Variability in Nutrient Intakes among Pregnant Women in Indonesia: Implications for the Design of Epidemiological Studies Using the 24-h Recall Method

Viveka Persson,*†2 Anna Winkvist,† T. Ninuk S. Hartini,†** Ted Greiner,* Mohammad Hakimi‡ and Hans Stenlund†

*Section for International Maternal and Child Health, Department of Women’s and Children’s Health, Uppsala University, Uppsala, Sweden; †Epidemiology, Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden; **Nutrition Academy, Ministry of Health, Yogyakarta, Indonesia and ‡Community Health and Nutrition Research Laboratory, Faculty of Medicine, University of Gadjah Mada, Yogyakarta, Indonesia

ABSTRACT Few studies have assessed the reliability of dietary intake methods during pregnancy. Between 1996 and 1998, a longitudinal study of dietary intake during pregnancy was carried out among 451 women in Central Java, Indonesia. Six 24-h recalls were performed each trimester. We report here on intraindividual and interindividual variability in energy and nutrient intakes, as well as the reliability of the 24-h diet recall method. Implications of the use of different numbers of replicate days for estimating dietary intake and the relationships between dietary intake and health outcomes are also discussed. Intravariance-to-intervariance ratios were <1 for energy and carbohydrates and >1 for all other nutrients throughout pregnancy. Reliability analyses found good agreement (reliability coefficient >0.7) with three replicates for the macronutrients, but at least six replicates were needed for an agreement of ≥0.6 for the micronutrients. To estimate true individual average intake with a precision of ±20%, six replicate recalls were sufficient for energy, carbohydrates, vitamin A, iron and vitamin C. In conclusion, mean intake of several nutrients can be reliably measured with the 24-h recall method, using a limited number of days. The nutrient of interest, the primary objectives and method of analyses should all be taken into account when planning sample size and number of replicates. J. Nutr. 131: 325–330, 2001.

KEY WORDS: • diet recall • pregnancy • reliability • variance • Indonesia • humans

Much research and intervention in low-income countries aim at ensuring optimal nutritional status and health of the pregnant woman for her sake as well as that of the newborn. One factor of great importance in achieving this is an adequate dietary intake during pregnancy. To assess dietary intake during pregnancy, as well as to establish links between diet and maternal and child health, reliable estimates of energy and nutrient intakes are often needed. However, estimating dietary intake during pregnancy is challenging, because pregnancy is a period when women’s dietary intakes may vary significantly over time. Factors that can affect dietary habits during pregnancy include nutritional requirements (FAO 1988, FAO/WHO/UNU 1985), activity (Banerjee et al. 1971), appetite (Coons 1933) and self-selected diet (Dickens and Trethowan 1971). Any dietary intake measurement is specific to the stage of pregnancy.

Information on variation in nutrient intake (intraindividual and interindividual) is required to guide decisions on the number of replicate measurements needed and on sample sizes. If too few replicate measures are taken or the sample size is too small, the statistical precision of intakes can be jeopardized and measures of diet-health outcome associations, such as correlations and relative risk, may be attenuated (Freudenberg et al. 1989, Liu et al. 1978, Sempos et al. 1985, Walker and Blettner 1985). In contrast, taking too many replications or too large a sample wastes resources and distracts respondents without serving any purpose.

Unfortunately, data assessing the precision of dietary intake methods during pregnancy are scarce and most come from the Western World (Nelson et al. 1989, Osofsky 1975, Rush and Kristal 1982). To our knowledge, only one study has reported patterns of dietary variability in women from a developing country (Launer et al. 1991). Food intake in that study was directly weighed for three consecutive d/mo from mo 6–9 in pregnancy among 743 Indonesian women.

The primary goal of the present study was to describe the intraindividual (within) and interindividual (between) vari-
ability in energy and nutrient intake in each trimester of pregnancy among women in a developing country. Second, because the weighing method presents difficulties in large studies, we wanted to elucidate how reliably the more practical 24-h recall method could be used in a developing country. Finally, we used our estimates of variance to show the implications of using different number of days for estimating true average intake as well as relationships between dietary intake and health outcomes.

