Methodology for the analysis of dietary data from the Mexican National Health and Nutrition Survey 2006

Sonia Rodríguez-Ramírez, MSc, (1) Verónica Mundo-Rosas, MSc, (1)
Alejandra Jiménez-Aguilar, MSc, (1) Teresa Shamah-Levy, MSc, (1)

Abstract
Objective. To describe the methodology for the analysis of dietary data from the Mexican National Health and NutritionSurvey 2006 (ENSANUT 2006) carried out in Mexico. Material and Methods. Dietary data from the population who participated in the ENSANUT 2006 were collected through a 7-day food-frequency questionnaire. Energy and nutrient intake of each food consumed and adequacy percentage by day were also estimated. Intakes and adequacy percentages > 5 SDs from the energy and nutrient general distribution and observations with energy adequacy percentages < 25% were excluded from the analysis. Results. Valid dietary data were obtained from 3 552 children aged 1 to 4 years, 8 716 children aged 5 to 11 years, 8 442 adolescents, 15 951 adults, and 3 357 older adults. Conclusions. It is important to detail the methodology for the analysis of dietary data to standardize data cleaning criteria and to be able to compare the results of different studies.

Key words: energy, nutrient intake analysis; nutrition surveys; Mexico

Received on: April 11, 2008 • Accepted on: November 19, 2008
Address reprint requests to: Mtra. Verónica Mundo. Instituto Nacional de Salud Pública, Av. Universidad 655, col. Santa María Ahuacatitlán. 62100 Cuernavaca, Morelos, México.
E-mail: vmundo@insp.mx
One of the main reasons for collecting dietary data through national surveys is to have an integral evaluation of the population nutrition status and to be able to provide information about the average dietary intake of a specific population group. The use of dietary data can be broadly classified into three categories: monitoring and nutrition surveillance, analysis of the relationship between diet and disease, and evaluation of the impact of policies and programs related to nutrition.

Using methods like 24-hour recall or food frequency allows the identification of dietary indicators that can be employed to compare the intake of nutrient-rich food and its association with risk factors. They also help to establish food-intake patterns in the population.

Care in dietary data processing (i.e. collection, input of data, cleaning and analysis) must be taken into consideration when designing national surveys to ensure data will be accurate and valid, which will help to adequately estimate energy and nutrient intakes.

Dietary data from several national nutrition surveys are available in Mexico. Dietary information from children and women of reproductive age was obtained in the 1988 survey, whereas dietary information from pre-school and school children, as well as from women of reproductive age, was collected in the 1999 national survey. The most recent available data come from the Mexican National Health and Nutrition Survey 2006 (ENSAUN 2006), in which dietary information was collected from pre-school and school children, adolescents, adults, and older adults, for both sexes. Data from this survey are representative at the national, regional, and local urban and rural levels.

It has become necessary to timely document the tools and processes created to analyze, in a confident manner, Mexican population dietary data taken from probabilistic national samples. This paper aims to describe the methodology used for the analysis of dietary data from the ENSANUT 2006.

**Material and Methods**

**Study population:** We included data from pre-school children aged 1 to 4 years, school children aged 5 to 11 years, adolescents aged 12 to 19 years, adults aged 20 to 59 years, and older adults aged 60 years and over.

**Sample selection:** The ENSANUT 2006 was a nationally representative probabilistic survey stratified by clusters. It was carried out from October 2005 to May 2006. Data from 48,304 households were obtained. Collected information included general data about households, health, anthropometry, biochemical tests, and dietary quality and quantity. For determining sample size it was considered that the minor importance proportion (minimum interest prevalence) should have a value of 8.1%. It was also considered that state estimators obtained through the survey should have a 25% maximum relative error, a 95% confidence interval, a 20% nonresponse rate, and a design effect of 1.7, determining a sample size of at least 1,476 households by state. A subsample corresponding to a third of the whole sample was selected for dietary information.

Informed consent was obtained from each subject or subject’s parent or guardian for their participation in the study. The survey protocol was approved by the Ethics Committee of the National Institute of Public Health, Mexico.

