Methods of the National Nutrition Survey 1999

Elsa Resano-Pérez, MD,[1] Ignacio Méndez-Ramírez, PhD,[2] Teresa Shamah-Levy, BSc,[3]
Juan A Rivera, MS, PhD,[3] Jaime Sepúlveda-Amor, MD, DSc.[3]

Resano-Pérez E, Méndez-Ramírez I, Shamah-Levy T, Rivera JA, Sepúlveda-Amor J.
Methods of the National Nutrition Survey 1999.
Salud Publica Mex 2003;45 suppl 4:S558-S564.
The English version of this paper is available too at:
http://www.insp.mx/salud/index.html

Abstract

Objective. To describe the methods and analyses of the 1999 National Nutrition Survey (NNS-99). **Material and Methods**. The 1999 National Nutrition Survey (NNS-99) is a probabilistic survey with nationwide representativity. The NNS-99 included four regions and urban and rural areas of Mexico. The last sampling units were households, selected through stratified cluster sampling. The study population consisted of children under five years of age, school-age children (6-11 years), and women of childbearing age (12-49 years). Data were collected on anthropometric measurements, hemoglobin levels, morbidity and its determinants, and socioeconomic and demographic characteristics. In addition, data on diet and micronutrients intakes (iron, zinc, vitamin A, folic acid, vitamin C, and iodine) were obtained in a sub-sample of subjects. **Results**. The response rate for the NNS-99 was 82.3%; the non-response rate was 5.9% and the remaining did not participate due to uninhabited houses. **Conclusions**. This survey updates the information on nutritional status in Mexico and should serve as the basis for food and nutrition policy-making and priority program design. The English version of this paper is available too at: http://www.insp.mx/salud/index.html

Key words: national surveys; nutrition; children; women; anthropometry; anemia; micronutrients; diet; Mexico

Resano-Pérez E, Méndez-Ramírez I, Shamah-Levy T, Rivera JA, Sepúlveda-Amor J.
Metodología de la Encuesta Nacional de Nutrición 1999.
Salud Publica Mex 2003;45 supl 4:S558-S564.
El texto completo en inglés de este artículo también está disponible en: http://www.insp.mx/salud/index.html

Resumen

Objetivo. Describir la metodología y análisis de la Encuesta Nacional de Nutrición 1999 (ENN-99). **Material y métodos**. La ENN-99 es probabilística, con representatividad nacional de cuatro regiones y estratos urbano/rural. Las unidades últimas de muestras son los hogares seleccionados por muestreo estratificado y por conglomerados. La población de estudio fueron los niños menores de cinco años de edad, los niños 6-11 años (escolares) y las mujeres de 12-49 años, en quienes se obtuvieron mediciones antropométricas, niveles de hemoglobina, información sobre morbilidad y sus determinantes, datos socieconómicos y demográficos, y en una submuestra se determinaron micronutrimentos (hierro, zinc, vitamina A, ácido fólico, vitamina C, yoduria) y se obtuvo información sobre dieta. **Resultados**. La Encuesta obtuvo una tasa de respuesta de 82.3%, la no respuesta se asoció en 5.9% a no respuesta del informante y el resto fue ocasionado por no estar habitadas las viviendas seleccionadas. **Conclusiones**. La información obtenida a través de esta encuesta actualiza la información existente en nuestro país y es la base para la formulación de políticas y programas prioritarios. El texto completo en inglés de este artículo también está disponible en: http://www.insp.mx/salud/index.html

Palabras clave: encuestas nacionales; nutrición; niños; mujeres; antropometría; anemia; micronutrimentos; dieta; México

(1) Secretaría Académica, Instituto Nacional de Salud Pública. Cuernavaca, Morelos, México.
(2) Instituto de Investigaciones en Matemáticas Aplicadas, Universidad Nacional Autónoma de México. México, DF, México.
(3) Centro de Investigación en Nutrición y Salud, Instituto Nacional de Salud Pública. Cuernavaca, Morelos, México.

