use the following Box-Cox functions:

$$g(x;\lambda) = (x^{\lambda} - 1)/\lambda, \text{ when } \lambda \neq 0 \quad (1)$$

$$g(x;\lambda) = \log(x), \text{ when } \lambda = 0 \quad (2)$$

$$g(x;\lambda) = sqrt(x), \text{ when } \lambda = 0.5 \quad (3)$$

to transform the input data to normality or near normality. The lambda variable associated with the Box-Cox transformation determines the strength of the transformation and is calculated as part of the overall model-fitting process.

The covariates included to represent the effect of personal characteristics were the province, the area of residence (formal urban, informal urban and rural areas), the gender of the child and whether the day of the 24-hour recall was a usual day. These are person-specific effects that allow an individual's usual intake to vary between persons. To model within-person variation, variety within the diet included an indicator as a covariate to indicate that the second and third recalls are being modelled, allowing for adjustment for repeat application of the recall (Tooze *et al*., 2006). The analysis was stratified, allowing for the three different age groups (1–2 years, 3–5 years and 6–9 years) and different transformation parameters per strata. This results in better approximations of normality for each age group to prevent highly skewed distributions.

The relationship of the covariates to the reported intakes is estimated by fitting a non-linear mixed effect model. The model can be written as:

$$g(R_{ij};\lambda) = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ki} + \sum_{l=1}^{L} \beta_l Z_{lij} + d_{ij} \quad (4)$$

$$d_{ij} = u_i + e_{ij} \quad (5)$$

where R_{ij} denotes the recall of individual i on Day j, $g(x;\lambda)=(x^{\lambda}-1)/\lambda$ is the Box-Cox transformation, X_{ki} is person i's value of the k-th person-level covariate, Z_{lij} is the value of the l-th temporal covariate for person i on Day j, β_k and β_l are regression coefficients and d_{ij} is a zero-mean regression error that is further decomposed into a zero-mean person-specific effect u_i and a zero-mean within-person error e_{ij}. The transformation is assumed to produce normally distributed terms u_i and e_{ij}, which implies that d_{ij} is also normally distributed. As mentioned previously, with three 24-hour recalls on a subset of individuals, it is possible to disaggregate the total residual variation (the variance of d_{ij}) into between-person and within-person components (the variances of u_i and e_{ij}, respectively) (Luo *et al*., 2019).

The NCI (amount-only) method can be applied via a set of macros written in the SAS programming language. The MIXTRAN macro evaluates the effects of individual covariates on usual intake and generates parameter estimates and linear predictor values used as inputs for the DISTRIB macro (next step) (Tooze *et al*., 2006). The MIXTRAN macro fits a non-linear mixed effects model to the three 24-hour recalls, using the SAS NLMIXED procedure. In this application, the model is for the amount of a nutrient consumed every day (i.e. amount-only) (NCI, 2019).

Step 3: Simulate usual intakes based on the fitted model. The DISTRIB macro incorporates parameter estimates and linear predictor values from the MIXTRAN macro in a Monte Carlo simulation to estimate the distribution of usual intake. The number of simulations per individual is 100.

Step 4: Back-transform to original scale. The simulated intake amounts are then back-transformed to the original scale, using the nine-point numerical integration approximation method, specifically the gauss-hermite quadrature method (as described in the distrib_bivariate.macro.v1.1.sas macro). The back-transformed values represent the usual intake distribution of a simulated population. This is also part of the DISTRIB macro.

Step 5: Derive percentiles and proportions above/below cut points. The estimated average requirement (EAR) value is the daily intake value of a nutrient that is estimated to meet the nutrient requirement of half the healthy individuals in a life stage and gender group. An intake below the EAR is deemed to reflect risk of deficiency in groups (Murphy *et al*., 2016).

After executing the DISTRIB macro, the Percentiles_Survey macro reads the back-transformed usual intake values calculated in the Monte Carlo simulation and calculates the percentiles of usual intake and the percentages below and above given EAR values. Sample weights are considered to ensure that the results are representative of the population.

The balanced repeated replication (BRR) method (Korn & Graubard, 1999) was calculated and used to do variance estimation with a Fay coefficient of 0.3 (Judkins, 1990). Two pseudo primary sampling units (PSUs) were created per stratum by randomly selecting half of the PSUs (or EAs) in each stratum into one pseudo-PSU and the rest in a second pseudo-PSU (Herrick *et al*., 2018). Therefore, six original strata were maintained with 12 pseudo-PSUs – two per stratum. Consequently, eight BRR weights were created, taking the original sampling weights as well as the age and gender of each child into consideration.[1] The SAS programs and the respective macros used are shown in Table 2.

Table 2: List of SAS programs and macros applied to calculate usual intakes

nlmixed_univariate_food1.sas (for the first nutrient component) • nlmixed_univariate_macro_v1.2.sas • boxcox_survey.macro.v1.2.sas
nlmixed_univariate_food2.sas (for kilojoules, if percentage of energy of a nutrient is required) • nlmixed_univariate_macro_v1.2.sas • boxcox_survey.macro.v1.2.sas
nlmixed_bivariate_food1_food2.sas (if percentage of energy of a nutrient is required) • nlmixed_bivariate_macro_v1.2.sas
distrib_bivariate_food1_food2.sas • distrib_bivariate.macro.v1.1.sas; • percentiles_survey.macro.v1.1.sas.
nlmixed_probabilities_bbr.sas (to incorporate balanced repeated replication weights) • brr_pvalue_ci_macro_v1.1.sas

[1] The website accessed can be visited at the following link: https://prevention.cancer.gov/research-groups/biometry/measurement-error-impact/software-measurement-error (US Department of Health and Human Services, 2019). The software selected is for estimating usual intake distribution, specifically for single regularly consumed nutrients.

2.4 Results of dietary intake analyses

The NCI method was applied to the PDIS data set (Senekal *et al*., 2020; Steyn *et al*., 2020) for selected macronutrients and micronutrients. The micronutrients will be addressed in this paper.

Equation (4) and equation (5) describe how using multiple 24-hour recalls on a subset of individuals can be used to disaggregate the total residual variation (the variance of d_{ij}) into between-person and within-person components (the variances of u_i and e_{ij}, respectively) (Luo *et al*., 2019). Table 3 shows the actual values of the between-person and within-person variances, by age group, for each of the relevant micronutrients. It is important to calculate the ratio of within-person to between-person variance. The results of the two provinces are combined.