SUBJECTS AND METHODS

Study site and sampling design. The study was conducted in Purworejo District, Central Java, Indonesia, which consists of 16 subdistricts and 494 villages. The total population is 750,000. Since 1994, the Faculty of Medicine at Gadjah Mada University has been operating a Community Health and Nutrition Research Laboratory in this area to support the Ministry of Health of Indonesia in developing and implementing a community health and nutrition surveillance program.

A two-stage cluster sampling method was used to select a 10% sample of households representative of the district. The sampling frame for the first stage consisted of a 20% sample of the enumeration areas or "wilayah," developed by the Central Bureau of Statistics for the 1990 census. In the second stage, the same number (101) of households was systematically sampled from each wilayah (Wilopo and team Community Health and Nutrition Laboratory 1997). From 1994 to 1998, each household was visited every third month for data collection. In addition, to identify new pregnancies, households with women likely to become pregnant (i.e., married women of reproductive age) were visited monthly. The nutritional status of women of reproductive age in the area is described in detail elsewhere (Nurdiati et al. 1998, Winkvist et al. 2000).

Between April 1996 and October 1998, a cohort of 846 women in early pregnancy was recruited for a study on nutritional status during reproduction. Within this framework, dietary data were collected among a subsample of 493 women. The remaining 353 women were not included for the following reasons: refusal (n = 42); abortions, stillbirths or death (n = 32); being deaf, too shy or mentally ill (n = 23); migration (n = 21); or because of difficulties with the fieldwork during the economic crisis or because they were recruited before dietary assessments were started (n = 235). We also excluded women who did not have six complete 24-h recalls per trimester (n = 28, 43 and 67 in trimesters 1, 2 and 3, respectively). However, these women were only excluded from analyses for those trimesters where fewer than six recalls were completed. This caused the total study sample to drop from 493 to 451. Of these 451 women, 122, 406 and 356 women in trimesters 1, 2 and 3, respectively, were included in the dietary analysis, referred to as the study sample. Hence, 451 women with data on at least one trimester were included.

Dietary intake. In each trimester, six 24-h recalls were used to estimate the dietary intake of the individual women. This number was based on the results of Launer et al. (1991) on random variation (CVw). The Variance Components procedure in SPSS was used to calculate the absolute values, and eq. 1 shows how CV was calculated:

$$ CV_{w} = \frac{s}{n\bar{x}} $$

where \( s \) is the square root of the estimated intraindividual variance.

To evaluate another source of variation, i.e., the trend in energy intake between each measurement occasion of each trimester (expressed as change per day), we applied a multilevel modeling technique (MLwiN 1.02). The first level was measurement occasion, and the second level was each woman. The model is a random regression model (Goldstein 1995):

$$ Y_{ij} = \beta_{0ij} + \beta_{1}X_{ij} + \epsilon_{ij} $$

where \( \beta_{0} \) and \( \beta_{1} \) are assumed to vary over the study population, and \( X_{ij} \) represents time between last menstrual period and measurement occasion.

To address our second goal, the reliability of 24-h recall, we used the intraclass correlation coefficient (\( \alpha \) coefficient, \( r_{\alpha} \)) as a function of number of recalls. This is a measure derived by Cronbach et al. (1972) that indicates the degree of agreement among repeated measures of some variable, in the present case, dietary intake. The Reliability Analysis procedure in SPSS was used for this purpose.

For the third goal, values of CV\(_{w}\) were used to illustrate the required number of recalls per individual for the various nutrients (Willett 1990):

$$ n = (Z_{\alpha}CV_{w}/D)^{2} $$

where \( n \) is the number of replicate days required, and \( Z_{\alpha} \) is the normal deviate for the percentage of times a confidence interval should cover the "true" average mean intake of an individual (e.g., \( Z_{\alpha} = 1.96 \) for 95% confidence). Finally, \( D \) is half the length of the interval, as a percentage of the mean.