Received on: August 20, 2002 • **Acepted on:** July 4, 2003
Address reprint requests to: Maestra Elsa Resano Pérez. Avenida Universidad 655, colonia Santa María Ahuacatitlán 62508. Cuernavaca, Morelos, México.
E-mail: eresano@correo.insp.mx

The first national probabilistic nutrition and diet survey in Mexico was carried out in 1988. The National Nutrition Survey 1988[1] (NNS-1988) described the nutrition situation for the first time, both nationwide and by region (the country was divided into four regions for that purpose). Surveys previous to 1988 and other recent surveys had excluded either urban or rural zones and had not included representative samples, thus limiting their scope. Unlike those surveys, the NNS-1988 included both urban and rural zones. Also, its probabilistic sampling design made it representative at the national and regional level.

During the NNS-1988, data were collected from children <5 years of age, and women of reproductive age, the groups considered to be the most nutritionally vulnerable. Reliability of anthropometric measures was assured through collection by well-trained and standardized personnel. Height and weight measurements, together with age data, could be used for constructing indicators to differentiate between types of malnutrition.

The NNS-1988 collected data from over 13 000 households, including 19 000 women aged 12-49 years and almost 7 500 children <5 years of age. Data were collected from women and children on anthropometric measures, dietary intake, sociodemographic characteristics, and health and illness indicators. Data on anemia were collected from women only.

A second NNS was carried out 10 years after the first one, to account for demographic and socioeconomic changes that were likely to be influencing the prevalence and distribution of malnutrition due to deficiency of nutrients and overweight. It was necessary to update the information on the nutritional status of the population, given the high prevalence of malnutrition and overweight found in the previous survey. This paper presents the methods of the 1999 National Nutrition Survey.

Material and Methods

The study population consisted of Mexican subjects residing in their households at the time of the study, of the following age groups: children <5 years of age, school-age children (5-11 years), and women aged 12-49 years. The survey was conducted between October 1998 to March 1999. It included the entire country, aggregated by localities with < 2 500 inhabitants,* <15 000 inhabitants, and 15 000 or more inhabitants, as well as into four regions.*

The survey included diverse data collection strategies for the different age groups. Table I describes collection strategies, surveyed population, and type of data collected. Three different samples are readily apparent in Table I. The original sample included the entire study, with the following data: household characteristics, morbidity, anthropometry, capillary blood sample, and breastfeeding and complementary feeding of children <2 years.

A sub-sample was characterized by application of a food questionnaire and a women's questionnaire. A sub-sample of the sub-sample selected for dietary assessment was utilized to obtain a venous blood sample, a urine sample, and table salt intake.

Sample design

The probabilistic nature of the sample allowed making inferences to the target population and ensured the quality of estimations, mainly presented as rates, means, and proportions.

The sampling frame corresponded to the National Household Sampling Frame of INEGI.[2] This sampling frame was constructed by stratifying within each Mexican state to demarcate zone grouping localities.‡

Construction of sampling units

Primary sampling units. The primary sampling units (PSU) consisted of one or part of one Basic Geographical Statistical Area (BGSA). This is described as the geographical reference area for statistical information, in localities with 2 500 or more inhabitants it is made up of a number of household blocks delimited by streets and avenues (urban BGSA). Localities with <2 500 inhabitants form a rural BGSA within a geograph-

* Instituto Nacional de Estadística Geografía e Informática establishes an operational definition of rural localities as those < 2 500 inhabitants and of urban localities as those > 2 500 inhabitants, according to the last National Household Sampling Frame of INEGI.

* North: Baja California, Baja California Sur, Coahuila, Chihuahua, Durango, Nuevo León, Sonora, Tamaulipas; Center: Aguascalientes, Colima, Guanajuato, Jalisco, México (excluding municipalities and localities adjoining México City), Michoacán, Morelos, Nayarit, Querétaro, San Luis Potosí, Sinaloa, Zacatecas; South: Campeche, Chiapas, Guerrero, Hidalgo, Oaxaca, Puebla, Quintana Roo, Tabasco, Tlaxcala, Veracruz, Yucatán and México City (México City and adjoining urban municipalities of the State of México).