Table 3: The relationship between within-person and between-person variation, by age group

Nutrient (Box-Cox transformation parameter)	Age: 1–2 years			Age: 3–5 years			Age: 6–9 years		
	Var_U*	Var_E	Ratio	Var_u	Var_E	Ratio	Var_u	Var_E	Ratio
Calcium (λ = 0.24)	5.78	9.24	1.60	1.42	9.39	6.61	3.19	5.99	1.88
Iron (λ = 0.28)	0.42	0.43	1.02	0.11	0.45	4.09	0.19	0.4	2.11
Zinc (λ = 0.23)	0.22	0.33	1.50	0.11	0.38	3.45	0.13	0.35	2.69
Vitamin A (λ = 0.01)	0.2	0.78	3.90	0.05	1.1	22.00	0.34	0.66	1.94
Vitamin D (λ = 0.26)	1.11	1.82	1.64	0.01	2.71	271.00	0.67	2.14	3.19
Vitamin E (λ = 0.16)	0.36	1.04	2.89	0.41	0.86	2.10	0.47	0.86	1.83
Vitamin C (λ = 0.29)	4.31	5.92	1.37	0.98	7.3	7.45	2.68	5.59	2.09
Thiamine (λ = 0.29)	0.1	0.1	1.00	0.06	0.1	1.67	0.07	0.09	1.29
Niacin (λ = 0.41)	0.2	1.7	8.50	0.57	1.32	2.32	0.7	1.49	2.13
Riboflavin (λ = 0.22)	0.18	0.27	1.50	0.06	0.27	4.50	0.17	0.19	1.12
Vitamin B6 (λ = 0.21)	0.06	0.23	3.83	0.06	0.3	5.00	0.14	0.29	2.07
Vitamin B12 (λ = 0.13)	0.62	1.27	2.05	0.03	2.48	82.67	0.71	2.01	2.83
Folate (λ = 0.10)	0.27	0.75	2.78	0.24	0.87	3.63	0.34	0.69	2.03

*Var_U: Between-person variation; Var_E: Within-person variation; Ratio: Ratio of within-person to between-person variance

According to Costa *et al.* (2008), the higher the ratio between within-person and between-person variance, the more records are needed. Vitamin A was described by Tooze *et al.* (2010) as a nutrient with a high ratio of within-person to between-person variance (in the order of 4:1), resulting in a distribution that is highly skewed and more difficult to transform to approximate normality. Similar consclusions were made in this study for one- to two-year-olds, with the low lambda value of almost zero and the high variance ratio of 3.9. Calcium was described by Tooze *et al.* (2006) as "well behaved", with a variance ratio around 1.5 and less skew than other nutrients. The conclusions here were similar (for one- to two-year-olds), with a variance ratio of 1.6 and a lambda value of 0.24.

Figure 1 demonstrates the distribution of calcium, for which the Box-Cox transformation resulted in a lambda value of 0.26. This distribution of Day 1 intake is more skew to the right and the distribution of the usual intake of calcium limits tail values. The mean values of Day 1 and usual intakes are similar, but the median value of the usual intake moved to the right. The EAR is above the mean and median values for Day 1 intakes as well as usual intakes. Consequently, the percentage below the EAR is higher for usual intake.

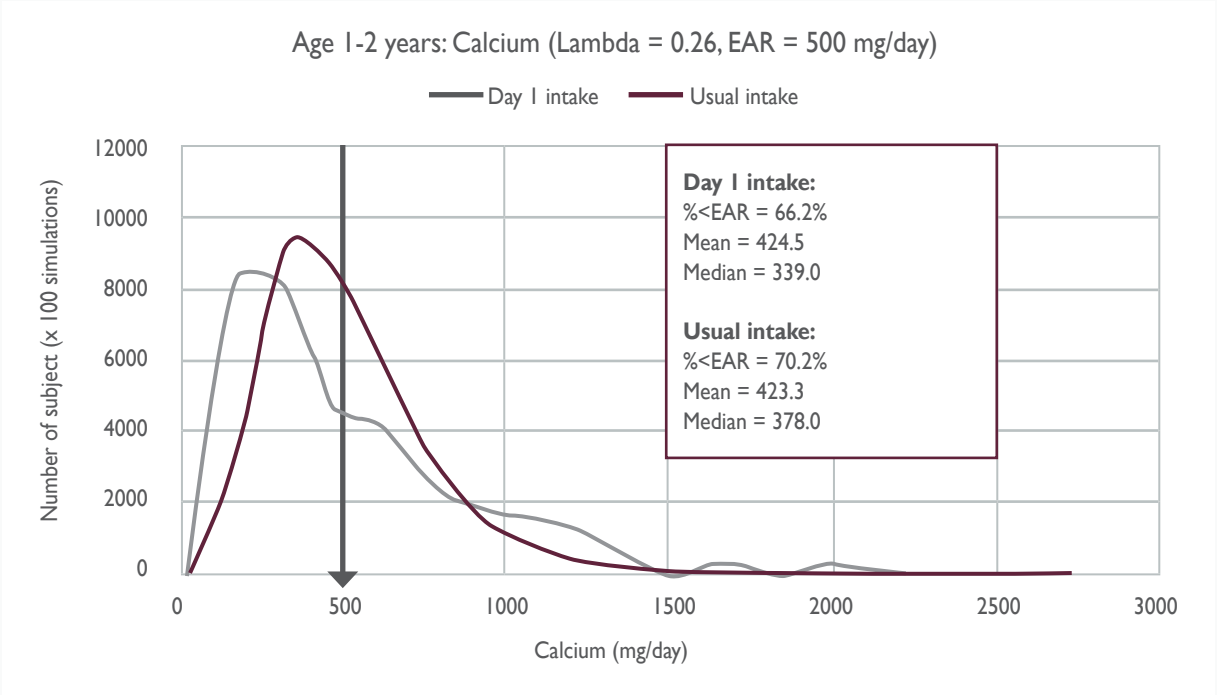


Figure 1: Distribution of Day 1 and usual intakes of calcium, children aged 1–2 years