In addition, the values of variance (\( s_{\alpha}^{2} \) and \( s_{\epsilon}^{2} \)) were used to illustrate the error in a regression analysis designed to look at the relation between dietary intake and some hypothetical outcome, with observed dietary intake as an independent variable. The error is measured by the ratio of the observed-to-true slope coefficients (Beaton et al. 1979), using eq. 3:

$$ b_{0}/b_{1} = r_{\epsilon}^{2}/(r_{\alpha}^{2} + s_{\epsilon}^{2}) $$

where \( b_{0} \) is observed regression coefficient, \( b_{1} \) is true regression coefficient, \( r \) is number of replications per individual, \( s_{\alpha}^{2} \) is estimated interindividual variance component and \( s_{\epsilon}^{2} \) is estimated intraindividual variance component. Independent sample t tests were used to

---

2 Abbreviations used: \( b_{0} \), observed regression coefficient; \( b_{1} \), true regression coefficient; \( CV_{w} \), coefficients of intraindividual (within) variation; MUAC, mid-upper arm circumference; \( Z_{\alpha} \), the normal deviate; \( r \), number of replications per individual; \( s_{\alpha}^{2} \), estimated interindividual variation component; \( s_{\epsilon}^{2} \), estimated intraindividual variation component; \( s_{\alpha} \), square root of the estimated intraindividual variation.
determine whether those excluded from dietary analyses (n = 395) were different from the study sample (n = 451) with respect to socioeconomic status and nutritional status.

**RESULTS**

**Sample characteristics.** The age of the study sample was 28.8 ± 5.4 y, and the mean parity was 1.6 ± 1.4. The mean height and mid-upper arm circumference (MUAC) at 2 mo of pregnancy were 150.0 ± 4.9 and 25.1 ± 2.9 cm, respectively. Most (90%) lived in the rural areas, and 46% had ≥7 y of schooling.

**Representativeness of the study sample.** The sample of 451 women from the pregnancy cohort included in the dietary intake analyses did not differ significantly from those excluded (n = 395) with respect to age, parity, height, MUAC at 2 mo of pregnancy, altitude of residence, education, type of toilet or water source (P > 0.05). However, a greater proportion of those included in the analyses lived in the urban areas (10% versus 4.5%, P < 0.01), and fewer worked with agriculture (38% versus 47%, P < 0.05).

The mean height and age of the total sample of women of reproductive age (n = 13,094) in Purworejo were 149.1 ± 5.1 cm and 30.4 ± 9.7 y, respectively, of which both were different from the study sample (P < 0.05). Forty-six percent of these women had ≥7 y of schooling, and 14% worked with agriculture, of which both were not different from the study sample. However, a greater proportion, 14% compared with 10%, lived in the urban areas (P < 0.01) (Nurdiati et al. 1998).

**Variance component analyses.** Intravariability and intervariability components for the nutrient intake of the Indonesian women are shown by trimester in Table 1. In all three trimesters, the ratio of intravariability to intervariability variation was <1.0 for energy and carbohydrates. Also in all trimesters, the ratios were greatest for the micronutrients (iron, vitamins A and C, thiamin and calcium). For most nutrients, the ratios were lowest in the first trimester.

Values of CV_w are shown in Table 2. For most nutrients, the CV_w was highest in the first trimester and lowest in the third. However, the differences in CV_w between the trimesters were not large. We also evaluated those women with 18 complete 24-h recalls (n = 84) to determine whether fluctuations in CV_w among these women were the same as the fluctuations in CV_w for the entire study sample. The differences between the two samples were always <10% in all three trimesters (data not shown).

Using random regression models, the slope (β_i) (SE), estimated during the first trimester, was 45.23 (7.82) (P < 0.001), indicating an average increase in intake of 45.23 kJ/d. For the second and third trimesters, the slope estimates were 12.62 (2.93) (P < 0.001) and 6.14 (3.13) (P = 0.05), respectively.