‡ Zone I: Cities and metropolitan areas selected for the National Survey of Urban Employment (ENEU); Zone II: Other cities of 100 000 and more inhabitants or state capitals; Zone III: Localities with 20 000 to 99 999 inhabitants; Zone IV: Localities with 15 000 to 19 999 inhabitants Zone V: Localities with 2 500 to 14 999 inhabitants; Zone VI: Localities with <2 500 inhabitants.

Table I
DATA COLLECTION STRATEGIES FOR THE DIFFERENT AGE GROUPS. MEXICO, 1998-1999

Strategy	*Surveyed population*	*Data collected*
Questionnaire applied at the household	All household members	Socioeconomic and demographic family characteristics
Questionnaire on morbidity	Children <5 years of age, school-age children (5-11 years), and women aged 12-49 years	Acute and chronic morbidity, its determinant factors and health-seeking behaviors
Anthropometric measurements	Children <12 years of age and women aged 12-49 years	Weight and height (in women, waist and hip circumferences)
Capillary blood samples	Children <12 years of age and women aged 12-49 years	Hemoglobin concentration
Questionnaire on children's feeding	Children <2 years of age	Complementary feeding and breastfeeding
Questionnaire on dietary intake 1 day before the visit	Sub-samples* (1/5) of 1- to 4- year-olds, 5- to 11-year-olds and 12- to 49-year-old women	Food, energy, macronutrient and micronutrient intake
Questionnaire on food intake frequency	Sub-sample* (1/5) of women aged 12-49 years	Food intake frequency
Women's questionnaire	Sub-sample* (1/5) of women aged 12-49 years	Obstetric history, physical activity, and use of tobacco and alcohol
Venous blood and urine samples	Sub-samples* (1/10) of 1- to 4- year-olds, 5- to 11-year-olds and 12- to 49-year-old-women	Assessment of micronutrient status

* Sub-samples of 1- to 4-year-old-children and 5- to 11-year-old-children included all children of selected households, while the sub-sample of 12- to 49-year-old-women included only one woman per household

ical area of about 10 000 hectares (BGSA rural) or of several adjoining BGSAs when the number of households was less than the required number to form a PSU. Depending on the zone, they were conformed as follows: zone I PSU: A BGSA with at least 480 households –the combination of two or more adjoining BGSA from the same stratum, with at least 480 households–; zones II to VI PSU: A BGSA or the combination of two or more BGSA with at least 280 households in localities with 2 500 or more inhabitants, or 100 households in localities with <2 500 inhabitants.

Secondary sampling units. The secondary sampling unit (SSU) in zone I was composed of one or more complete and adjacent blocks with at least 40 inhabited households. In zones II through VI, the SSU consisted of private households.

Tertiary sampling units. The tertiary sampling units (TSU) were found only in zone I and consisted of private households.

Stratification

Once constructed, the PSUs were stratified by an index of socioeconomic status for each locality and zone using the following variables from the Population and Household Census of 1995: percent of literate population aged 6 to 14 years, percent of literate population aged 15 years, percent of households with public sewage system, percent of households with potable water. For zone VI, the variable "percent of households with electricity" was added. The index was the first principal component obtained with the variables for each zone and locality.

Sample size calculation

The sample size should be sufficient to ensure that estimations are of good statistical quality. In a multiple-purpose survey, the calculation of an adequate sample size for each single purpose may be quite complex due to the number of variables and relationships of interest. Multiple analyses should be approached by selecting the most important variables that have the lowest values, to obtain the minimum sample size required when the event of interest is uncommon in the total population, thus including by default other more common events. The sample size was calculated using the formula:[3]

$$n = \frac{k^2}{r^2} \cdot \frac{Q}{P} \cdot \frac{DEFF}{(1\text{-TNR})\ PHV}$$

where:

n= sample size in number of households;
P= proportion of event of interest;
Q= 1 – P;
r= maximum acceptable relative error;
k= preset value for appropriate statistical estimation;
DEFF= design effect from cluster sampling;
TNR= non-response rate, and
PHV= average number of inhabitants per household.
Substituting the expression yields:
p= .0575;
Q= 0.9425;
r= 0.165;
k= 1.645 (value of tables with a 90% confidence);
DEFF= 2.3 (from an average of DEFF of above variables from NNS-1988);
TNR= 0.15, and
PHV= 0.21 children <2 years of age per household.