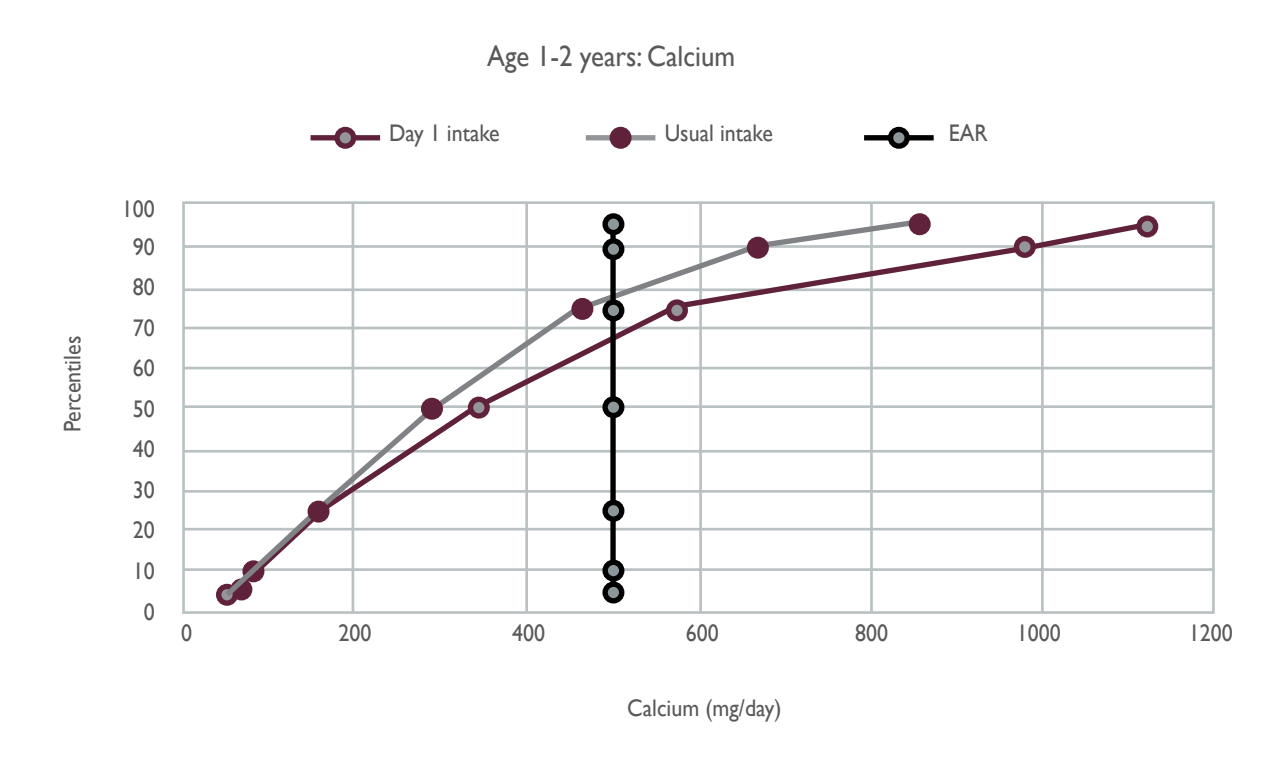


Figure 2: Cumulative distribution of Day 1 and usual intakes of calcium, children aged 1–2 years

Figure 2 demonstrates for calcium the spread of the distributions of the Day 1 intake vs. the smaller spread of the usual intake as well as the proportion below the EAR. For the Day 1 intake, the proportion below the EAR value is 66%, and for the usual intake, the proportion below the EAR is 70% (the slight difference is a result of the inclusion of the BRR weights, which were not incorporated in the graph). The ratio of the within-person variance to between-person variance is 1.6, and the within variation mostly affected the higher intake values.

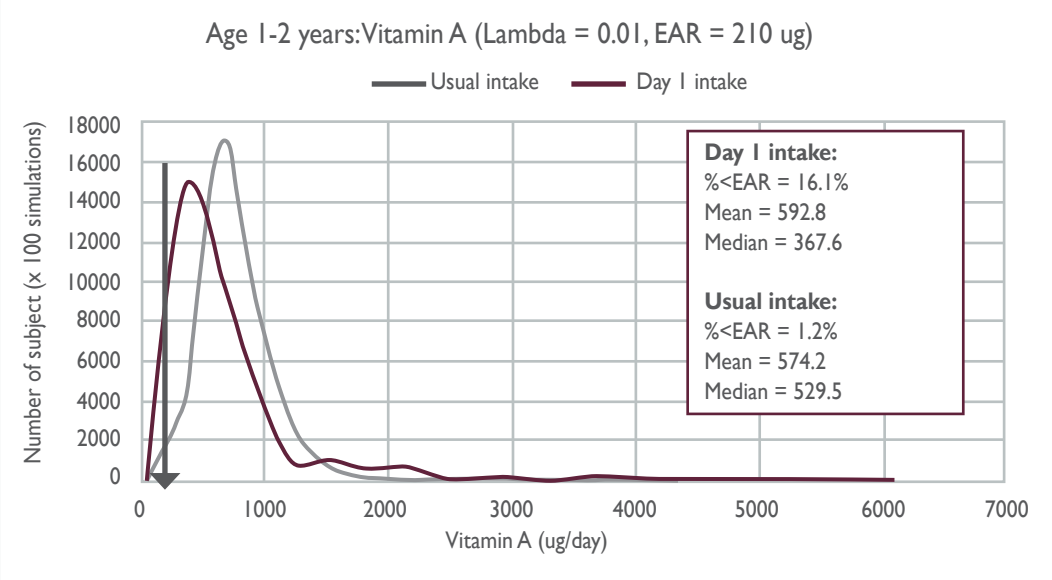


Figure 3: Distribution of Day 1 and usual intakes of vitamin A, children aged 1–2 years

In the case of vitamin A (Figure 3), the lambda value is almost zero, indicating that the distribution of Day 1 intakes is extremely skew and that the results of the transformation and the usual intake should be used with caution. There is a large positive shift in the median values from the Day 1 to the usual intake, and the percentage below the EAR is only 1.2% for the usual intake as opposed to 16.1% for the Day 1 intake. The minimum lambda bound is set at 0.01 for vitamin A, because the optimum lambda values for the Box-Cox transformation were outside the permitted lambda bounds. Other studies also obtained the same lambda values for vitamin A (Australian Bureau of Statistics, 2015).