**Reliability of 24-h recalls.** Table 3 indicates the α coefficient as a function of the number of 24-h recalls in trimester 2. The reliabilities for two recalls are low for the micronutrients but respectable for the macronutrients. When six recalls are used, the α coefficients for all nutrients fall above 0.55.
Estimating true average intake. The number of 24-h recalls needed to estimate true average intakes of individuals is presented in Table 2. The number of replicates needed to estimate true average intake within an error of $\pm 10\%$ would be beyond the scope of most surveys for all nutrients. However, if an error range of $\pm 20\%$ is accepted, six replicates would be sufficient to estimate energy, carbohydrates, iron and vitamins A and C.

Dietary intake in regression analyses. Relationships between dietary intake and health outcomes are often evaluated with regression analyses. The ratio of the observed to true regression slope as a function of the number of replicates, with the different nutrients as the independent variable, is shown in Table 4. The results are based on the second trimester, but similar results were found for the other two trimesters. In contrast to $\text{CV}_w$, the ratio of the observed to true regression coefficient is a direct function of the ratio of intravariation to intervariation. The attenuation of the true regression coefficient with decreased replicates was greatest for the micronutrients. The regression coefficient for energy and carbohydrates did not change substantially by reducing the number of days from 6 to 4.

### TABLE 2

Number of 24-h recalls needed to estimate true average intake of pregnant women in Indonesia

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV$_w^1$ in trimester</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy</td>
<td>0.26</td>
<td>0.22</td>
<td>0.20</td>
</tr>
<tr>
<td>Fat</td>
<td>0.47</td>
<td>0.41</td>
<td>0.36</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>0.27</td>
<td>0.23</td>
<td>0.21</td>
</tr>
<tr>
<td>Protein</td>
<td>0.35</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>Iron$^4$</td>
<td>0.25</td>
<td>0.20</td>
<td>0.22</td>
</tr>
<tr>
<td>Vitamin A$^4$</td>
<td>0.24</td>
<td>0.23</td>
<td>0.21</td>
</tr>
<tr>
<td>Vitamin C$^4$</td>
<td>0.28</td>
<td>0.26</td>
<td>0.25</td>
</tr>
<tr>
<td>Calcium</td>
<td>0.54</td>
<td>0.51</td>
<td>0.50</td>
</tr>
<tr>
<td>Thiamin</td>
<td>0.49</td>
<td>0.40</td>
<td>0.60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>D$^2 = 10%$</th>
<th>D = 20%</th>
<th>D = 30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>15–28$^3$</td>
<td>4–7</td>
<td>2–3</td>
</tr>
<tr>
<td>Fat</td>
<td>50–85</td>
<td>12–21</td>
<td>6–9</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>17–28</td>
<td>4–7</td>
<td>2–3</td>
</tr>
<tr>
<td>Protein</td>
<td>37–47</td>
<td>9–12</td>
<td>4–5</td>
</tr>
<tr>
<td>Iron$^4$</td>
<td>15–24</td>
<td>4–6</td>
<td>2–3</td>
</tr>
<tr>
<td>Vitamin A$^4$</td>
<td>17–22</td>
<td>4–6</td>
<td>2</td>
</tr>
<tr>
<td>Vitamin C$^4$</td>
<td>24–30</td>
<td>6–8</td>
<td>3–4</td>
</tr>
<tr>
<td>Calcium</td>
<td>96–112</td>
<td>24–28</td>
<td>11–12</td>
</tr>
<tr>
<td>Thiamin</td>
<td>61–138</td>
<td>15–35</td>
<td>7–15</td>
</tr>
</tbody>
</table>

1 Coefficient of variation and calculated from $s_w$/mean intake (of that nutrient), where $s_w$ is the square root of the estimated intraindividual variance.
2 Half-length of the confidence interval as percentage of the mean.
3 Range indicates the number of days needed based on the CV$_w$ in trimester 1–3.
4 Due to skewed values, log-transformed data are shown.

### DISCUSSION

This is the first population-based study to report nutrient variance ratios for women in a developing country throughout pregnancy. It is also the first study to investigate the reliability of the 24-h recall method in pregnancy in a developing country.