The sample size was n=2 000 households. The proportions utilized as reference to calculate the sample size came from the National Nutrition Survey 1988 (NNS-1988) and are presented in Table II. A basic variable to calculate the sample size was the proportion of children <2 years of age starting milk formula feeding in the first 4 to 5 months of life, since this is an uncommon event representing the smallest study group.

Table III shows the sample distribution by region and locality. Finally, the sub-sample for diet was applied in 4 200 households for each of the three intended populations, and the sub-sample for venous blood collection in 2 000 households for those three populations.

A multistage, stratified, cluster sampling selection was conducted. In each household, data were collected from all children <5 years of age, all school-age children (5-11 years), but only for one woman aged 12-49 years.

Households were selected independently in each state/zone. This process varied by zone.

Zone I. In the first stage n_{gh} PSU were selected with probability proportional to size (number of households in the PSU in the sampling frame). In the second stage k SSU were selected with probability proportional to size (number of households in the SSU in the sampling frame) in each PSU selected in the first stage.

In each SSU, five households were selected without replacement with equal probability. Thus, the probability of selecting a household was given by:

Table II
PROPORTIONS UTILIZED AS REFERENCE TO CALCULATE THE SAMPLE SIZE AS OBTAINED FROM THE NNS-1998. MEXICO, 1998-1999

Population and type of indicator	*Proportion of interest*
Population <2 years of age	
Perinatal pathologic conditions	0.1530
Absence of breastfeeding	0.1160
Artificial breastfeeding	
In the first 4-5 months	0.0575
Breastfeeding 4-5 months	0.0430
Population <5 years of age	
Diarrhea and depressed fontanel	0.0088
Diarrhea and used antidiarrheal medication	0.0489
Diarrhea, the family had no knowledge about oral rehydration therapy	0.0611
Other acute disease	0.1640
Women aged 12-49 years	
Attended ISSSTE for a medical visit	0.0373
Presence of diabetes	0.0130
Pregnant with 31-36 months of time elapsed between deliveries	0.0074
Receives nutritional supplements from ISSSTE	0.0060

ISSSTE: Social Security Institute of State Employees

Table III
SAMPLE DISTRIBUTION BY REGION AND LOCALITY. MEXICO, 1998-1999

Region	*Sample size in households*	*Locality sample*	*Sample size in households*
North	6 200	15 000 or more inhabitants	10 160
Center	6 200	<15 000 inhabitants	10 840
South	6 200	National	21 000
Mexico City	2 400		
National	21 000		

$$P(V_{ghij}) = \frac{n_{gh} m_{ghi}}{m_{gh}} \cdot \frac{k m_{ghij}}{m_{ghi}} \cdot \frac{5}{m^*_{ghij}} = \frac{5\ k\ n_{gh}\ m_{ghij}}{m_{gh}\ m^*_{ghij}}$$

Where:

n_{gh}= PSU sample size;
m_{ghi}= number of households of the i-th PSU in the h-th stratum of the g-th state;

m_{gh}= number of households in the h-th stratum of the g-th state;
m_{ghij}= number of households in the sampling frame in the j-th SSU of the i-th PSU in the h-th stratum of the g-th state;
k = SSU sample size, and
m^*_{ghij}= number of households listed in field work for the j-th SSU of the i-th PSU in the h-th stratum for the g-th state.