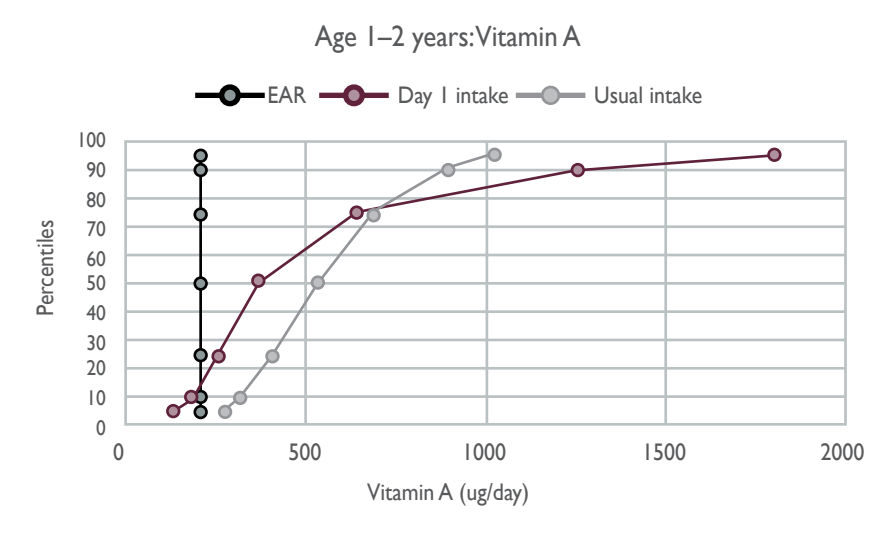


Figure 4: Cumulative distribution of Day 1 and usual intakes of vitamin A, children aged 1–2 years

Figure 4 demonstrates how the spread of the usual intake is decreased when the within-person variance is taken care of. Both the tails at the lower and upper ends were decreased. The variance ratio is 3.9, indicating the high impact of within-person variation on the Day 1 intakes. After getting rid of the tails, as a result of the high within-person variation, the percentage below the EAR reduced from 16.1% to 1.2%.

Tables 4, 5 and 6 summarise the differences between Day 1 intake and usual intake per age group for 13 micronutrients for which EAR values are available. Prevalence of inadequate intakes (%<EAR) is also included. For this paper, prevalence of excessive intake was not considered.

Vitamin A (Table 5), vitamin B12 and folate (Table 6) have lambda values <0.15, indicating that the distributions are skew to the right (mean values>median values) and that the results should be treated with caution. The skew distributions and the large within-person variation, as reflected in high ratios of within-person to between-person variance, especially in the three- to five-year age group, are indicative of the problems experienced with modelling these nutrients. The prevalences of inadequacy (%<EAR) for these nutrients differ by more than 10% between Day 1 and usual intakes for all age groups. Therefore, it will not be advised to analyse these nutrients using Day 1 intakes only, not even for larger samples.

Calcium (Table 4) and vitamin D (Table 5) intake values are still unacceptably low. Note that the %<EAR for these two nutrients are higher for usual intake (the median intake values are lower than the EAR). Again, large within-person variations, especially in the three- to five-year age group, were observed.

Vitamins E and C (Table 5) and riboflavin (Table 6) have a large difference between %<EAR for Day 1 and usual intake. Again, the three- to five-year age group reflects the highest ratio of with-in-person to between-person variance.

Table 4: Mean, median and percentage intake below the EAR for mineral intake (usual intake and Day 1 intake) of children aged 1–9 years

		Age 1–2 years N = 333		Age 3–5 years N = 514		Age 6–9 years N = 479	
Nutrient		Usual intake	Day 1 intake	Usual intake	Day 1 intake	Usual intake	Day 1 intake
Calcium (mg/day) EAR:	Mean (SE)	423.3 (45.3)	424.5 (19.1)	350.9 (6.4)	348.5 (19.2)	351.8 (13.6)	352.2 (13.5)
1–3 yrs = 500 mg	Median (SE)	378.0 (39.2)	339.0 (19.9)	329.2 (4.4)	288.3 (22.9)	331.3 (15.0)	299.9 (15.0)
4–8 yrs = 800 mg							
9–<10 yrs = 1 100 mg	%<EAR	70.2	66.2	94.8	87.3	99.4	95.9
(lambda = 0.24)							
Iron (mg/day) EAR:	Mean	7.8 (0.5)	7.7 (0.3)	8.9 (0.1)	8.9 (0.3)	10.6 (0.1)	10.6 (0.2)
1–3 yrs = 3.0 mg	Median	7.3 (0.4)	7.2 (0.3)	8.8 (0.1)	8.6 (0.3)	10.3 (0.1)	9.7 (0.2)
4–8 yrs = 4.1 mg							
9–<10 yrs = 5.9 mg	%<EAR	1.0	3.4	0.01	2.7	0.3	2.6
(lambda = 0.28)							
Zinc (mg/day) EAR:	Mean	6.5 (0.4)	6.4 (0.2)	7.3 (0.1)	7.3 (0.2)	8.5 (0.2)	8.5 (0.2)
1–3 yrs = 2.2 mg	Median	6.2 (0.4)	6.0 (0.2)	7.1 (0.1)	6.8 (0.3)	8.3 (0.1)	7.9 (0.2)
4–8 yrs = 4.0 mg							
9–10 yrs = 7.0 mg	%<EAR	0.1	1.7	0.5	8.9	4.9	12.4
(lambda = 0.23)							

A lambda value <0.15 reflects a larger mean bias as a result of sensitivity to the transformation applied.

Table 5: Mean, median and percentage intake below the EAR for vitamins A, D, E and C (usual intake and Day 1 intake) of children aged 1–9 years

		Age 1–2 years N = 333		Age 3–5 years N = 514		Age 6–9 years N = 479	
Nutrient		**Usual intake**	**Day 1 intake**	**Usual intake**	**Day 1 intake**	**Usual intake**	**Day 1 intake**
Vitamin A (ug/day) EAR: 1–3 yrs = 210 ug 4–8 yrs = 275 ug 9–<10 yrs = 445 ug (lambda = 0)	Mean	574.2 (67.5)	592.8 (41.5)	607.0 (23.6)	639.2 (50.2)	623.8 (61.7)	694.3 (58.8)
	Median	529.5 (54.3)	367.6 (22.2)	580.5 (47.8)	400.7 (19.9)	550.3 (31.7)	433.2 (23.7)
	%<EAR	1.2	16.1	0.4	24.5	12.0	29.4
Vitamin D (ug/day) EAR: 1–<10 yrs = 10 ug (lambda = 0.26)	Mean	2.8 (0.3)	2.9 (0.3)	2.4 (0.1)	2.4 (0.2)	3.3 (0.1)	3.2 (0.2)
	Median	2.2 (0.3)	1.1 (0.1)	2.3 (0.1)	1.2 (0.1)	2.9 (0.2)	2.0 (0.2)
	%<EAR	98.2	79.2	100.0	85.0	99.3	78.0
Vitamin E (mg/day) EAR: 1–3 yrs = 5 mg 4–8 yrs = 6 mg 9–<10 yrs = 9.0 mg (lambda = 0.16)	Mean	8.1 (0.3)	7.9 (0.5)	8.2 (0.2)	8.2 (0.4)	11.1 (0.4)	11.0 (0.5)
	Median	7.3 (0.3)	6.2 (0.3)	7.5 (0.4)	6.0 (0.3)	10.1 (0.5)	8.2 (0.4)
	%<EAR	18.2	36.2	26.9	46.6	18.8	35.2
Vitamin C (mg/day) EAR: 1–3 yrs = 13 mg 4–8 yrs = 22 mg 9–<10 yrs = 39 mg (lambda = 0.29)	Mean	47.6 (2.9)	46.6 (3.4)	39.4 (1.5)	40.8 (3.4)	42.4 (2.9)	43.6 (3.8)
	Median	40.2 (2.6)	32.7 (4.0)	36.6 (1.4)	23.6 (2.0)	37.2 (2.0)	27.3 (1.7)
	%<EAR	7.4	21.3	9.0	39.1	25.3	40.3