A major finding was that the intravariance-to-intervariance ratios were $\leq 1$ for energy and carbohydrates in all three trimesters but were $>1$ for all other nutrients, except for protein in the first trimester. The largest variance ratios were found for the micronutrients. Hence, micronutrients are measured with greater error, meaning it will be more difficult to discover associations between micronutrient intake and certain outcomes than is the case for macronutrients. However, this study also showed that to estimate the mean intake of individuals with a precision of $\pm 20\%$ of the true intake, the current number of replicates (six) per individual would be sufficient for energy, carbohydrates, iron and vitamins A and C.

For most nutrients, a tendency toward larger $\text{CV}_w$ values in the first trimester was seen, perhaps because for the majority...
of nutrients, the intrapersonal variation was largest during the first trimester. This may in turn have been because many women experienced nausea at that time. In addition, the nutrient intake was smallest that trimester, as shown in Table 1.

In this study, we only accounted for intravariation and interviation in nutrient intake. We could not examine the extent to which the estimate of usual intake contained other errors, such as nutrient content of foods, bioavailability of nutrients, nutrient interactions or errors in recording. We do believe that the food composition data we used were more accurate and complete than those used in most studies from Indonesia, because our project included extensive efforts to obtain detailed information on ingredients and to conduct additional examinations on the composition of foods that have not been analyzed previously.

Intravariation has in earlier studies been shown to consist of several different components; these include method of data collection (Tarasuk and Beaton 1991), sequence of observation (Beaton et al. 1979, Hartman et al. 1990), day-of-the-week effect (Beaton et al. 1979, Hankin et al. 1967, Hartman et al. 1990, McGee et al. 1982) and seasonal differences (Hartman et al. 1990).

Intrapersonal variation may also vary with changes in appetite, physical activity and intake that occur during pregnancy. We found an average increase of 45.23, 12.62 and 6.14 kJ/d during each trimester, although individual variation around this mean increase was large. Thus, there was no indication of fatigue in repeated recalls. Also, for the other nutrients, no trend of declining intake was seen according to sequence of the recall (data not shown).

To minimize the days of the week effect, we included all 5 d in the Javanese calendar, accounting for market days when dietary habits may change. The variation due to interview occasion was partitioned from the total intravariance but was found to be <5% of the total intravariance. The effect of season was not investigated in this study, although the effect on macronutrient intake in this area is likely to be small. The economic crisis in Indonesia, which started approximately in October 1997, could also have contributed to variation in intake. However, we compared the variance ratios of women who completed their pregnancy before the crisis (~50%) with those after the crisis and found them to be quite similar.

To our knowledge, the only other study that has evaluated variance components of pregnant women from the developing world is that of Launer and colleagues (1991) from East Java, Indonesia. Their results are derived from 4 mo –9 in pregnancy among 743 women, whereas we present one estimate per trimester among 451 women. Launer and colleagues used a 3-consecutive-d food weighing method every month, whereas we used six nonconsecutive 24-h recalls within ~1 mo in each trimester. Although error may be greater (likely toward underreporting) in measuring individual intake with the 24-h recall method than with the direct weighing method (Block 1982), recall is cheaper and simpler, making it more suitable for larger studies. A recent study (Harrison et al. 2000) suggests that underreporting was less of a problem in one developing country, Egypt. It may also be easier with the 24-h recall method to assess nonconsecutive days, which gives a better estimate of the variance ratio (Block 1982, Tarasuk and Beaton 1991). The 24-h recall method is considered suitable for measuring change over time (Tarasuk and Beaton 1992) and does not require literate respondents.