**Dietary data collection instrument:** Dietary information was collected with an adapted version of the semi-quantitative food-frequency questionnaire found in the Procedure Handbook for Nutrition Projects, published by the National Institute of Public Health (INSP). The questionnaire included 101 foods classified into 14 groups. For each food, number of days of intake per week, times a day, portion size, and number of portions consumed within the seven days before the date of interview were asked.

The food table included in the questionnaire, as well as the portion size (very small, small, medium, large, and very large) and mean weight of portion, for each population group, were estimated by a group of INSP researchers from the analysis of the most-consumed-foods data obtained from the National Nutrition Survey 1999. It is worth mentioning that portion sizes were different for each age group studied.

**Personnel training:** The food-frequency questionnaire was administered by personnel trained and standardized on collection and input of data. Details on the data collection method are available in the above-mentioned procedure handbook. Data were entered into HP Compaq nx 6120 and Dell Latitude D510 laptop computers. Personnel from the INSP developed and validated the electronic questionnaires for input of data into the Fox Pro program version 7. Input was backed up and data were sent daily to the INSP.

**Quantity of food consumed:** Per-week and per-day quantities of each food consumed was obtained from the number of days per-week, times a day, portion size (weight) and number of portions.

Cleaning of the quantity of each food consumed by setting the maximum values that corresponded to a plausible intake for each age, sex and physiological status group was carried out. Food-frequency tables for every food analyzed were generated to establish plausible intake of each food in each population group (pre-school children, school children, adolescents, adults,...

---

Rodríguez-Ramírez S et al.
and older adults), with the goal of observing maximum and minimum number of portions consumed per-week of each food included in the questionnaire. Maximum values considered non-biologically plausible were excluded, although exclusion should not go beyond 1% of accumulated percentage. No minimum value was excluded and all quantities below the maximum cutoff points for each food were considered as valid.

Cleaning of maximum and minimum values of corn flour and wheat flour tortillas: Because tortilla is a staple for the Mexican population, an additional cleaning procedure was performed considering the weight of both corn and wheat tortillas, with a valid interval of 10 g minimum value and 500 g maximum value. Those values came from the mean range observed on tortilla weight reported by the same population. On the basis of that range, mean weight of both corn and wheat tortillas was calculated for each state. Mean weight was also imputed to those cases outside the valid interval.

Energy, macronutrient and micronutrient intakes: Total energy and nutrients contained in each food consumed and in total diet for each individual were estimated through the nutrient composition database compiled by the INSP.* Energy, carbohydrate, protein, fat, fiber, iron (total, heme, and nonheme), zinc, vitamin C, vitamin A, folate and calcium intakes were calculated. Because information was organized by week of intake, total quantities of energy and nutrients were divided by 7 to obtain an average intake per day. With respect to vitamin C, loss during cooking was considered, so the quantity of vitamin C was adjusted depending on type of food and cooking method commonly used.10-13

Some of the items on the food-frequency table were made up of several ingredients, so an average estimation of their nutritional value was performed. That value was calculated by adding the nutrients of each ingredient or food (as specified in the nutrient composition database) employed in Mexican dishes according to the proportions indicated in the recipes. Food was considered in raw condition. Although the estimation is imprecise, not having come from a bromatological analysis, it is a good approximation of the quantity of nutrients contained in some dishes, as it takes into account loss factors and gram conversion, such as densities, edible portion and cooked-raw/ raw-cooked factor.

Energy and nutrient adequacy percentage: Nutrient adequacy percentage for children, adolescents and adults was calculated from intake data using the reference values proposed by the United States Institute of Medicine (IOM). The estimation of protein, iron, zinc, vitamin C, retinol and folate adequacy was done using the Estimated Average Requirement.14-19 For calcium, Adequate Intake value was used as reference since the Estimated Average Requirement value has not yet been established because of lack of information for such calculations.20 For carbohydrates and fats, 50 and 30%, respectively, of the energy derived from those macronutrients were used as adequacy values. Reference values for the estimation of protein and micronutrient adequacy are shown in Table I.