A lambda value <0.15 reflects a larger mean bias as a result of sensitivity to the transformation applied.

Table 6: Mean, median and percentage intake below the EAR for B vitamins (usual intake and Day 1 intake) of children aged 1–9 years

		Age 1–2 years N = 333		Age 3–5 years N = 514		Age 6–9 years N = 479	
Nutrient		**Usual intake**	**Day 1 intake**	**Usual intake**	**Day 1 intake**	**Usual intake**	**Day 1 intake**
Thiamine (mg/day) EAR: 1–3 yrs = 0.4 mg 4–8 yrs = 0.5 mg 9–<10 yrs = 0.7 mg (lambda = 0.29)	Mean	1.0 (0.03)	1.0 (0.04)	1.0 (0.01)	1.0 (0.03)	1.2 (0.03)	1.2 (0.02)
	Median	0.9 (0.01)	0.9 (0.04)	1.0 (0.02)	0.9 (0.03)	1.1 (0.03)	1.1 (0.04)
	%<EAR	1.6	4.8	0.7	6.9	1.3	4.8
Niacin (mgNE/day) EAR: 1–3 yrs = 5.0 mg NE 4–8 yrs = 6.0 mg NE 9–<10 yrs = 9.0 mg NE (lambda = 0.41)	Mean	11.5 (0.4)	11.5 (0.4)	14.2 (0.2)	14.2 (0.4)	17.2 (0.4)	17.3 (0.4)
	Median	11.3 (0.3)	10.6 (0.4)	13.8 (0.2)	13.2 (0.4)	16.8 (0.4)	16.7 (0.5)
	%<EAR	0.05	10.7	0.3	5.9	0.5	5.4
Riboflavin (mg/day) EAR: 1–3 yrs = 0.4 mg 4–8 yrs = 0.5 mg 9–<10 yrs = 0.8 mg (lambda = 0.22)	Mean	0.9 (0.1)	0.9 (0.04)	0.9 (0.02)	0.9 (0.03)	1.0 (0.03)	1.0 (0.04)
	Median	0.8 (0.1)	0.8 (0.04)	0.9 (0.02)	0.8 (0.05)	0.9 (0.02)	0.9 (0.04)
	%<EAR	6.0	17.5	2.3	19.5	11.4	23.4
Vitamin B_6 (mg/day) EAR: 1–3 yrs = 0.4 mg 4–8 yrs = 0.5 mg 9–<10 yrs = 0.8 mg (lambda = 0.21)	Mean	1.4 (0.05)	1.4 (0.05)	1.8 (0.01)	1.8 (0.04)	2.5 (0.04)	2.5 (0.1)
	Median	1.3 (0.04)	1.2 (0.1)	1.8 (0.02)	1.7 (0.05)	2.4 (0.04)	2.2 (0.1)
	%<EAR	0.0	2.5	0.0	2.7	0.0	0.8
Vitamin B_{12} (ug/day) EAR: 1–3 yrs = 0.7 ug 4–8 yrs = 1.0 ug 9–<10 yrs = 1.5 ug (lambda = 0.13)	Mean	2.2 (0.1)	2.3 (0.3)	2.9 (0.1)	3.3 (0.4)	4.3 (0.4)	4.7 (0.6)
	Median	1.7 (0.5)	1.1 (0.1)	2.9 (0.2)	1.3 (0.1)	3.5 (0.3)	1.7 (0.1)
	%<EAR	14.3	34.4	0.0	36.7	5.4	35.0
Folate (ug/day) EAR: 1–3 yrs = 120 ug 4–8 yrs = 160 ug 9–<10 yrs = 250 ug (lambda = 0.1)	Mean	225.4 (12.7)	225.0 (12.1)	253.2 (4.0)	253.2 (11.6)	282.1 (7.5)	284.6 (7.9)
	Median	210.1 (10.5)	200.0 (9.8)	238.7 (9.9)	202.3 (11.8)	266.0 (5.5)	242.9 (5.5)
	%<EAR	9.6	22.9	10.3	26.5	15.5	27.8

A lambda value <0.15 reflects a larger mean bias as a result of sensitivity to the transformation applied.

Figures 5, 6 and 7 graphically show the prevalence of intake below the EAR in all three age groups for the selected minerals and vitamins.