Launer and coworkers (1991) suggest that "whether or not the ratio of intrapersonal to interindividual variation derived from recall data would differ from those reported here depends on the type of error." For example, if errors in the recall data cause only a systematic shift away from the true mean, this may not affect the variance components or their ratio. However, other types of error could inflate or deflate the ratios. Because the two studies not only used different methods but were conducted in different areas of Central Java 14 y apart, a comparison of results should be made with caution. Most likely, the socioeconomic situation has improved, resulting in the higher energy and fat intakes seen among our women. For protein and vitamin A, the variance ratios are comparable (ranging from 0.82 to 1.50 and 3.12 to 3.44, respectively, for the three trimesters in our study, compared with 1.28 and 3.8, in their study), as were the CV_w’s for energy (0.20 – 0.26, compared with 0.24 in their study). However, for energy and fat, our ratios were lower (0.57–0.78 and 1.30–1.50, respectively, compared with 1.35 and 3.24 in theirs). In contrast, the CV_w for protein and vitamin A in our study was slightly higher (0.31–0.35 and 0.21–0.24, respectively, compared with 0.25 and 0.19 in their study).

The studies available from Western countries on pregnant women are generally smaller (n = 60–225) and less complete than these two from Indonesia. Two studies used the 24-h recall method, either up to four or to four recalls taken during the second trimester (Rush and Kristal 1982) or up to four recalls before delivery without specifying time points (Ososfky 1975). A third one used weighed diet records (4 d at approximately monthly intervals from week 12–16) (Nelson et al. 1989). Their intravariance-to-intervariance ratios for energy (1.14–1.4), carbohydrates (1.18–1.2) (Nelson et al. 1989, Ososfky 1975, Rush and Kristal 1982) and vitamin A (4.9) (Nelson et al. 1989) were all higher than ours. For protein, they were either similar (1.38) (Rush and Kristal 1982) or higher (1.7–1.9) (Nelson et al. 1989, Ososfky 1975). For vitamin C (1.2) (Nelson et al. 1989) and calcium, they were lower (0.98–1.8) (Nelson et al. 1989, Rush and Kristal 1982).

The reasons behind the mostly lower variance component ratios seen in our study could be that in developing countries, diets tend to be more monotonous and what people eat is more closely linked to income. In both cases, this would increase interindividual variation in relation to intrapersonal variation.

Implications regarding the design of future studies. Our population-based study using the 24-h recall method has generated results similar to a population-based study conducted in the same country using the more accurate weighing method. This has several implications for future research.

First, it suggests that the usual mean intake of several nutrients among pregnant women can be reliably measured using the more practical 24-h recall method. The reliability analysis (Table 3) indicates that it is possible to obtain good agreement with two or three repeated recalls for the macronutrients. However, for vitamin A and vitamin C, six replications are not sufficient to obtain an "acceptable" value of ~0.7 for the α coefficient.

Second, our data on attenuation of the simple regression coefficients of the macronutrients suggest that it should be possible to use dietary intake data from 24-h recalls during pregnancy when evaluating the effect of diet on different outcomes of pregnancy. However, the regression coefficients of the micronutrients showed a much larger attenuation. Thus, for example, an association between the mother's vitamin A intake and her breast milk vitamin A content would be difficult to find even with six replicate 24-h recalls, because the observed regression coefficient would be attenuated almost 35%.

Third, in our analyses, we evaluated the effect of using
different numbers of replications on CV_w, regression coefficients and α coefficients. When the primary objective of a study is to detect differences between groups of individuals, the choice between increasing sample size or increasing number of replicates would depend on factors such as costs of repeated dietary sampling relative to the cost of recruiting additional subjects as well as availability of subjects. However, Freudenheim et al. (1989) showed that for weaker underlying associations, non-differential misclassification due to intra-individual variability induced bias in the odds ratio (toward unity) would persist, even with a larger sample size.

Fourth, our variance component ratios for macronutrients were generally lower than those reported for pregnant women in the industrialized world (Nelson et al. 1989, Ososky 1975, Rush and Kristal 1982). Thus, it may not be appropriate to generalize findings on intra-individual-to-interindividual variance derived from the Western world to populations in low-income countries.

Finally, when using the 24-h recall method, the nutrient of interest, the primary objectives of the study and the methods of analysis to be used should all be taken into account when planning the sample size and number of replicate measures.

LITERATURE CITED