Regarding older adults over 60 years of age, adequacy percentages were calculated in accordance with the guide of nutrition needs for older adults published by the WHO.21

Energy adequacy for each individual was calculated employing the Estimated Energy Requirements as the reference.22 Therefore, it was necessary to make use of several data included in the calculation equations; those data were weight, height, age, and physiological status of each individual. Energy requirements were calculated for those individuals with a body mass index (BMI) within the normal limits for their age and sex. For adults, BMI limits were ≥ 18.5 and < 25; for ≥ 2-years-old children and for adolescents those limits were based on data distribution, considering as minimum value above 1% of the distribution, and as maximum value the cutoff points proposed by Cole et al.23

For those individuals with a BMI outside the normality range, weight corresponding to the BMI limit value, considering the height of the individual, was imputed. For instance, in adults with BMI ≥ 25 the weight corresponding to a BMI of 24.9 was imputed. The same was done for the lower limit value, BMI < 18.5, by imputing the weight corresponding to a BMI of 18.5.

For those individuals without available height and weight data, median value of height and weight for the population with same age, sex, and physiological status, and with data availability, was imputed.

On the other hand, as it is well known, equations require a specific physical activity factor. There was no data on physical activity for this analysis; therefore, we used data from the 1999 National Nutrition Survey that showed that women led a sedentary life.24 Likewise, children were found to have low physical activity, in accordance with a Torun et al. study.25 Regarding men, they also showed low physical activity.

Cleaning of data on intake and percentages of energy and nutrient adequacy: As suggested in dietary analyses, intake and adequacy observations greater than 5 SDs

Table I
REFERENCE VALUES OF ESTIMATED AVERAGE REQUIREMENTS OF MICRONUTRIENTS AND PROTEINS FOR THE ESTIMATION OF DIET ADEQUACY IN MEXICAN POPULATION. MEXICO, ENSANUT 2006