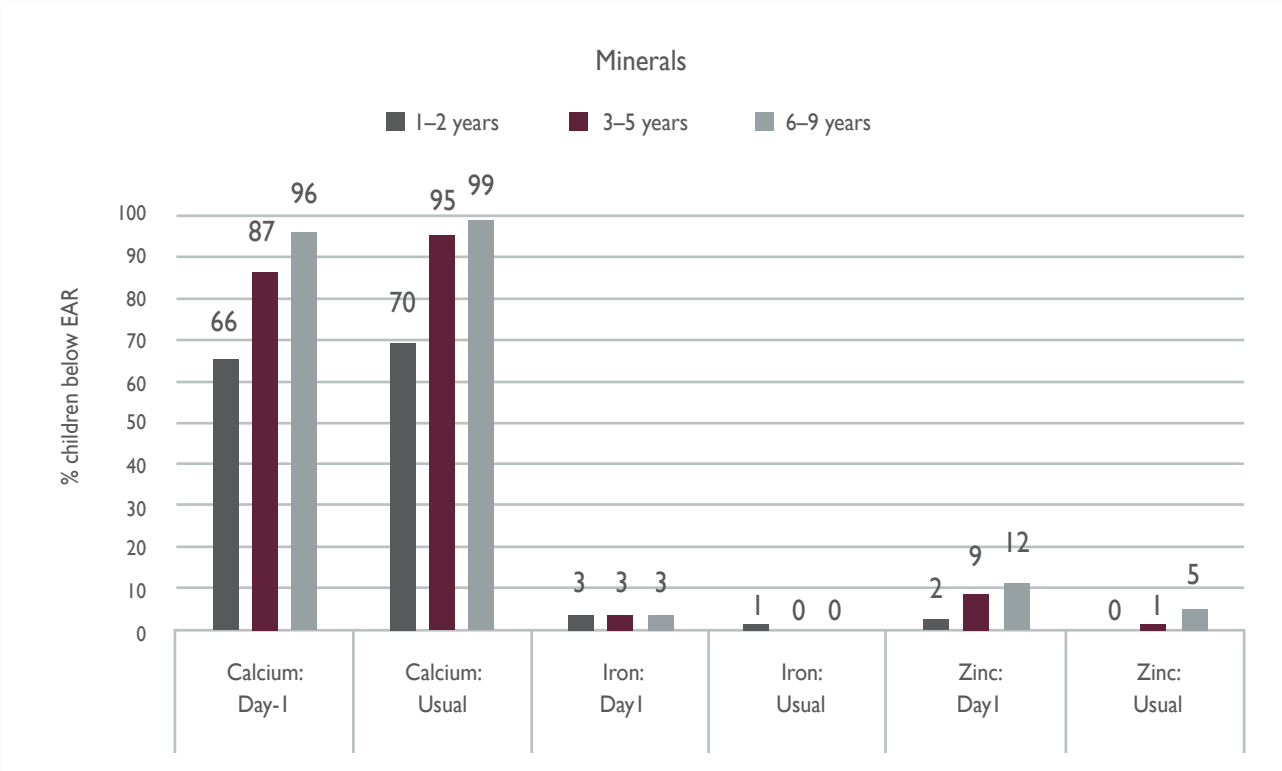


Figure 5: Percentage below EAR for mineral intake

Figure 5 reflects the high %<EAR for calcium relative to the other minerals. Intakes of iron and zinc appear to be adequate in both Day 1 and usual intakes.

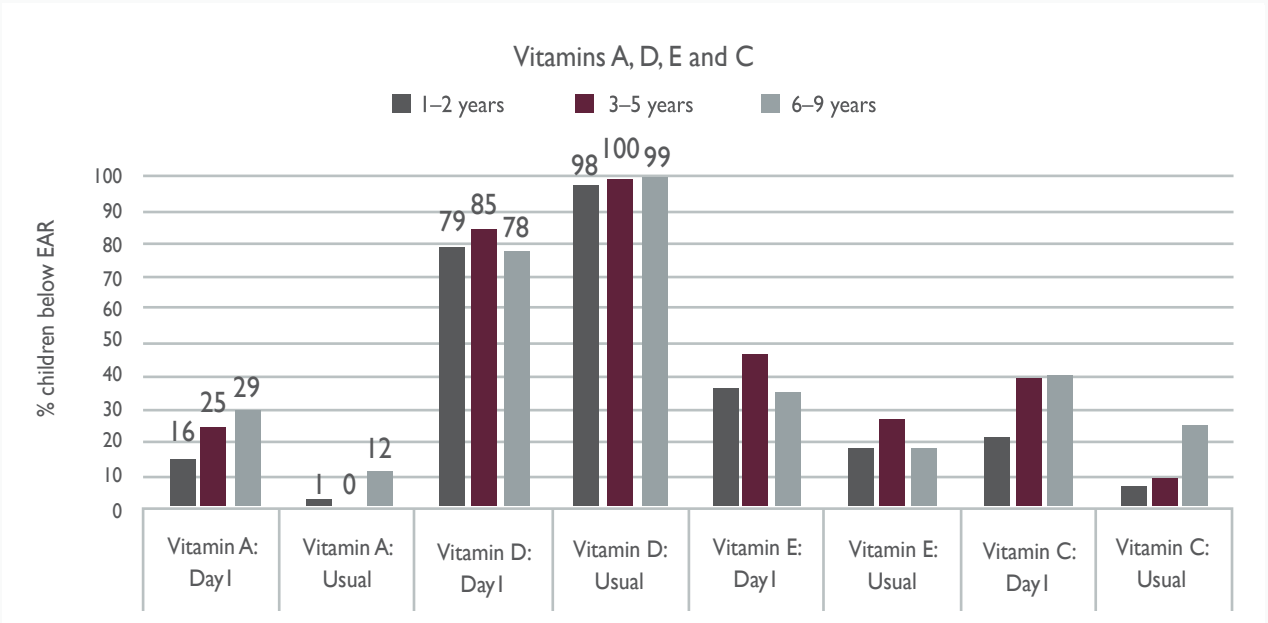


Figure 6: Percentage below EAR for the intake of vitamins A, D, E and C

The usual intakes of vitamin A and C appear to be adequate in the younger children, but for the six- to nine-year-old children, the %<EAR is still estimated at 12% for vitamin A and 25% for vitamin C (Figure 6).

Of note is that for 16% of food items included in the SAFCT (Van Graan & Chetty, 2017), the vitamin D content values are not yet known (missing) (similar for vitamin E), resulting in unavoidable underestimations of dietary vitamin D intake (personal communication with Ms J Chetty of the SAFCT division, South African Medical Research Council). Results on the adequacy of vitamin D and E intake should therefore be interpreted with caution.

In Figure 6, the high %<EAR for vitamin D, especially for the usual intake, is evident.