Zones II to V. In the first stage n_{gh} PSU were selected with probability proportional to size (number of households in the PSU in the sampling frame). In the second stage *t* households were selected without replacement with equal probability within each PSU selected in the first stage. Thus, the probability of household selection in the PSU_i is given by:

$$P(V_{ghi})= \frac{n_{gh}}{m_{gh}} \cdot \frac{m_{ghi} \cdot t}{m^*_{ghi}}$$

Where:

n_{gh}= PSU sample size;
m_{ghi}= number of households for the i-th PSU within the h-th stratum of the g-th state;
m_{gh}= number of households within the h-th stratum of the g-th state, and
m^*_{ghi}= total number of households listed in field work for the i-th PSU within the h-th stratum of the g-th state.

Zone VI. In the first stage n_{gh} PSU were selected with probability proportional to size (number of households in the PSU in the sampling frame). In the second stage, there were two or four segments of 10 households on average in each PSU. Therefore, the probability of selecting a household in the PSU_i of the h stratum in the g state was given by:

$$P(V_{ghi})= \frac{n_{gh}}{m_{gh}} \cdot \frac{m_{ghi} \cdot n_{seg}}{N_{seg}}$$

Where:

n_{gh}= PSU sample size;
m_{ghi}= number of households for the i-th PSU within the h-th stratum of the g-th state;
m_{gh}= number of households in the h-th stratum of the g-th state;
n_{seg}= sample size of segments, and
N_{seg}= total number of segments constructed in the PSU.

Selection of the sub-sample for dietary data. The sub-sample for diet was selected independently for each study population; one of every five of the 21 000 households was selected. In this way, the probability of selecting a household where dietary data were collected for each of the three target populations was given by:

$$PD(V_{ghij})= P(V_{ghij}) \cdot \frac{1}{5}$$

Where:

$PD(V_{ghij})$= probability of selecting a household for the sub-sample on diet, and
$P(V_{ghij})$= probability of selecting a household within the j-th SSU within the i-th PSU in the h stratum of the g state.

For the subgroup of women only one woman was selected from those living in a given household; thus, the probability of selecting a woman is given by:

$$PM= PD(V_{ghij}) \cdot \frac{1}{NMV}$$

Where:

PM= probability of selecting a woman, and
NMV= number of women in the household.

Selection of the sub-sample for venous blood data. For each target population, one of every three households was selected from those selected for the diet sub-sample. Venous blood specimens were collected from the target population.

The probability of selecting the household for the venous blood specimen collection was:

$$PS(V_{ghij})= PD(V_{ghij}) \cdot \frac{1}{3}$$

Where:

$PS(V_{ghij})$= probability of selecting a household for venous blood collection.

Estimates

The estimate of the total number of characteristic X is given by:

$$\hat{X}=\sum_g \sum_h \sum_i \left(F_{ghi} \sum_k X_{ghik} \right)$$

Where:

Table IV
FITTING OF SAMPLE BY REGION. MEXICO, 1998-1999

State	Sample fitting in households			
	<2 500 inhabitants	*2 500 to 14 999 inhabitants*	*15 000 or more inhabitants*	*Total*
North				
Baja California	160	160	350	670
Baja California Sur	80	40	50	170
Coahuila	240	120	360	720
Chihuahua	480	160	410	1 050
Durango	480	200	130	810
Nuevo León	240	160	630	1 030
Sonora	360	200	290	850
Tamaulipas	360	160	380	900
Total for region	2 400	1 200	2 600	6 200
Center				
Aguascalientes	40	40	70	150
Colima	40	40	40	120
Guanajuato	400	80	290	770
Jalisco	280	200	480	960
México*	440	320	990	1 750
Michoacán	360	200	190	750
Morelos	40	80	100	220
Nayarit	80	40	50	170
Querétaro	120	40	70	230
San Luis Potosí	240	40	120	400
Sinaloa	200	80	150	430
Zacatecas	160	40	50	250
Total for region	2 400	1 200	2 600	6 200
South				
Campeche	40	40	80	160
Chiapas	400	120	230	750
Guerrero	240	80	270	590
Hidalgo	200	80	150	430
Oaxaca	360	160	170	690
Puebla	320	240	480	1 040
Quintana Roo	40	40	120	200
Tabasco	160	40	140	340
Tlaxcala	40	80	80	200
Veracruz	560	240	660	1 460
Yucatán	40	80	220	340
Total for region	2 400	1 200	2 600	6 200
Mexico City		0	2 360	2 400
Federal District	40			
Total for region	40	0	2 360	2 400
Total number of households	7 240	3 600	10 160	21 000