<table>
<thead>
<tr>
<th>Group</th>
<th>Age (Years)</th>
<th>Protein (g)</th>
<th>Folate (μg)</th>
<th>Vit. C (mg)</th>
<th>Retinol (RE)* (μg)</th>
<th>Iron (mg)</th>
<th>Zinc (mg)</th>
<th>Calcium‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys and girls†</td>
<td>1 - 3</td>
<td>10.44</td>
<td>120</td>
<td>13.0</td>
<td>210</td>
<td>3.0</td>
<td>2.5</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>4 - 8</td>
<td>15.2</td>
<td>160</td>
<td>22.0</td>
<td>275</td>
<td>4.1</td>
<td>4.0</td>
<td>800</td>
</tr>
<tr>
<td>Boys†</td>
<td>9 - 13</td>
<td>27.36</td>
<td>250</td>
<td>39.0</td>
<td>445</td>
<td>5.9</td>
<td>7.0</td>
<td>1300</td>
</tr>
<tr>
<td></td>
<td>14 - 18</td>
<td>44.53</td>
<td>330</td>
<td>63.0</td>
<td>630</td>
<td>7.7</td>
<td>8.5</td>
<td>1300</td>
</tr>
<tr>
<td>Girls†</td>
<td>9 - 13</td>
<td>27.36</td>
<td>250</td>
<td>39.0</td>
<td>420</td>
<td>5.7</td>
<td>7.0</td>
<td>1300</td>
</tr>
<tr>
<td></td>
<td>14 - 18</td>
<td>38.34</td>
<td>330</td>
<td>56.0</td>
<td>485</td>
<td>7.9</td>
<td>7.3</td>
<td>1300</td>
</tr>
<tr>
<td>Women†</td>
<td>19 - 30</td>
<td>38.0</td>
<td>320</td>
<td>60.0</td>
<td>500</td>
<td>8.1</td>
<td>6.8</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>31 - 50</td>
<td>38.0</td>
<td>320</td>
<td>60.0</td>
<td>500</td>
<td>8.1</td>
<td>6.8</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>51 - 60</td>
<td>38.0</td>
<td>320</td>
<td>60.0</td>
<td>500</td>
<td>5.0</td>
<td>6.8</td>
<td>1200</td>
</tr>
<tr>
<td>Pregnant‡</td>
<td>14 - 18</td>
<td>50.1</td>
<td>520</td>
<td>66.0</td>
<td>530</td>
<td>23.0</td>
<td>10.0</td>
<td>1300</td>
</tr>
<tr>
<td></td>
<td>19 - 30</td>
<td>50.1</td>
<td>520</td>
<td>70.0</td>
<td>550</td>
<td>22.0</td>
<td>9.5</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>31 - 50</td>
<td>50.1</td>
<td>520</td>
<td>70.0</td>
<td>550</td>
<td>22.0</td>
<td>9.5</td>
<td>1000</td>
</tr>
<tr>
<td>Breastfeeding‡</td>
<td>14 - 18</td>
<td>59.9</td>
<td>450</td>
<td>96.0</td>
<td>885</td>
<td>7.0</td>
<td>10.9</td>
<td>1300</td>
</tr>
<tr>
<td></td>
<td>19 - 30</td>
<td>59.9</td>
<td>450</td>
<td>100.0</td>
<td>900</td>
<td>6.5</td>
<td>10.4</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>31 - 50</td>
<td>59.9</td>
<td>450</td>
<td>100.0</td>
<td>900</td>
<td>6.5</td>
<td>10.4</td>
<td>1000</td>
</tr>
<tr>
<td>Men†</td>
<td>19 - 30</td>
<td>47.0</td>
<td>320</td>
<td>75.0</td>
<td>625</td>
<td>6.0</td>
<td>9.4</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>31 - 60</td>
<td>47.0</td>
<td>320</td>
<td>75.0</td>
<td>625</td>
<td>6.0</td>
<td>9.4</td>
<td>1000</td>
</tr>
<tr>
<td>Older adults*</td>
<td>≥60 years</td>
<td>1 g/kg of weight by day</td>
<td>400</td>
<td>80.0</td>
<td>650</td>
<td>10.0</td>
<td>7.0 (men)</td>
<td>4.9 (women)</td>
</tr>
</tbody>
</table>

* Retinol equivalents
† Adequate intake, as data of estimated average requirements are unavailable
‡ Requirement proposed by the United States Institute of Medicine
§ As recommended by the World Health Organization

from the general distribution of energy, macronutrients, iron, zinc, vitamin C, vitamin A, foyltes and calcium were excluded from the analysis. Further on, data distribution was graphically analyzed, and those observations within 5 SDs but graphically separated from the conglomerate of the majority of data were eliminated as well. Concerning lower values, energy adequacy values less than 25% were eliminated for all age groups, as adequacy percentages less than that could not represent an intake compatible with life.

Due to the data elimination during the cleaning process, we calculated new expansion factors.

Because dietary data distribution is usually biased toward high values, data are presented as medians, 25 and 75 percentiles of the distribution. The process to obtain nutrient vectors from data was done in Access (Microsoft Office, 2003) using the Visual Basic code and SQL consults. Data cleaning was performed with Stata, version 9.2 (Stata Corporation, TX, USA) and SPSS, version 15.0 (SPSS Inc., Chicago, IL, USA).
Results

We analyzed dietary information of 39,998 individuals. Data lost by the cleaning process were between 3.3 and 11.9% among the different groups of population.

Pre-school children: Dietary data from 3,959 children less than 5 years old were obtained. Because 461 observations lacked data on weight and height, it was necessary to impute data from the same study population in accordance with age and sex so as to be able to calculate the percentage of energy adequacy.