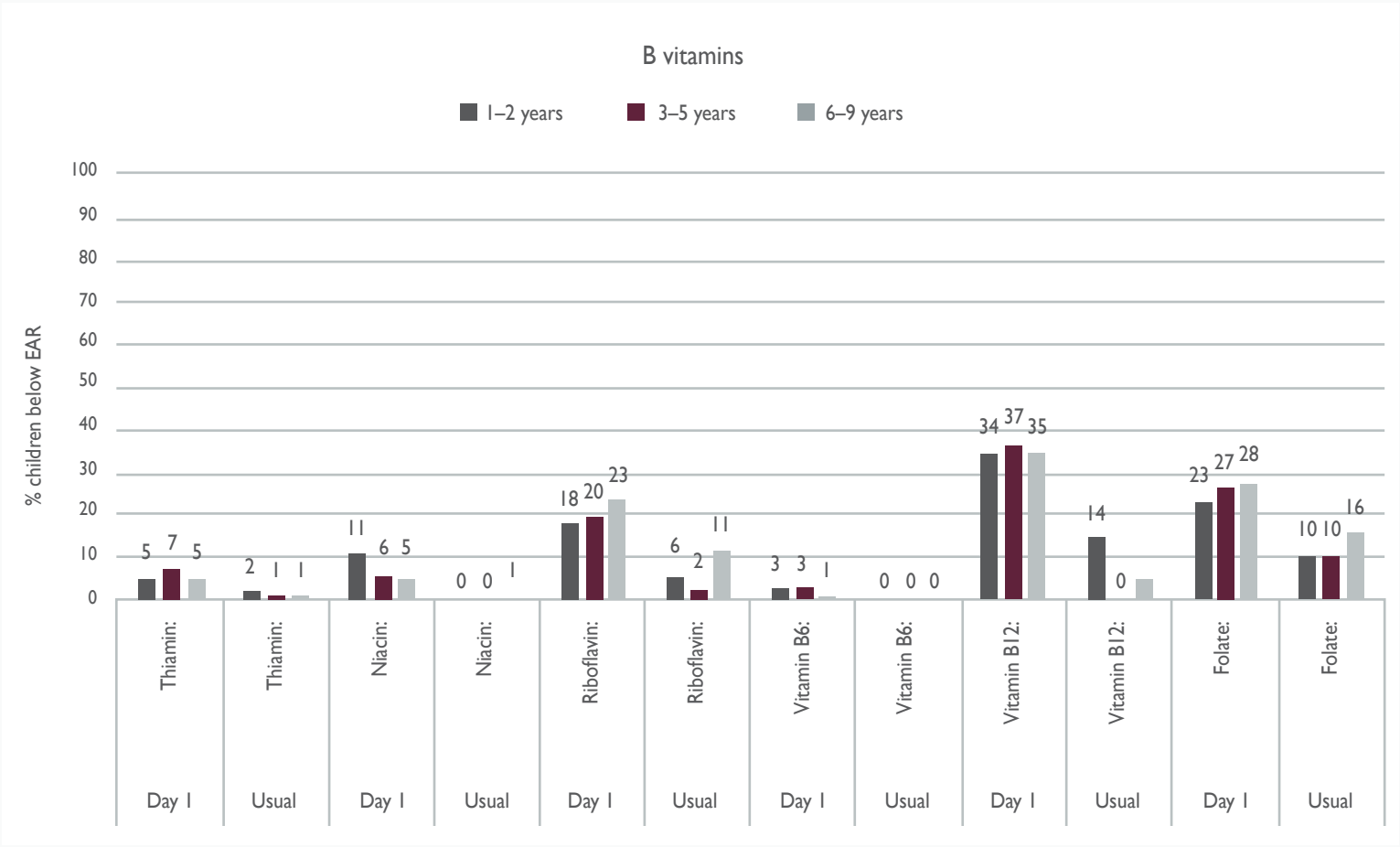


Figure 7: Percentage below EAR for the intake of the B vitamins

Figure 7 reflects the %<EAR for the B vitamins. Riboflavin, vitamin B12 and folate reflect a difference of more than 10% below %<EAR, between the Day 1 and usual intakes. The usual intakes appear to be adequate, but 16% of children in the six- to nine-year group have %<EAR for folate.

3. DISCUSSION AND CONCLUSION

The PDIS is the first study in South Africa to implement statistical methods to investigate how adjustment for within-person variation affected nutrient intake and dietary adequacy of selected nutrients. The NCI method was implemented to correct for within-person variation for a better approximation of usual intake. Overall, after including a second and third day of dietary intake and adjusting for selected covariates, a reduced variability was reported for all estimates, as shown by the narrower tails of the distributions across all age groups. The NCI method was originally developed to be applied to the US Department of Agriculture's National Health and Nutrition Examination Surveys, with its unique PSU, strata system and weighting factors, and where the majority of the participants usually provide two days of dietary intake (Piernas *et al*., 2015). The PDIS has a different PSU scheme and included two additional days of intake for approximately 11% of the total sample. The smaller subsample might affect the approximation of the within-person variance. In addition, we had to design a unique set of BRR weights (Herrick *et al*., 2018). These methodological and technical challenges were faced, however.

Regarding micronutrient intakes, we found small or no differences in the mean values of the Day 1 and usual intakes. Even though relatively accurate estimates of mean nutrient intakes at the population level are obtained using only Day 1 intakes, the effect of day-to-day, or within-person, variation is mainly reflected by the overestimated standard errors and high within-person to between-person variance ratios. Therefore, estimates of the prevalence of inadequate intake can be biased if only one day of intake is available. Compared with usual intake, estimates from Day 1 intake in the PDIS tended to yield overestimated values of inadequate micronutrient intakes, except for calcium and vitamin D. This is because for calcium and vitamin D, median intakes are larger than EAR values.

Prevalence of dietary insufficiency could be overestimated if intakes derived from a single 24-hour recall are not adjusted. The NCI method allows for estimation of population exposure without conducting repeated 24-hour recalls on the total sample. The PDIS demonstrates the importance of repeated 24-hour recalls to adjust for within-person variability to study dietary adequacy of critical nutrients. This study also shows that a small subsample (11%) of repeated 24-hour recalls may be sufficient to conduct the described adjustments. The study also demonstrates that the usual intake distributions are narrower, hence the prevalence of inadequate intake may be biased when estimating nutrient adequacies if one day of dietary data is used.

Funding: This research was mainly funded by the International Life Sciences Institute of South Africa and their industry partners, the National Research Foundation and UCT.

Acknowledgements: The author acknowledges Profs Nelia Steyn and Marjanne Senekal as co-workers on the project.

Conflict of interest: The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of the data; in the writing of the paper; or in the decision to publish the results.

REFERENCE LIST

Afshin, A., Sur, P.J., Fay, K.A. *et al*. 2019. Health effects of dietary risks in 195 countries, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 393:1958–1972.

Australian Bureau of Statistics. 2015. *Australian Health Survey: User's guide, 2011-13*. Available at: https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4363.0.55.001Chapter651042011-13 [Accessed 21 August 2020].

Basiotis, P., Welsh, S.O., Cronin, F.J., Kelsay, J.L. & Mertz, W. 1987. Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence. *The Journal of Nutrition*, 117(9):1638–1641.

Bingham, S. 2003. Urine nitrogen as a biomarker for the validation of dietary protein intake. *The Journal of Nutrition*, 133(3), Suppl. 3:921S–924S.