* Excluding municipalities and localities merged with urban areas of Mexico City

F_{ghi}= expansion factor of the household within the i-th PSU of the h stratum in the g state (inverse of the probability of selecting a household), and

X_{ghik}= value of the characteristic of interest in the k-th interview within the i-th PSU in the h stratum of the g state.

The estimation of proportions, rates, and means uses the combined rate estimator $\hat{R}$, given by:

$$\hat{R}=\frac{\hat{Y}}{\hat{X}}$$

Where $\hat{Y}$ is defined in a way similar to $\hat{X}$.

The mean squared error for $\hat{R}$ (its squared root is used as a standard error) was obtained with

$$MSE(\hat{R})=\sum_g\sum_h \frac{n_h}{n_h-1}\sum_i^{nh}(d_{ghi}-\bar{d}_{ghi})^2 \text{ where } d_{ghi}=\sum_i F_{hgi}\sum_k \frac{(Y_{ghik}-\hat{R}X_{ghik})}{\hat{X}}$$

The fitting of sample per region is shown in Table IV.

Data collection results

Table V shows that the response rate was 82% of households. The non-response in 4.46% of households was due to refusal or to the absence of a proper informant; 13.3% of households were uninhabited or just temporally inhabited. The response rate per region is also shown.

Data analysis. Data analysis included a socioeconomic index, modified after that proposed by Bronfman,[4] by weighting different predictors using the principal components method. Predictor variables were: household flooring material, potable water, and ownership of electrical home appliances (washing machine, refrigerator, television, radio, and stove). This index explained 51.6 % of the variance of the set of variables, with a single component scoring poverty (SES –2.28797 to 1.54518).

References

1. Sepúlveda-Amor J, Lezana MA, Tapia-Conyer R, Valdespino JL, Madrigal H, Kumate J. Estado nutricional de prescolares y mujeres en México: resultados de una encuesta probabilística nacional. Gac Med Mex 1990;126(3):207-225.

Table V
RESPONSE RATE IN THE NATIONAL NUTRITION SURVEY, MEXICO 1999

National Region	Planned no. of households	Households visited	Households visited according to interview results: Interview results in households: Response Total	Response %	Information-related non-response Total	Information-related non-response %	Uninhabited temporal use Total	Uninhabited temporal use %	Prob. of frame Total	Prob. of frame %	Households with respondent
Republic of Mexico	21 000	21 503	17 716	82.39	1 260	5.86	2 109	9.81	418	1.94	17 944
North	6 200	6 361	5 430	85.37	229	3.6	602	9.46	100	1.57	5 471
Center	6 200	6 381	5 194	81.4	429	6.72	652	10.22	106	1.66	5 212
Mexico City	2 400	2 411	1 829	75.86	285	11.82	245	10.16	52	2.16	1 930
South	6 200	6 350	5 263	82.89	317	4.99	610	9.6	160	2.52	5 331

* Difference with planned number of households resulting from the updating of the sampling frame during data collection

2. Instituto Nacional de Estadística, Geografía, e Informática. Resultados del Conteo de Población y Vivienda de 1995. México, DF: INEGI, diciembre 1995.
3. Cochran W. Sampling techniques. Third edition. Otawa, Canada: John Wiley & Sons, 1977.
4. Bronfman M, Guiscafré H, Castro V, Castro R, Gutiérrez G. La medición de la desigualdad: una estrategia metodológica, análisis de las características socioeconómicas de la muestra. Arch Invest Med 1988;19:351-360.