After cleaning, 407 observations were excluded (10.3%). In 203 of those, there was low energy (less than 25%) or energy adequacy greater than 5 SDs from the adequacy mean. The other 204 observations had adequies above 5 SDs from the adequacy of several nutrients. Final sample size after cleaning of all the main variables (energy, macronutrients, total iron, zinc, vitamin C, retinol equivalents, folate and calcium) was 3,552 observations, which represented a total of 7,836,674 children aged 1 to 4 years countrywide.

School children: Dietary data were collected from 9,383 children between 5 and 11 years of age. In 396 observations it was necessary to impute height and weight of the distribution median according to age and sex since that data was not available.

Because 667 observations (7.1%) had adequacy percentages above 5 SDs for one or more of the nutrients studied, or an energy adequacy percentage less than 25%, they had to be excluded.

In total, 8,716 valid complete data were analyzed, representing 15,494,558 children aged 5 to 11 years (Table II).

Adolescents: Data from 8,768 adolescents aged 12 to 19 years were obtained. It was necessary to impute height and weight of the distribution median in 532 and 545 observations, respectively, in accordance with age and sex, as those observations lacked that information.

After cleaning, 326 observations (3.7%) were eliminated because of energy and/or nutrient adequacy percentages above 5SDs or energy adequacy percentage less than 25%. Therefore, 8,442 observations with valid dietary information were considered, which represented 18,276,531 adolescents at a national level.

Adults aged 20 to 59 years: Data from 16,494 adults were obtained, and 543 (3.3%) observations were eliminated as they presented intakes and adequacy percentages above 5 SDs or energy intakes less than 25%. Thus, valid information corresponded to 15,951 individuals aged 20 to 59 years, who represented a population of 47,946,764 nationwide.

Older adults: Data from 3,812 adults between 60 to 99 years old were collected, and 455 observations were eliminated (11.9%) owing to adequacies above 5 SDs from the energy and nutrient adequacy mean, or energy adequacies less than 25% of those recommended.

In total, valid information was obtained from 3,357 adults who represented 11,084,070 older adults aged 60 years and over nationwide.

Table II shows original sample size and total number of feasible data, by age group and with valid dietary information, obtained through the different cleaning stages.

Discussion

The contribution of this paper is that it shows in detail how the methodology for processing and analyzing dietary data from the ENSANUT 2006 in Mexico was established, from its planning to the presentation of results. This methodology has been used in the presentation of descriptive dietary information of different groups from the population in the ENSANUT 2006.26-29 This paper can be useful to standardize data cleaning in different population groups and in different studies, making it possible to compare results and to handle data easily, as well as to establish criteria and to validate...
steps for data processing. This information is also very valuable as it comes from a nationally representative survey, which permits the extrapolation of results to all the population.

This methodology is similar to the one employed in other national nutrition surveys. This is the case of the National Nutrition Survey 2005 of Colombia and the NNS-99 in Mexico where, for several nutrients, the reference values proposed by the United States Institute of Medicine were used to estimate adequacy percentage and population at risk of dietary inadequacy. Also, weight limits and physical activity assumption were used to calculate energy adequacy. Unlike ENSANUT 2006, in Colombia 24-hours recall method was employed to obtain the dietary information.30

With regard to the cleaning process, this is similar to the process used in the NNS-99.6,7 We found that the dietary data quality is good, since we had low percentage of losses in the cleaning process (3.3 to 11.9%); the highest losses were for pre-school children (10.3%) and older adults (11.9%), possibly due to bias in the reporting of dietary information.

We consider that the methodology employed is valid since it has been used in other studies with optimal/positive results.

When developing the methodology for processing and analyzing dietary data, researchers must warrant the validity of results since its use is of great importance for evaluating population nutrition status as well as for further planning and evaluation of programs and policies related to food production and marketing, nutrition education and food assistance and security.

Acknowledgements

We are grateful to the group of researchers from the Nutrition and Health Research Center of the National Institute of Public Health, particularly to Alfonso Jesús Mendoza Ramírez, Claudia Ivonne Ramírez Silva, Juan Espinosa Montero, Lucía Hernández Barrera, and Xóchitl Ponce Martínez, for their remarks, which helped to improve the methodology.

References