Costa, M. de M.F., Takeyama, L., Voci, S.M., Slater, B. & Silva, M.V. 2008. Within- and be-tween-person variations as determinant factors to calculate the number of observations to estimate usual dietary intake of adolescents. *Revista Brasileira de Epidemiologia*, 11(4):541–548.

Department of Health. 2003. *Regulations relating to the fortification of certain foodstuffs*. Government Notice No. R 2003. Pretoria.

Gibson, R. 2005. *Principles of nutritional assessment*. New York, NY: Oxford University Press.

Herrick, K., Rossen, L., Parsons, R. & Dodd, K. 2018. Estimating usual dietary intake from National Health and Nutrition Examination Survey data using the National Cancer Institute method. National Center for Health Statistics. *Vital Health Statistics*, 2(178).

ICF International. 2012. *Demographic and health survey sampling and household listing manual*. Calverton, MD.

International Dietary Data Expansion Project. 2020. *Weighed food record (WFR)*. Available at: https://inddex.nutrition.tufts.edu/data4diets/data-source/weighed-food-record-wfr [Accessed 18 August 2020].

Johnson, R. 2002. Dietary intake: How do we measure what people are really eating? *Obesity Research*, 10, Suppl. 1: 63S–68S.

Judkins, D. 1990. Fay's method for variance estimation. *Journal of Official Statistics*, 6(3):223–239.

Katz, D. & Meller, S. 2014. Can we say what diet is best for health? *Annual Review of Public Health*, 35:83–103.

Knuppel, S., Norman, K. & Boeing, H. 2019. Is a single 24-hour dietary recall per person sufficient to estimate the population distribution of usual dietary intake? *The Journal of Nutrition*, 149(9):1491–1492.

Korn, E. & Graubard, B. 1999. *Analysis of health surveys*. New York, NY: Wiley.

Labadarios, D., Steyn, N.P., Maunder, E. *et al*. 2005. The National Food Consumption Survey (NFCS): South Africa 1999. *Public Health Nutrition*, 8(5):533–543.

Lee, R. & Nieman, D. 2013. *Nutritional assessment*. New York, NY: McGraw-Hill.

Luo, H., Dodd, K., Arnold, C. & Engle-Stone, R. 2019. A new statistical method for estimating usual intakes of nearly-daily consumed foods and nutrients through use of only one 24-hour dietary recall. *The Journal of Nutrition*, 149(9):1667–1673.

Moshfegh, A.J., Rhodes, G.G., Baer, D.J. *et al*. 2008. The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. *American Journal of Clinical Nutrition*, 88:324–332.

Murphy, A., Yates, A.A., Atkinson, S.A., Barr, S.I. & Dwyer, J. 2016. History of nutrition: The long road leading to the dietary reference intakes for the United States and Canada. *Advances in Nutrition*, 7(1):157–168.

NCI (National Cancer Institute). 2019. *User's guide for analysis of usual intakes: For use with versions 2.1 of the MIXTRAN, DISTRIB, and INDIVINT SAS* macros*. Available at: https://epi.grants.cancer.gov/diet/usualintakes/Users_Guide_v2.1.pdf [Accessed 25 February 2019].

NCI (National Cancer Institute). 2020. *Usual dietary intakes*. Division of Cancer Control & Population Sciences. Available at: https://epi.grants.cancer.gov/diet/usualintakes/#overview [Accessed 24 July 2020].

Piernas, C., Miles, D.R., Deming, D.M., Reidy, K.C. & Popkin, B.M. 2015. Estimating usual intakes mainly affects the micronutrient distribution among infants, toddlers and pre-schoolers from the 2012 Mexican National Health and Nutrition Survey. *Public Health Nutrition*, 19(6):1017–1026.

Senekal, M., Nel, J.H., Malczyk, S., Drummond, L., Harbron, J. & Steyn, N.P. 2019. Provincial Dietary Intake Study (PDIS): Prevalence and sociodemographic determinants of the double burden of malnutrition in a representative sample of 1 to under 10-year-old children from two urbanized and economically active provinces in South Africa. *International Journal of Environmental Research and Public Health*, 16(3334):1–27.

Senekal, M., Nel, J.H., Malczyk, S., Drummond, L. & Steyn, N.P. 2020. Provincial Dietary Intake Study (PDIS): Micronutrient intakes of children in a representative/random sample of 1- to <10-year-old children in two economically active and urbanized provinces in South Africa. *International Journal of Environmental Research and Public Health*, 17(5924):1–27.

Statistics South Africa. 2020. *Mid-year population estimates 2020*. Pretoria.

Steyn, N.P., Nel, J.H., Malczyk, S., Drummond, L. & Senekal, M. 2020. Provincial Dietary Intake Study (PDIS): Energy and macronutrient intakes of children in a representative/random sample of 1–<10-year-old children in two economically active and urbanized provinces in South Africa. *International Journal of Environmental Research and Public Health*, 17(1717):1–37.

Subar, A.F., Kipnis, V., Troiano, R.P. *et al*. 2003. Using intake biomarkers to evaluate the ex-tent of dietary misreporting in a large sample of adults: The OPEN Study. *American Journal of Epidemiology*, 158(1):1–13.

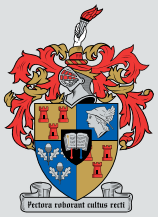
Tooze, J.A., Kipnis, V., Buckman, D.W. *et al*. 2010. A mixed-effects model approach for estimating the distribution of usual intake of nutrients: The NCI method. *Statistics in Medicine*, 29:2857–2868.

Tooze, J.A., Midthune, D., Dodd, K.W. *et al*. 2006. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *Journal of the American Diet Association*, 106(10): 1575–1587.

US Department of Health and Human Services. 2019. *Software for measurement error in nutrition research*. Available at:https://prevention.cancer.gov/research-groups/biometry/measurement-error-impact/software-measurement-error [Accessed 18 September 2019].

Van Graan, A. & Chetty, J.L.M. 2017. *Food Composition Tables for South Africa*. Fifth edition. Cape Town: South African Medical Research Council.

Wark, P.A., Hardie, L.J., Frost, G.S. *et al*. 2018. Validity of an online 24-h recall tool (my-food24) for dietary assessment in population studies: Comparison with biomarkers and standard interviews. *BMC Medicine*, 16(136):1–14.



UNIVERSITEIT • iYUNIVESITHI • STELLENBOSCH • UNIVERSITY

100
1918 · 2018