Source of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Carbohydrate sources, vitamins, and minerals

George H Beaton, PhD, Jean Milner, BA, MSA, Valerie McGuire, BSc, RPDt, Thomas E Feather, BSc, and J Alick Little, MA, MD, FRCP(C)

ABSTRACT Beaton et al (Am J Clin Nutr 1979:32:2546-59) reported on the partitioning of variance in 1-day dietary data for the intake of energy, protein, total carbohydrate, total fat, classes of fatty acids, cholesterol, and alcohol. Using the same food intake data and the expanded National Heart, Lung and Blood Institute food composition data base, these analyses of sources of variance have been expanded to include classes of carbohydrate, vitamin A, vitamin C, thiamin, riboflavin, niacin, calcium, iron, total ash, caffeine, and crude fiber. The analyses relate to observed intakes (replicated six times) of 30 adult males and 30 adult females obtained under a paired Graeco-Latin square design with sequence of interview, interviewer, and day of the week as determinants. Neither sequence nor interviewer made consistent contribution to variance. In females, day of the week had a significant effect for several nutrients. The major partitioning of variance was between interindividual variation (between subjects) and intraindividual variation (within subjects) which included both true day-to-day variation in intake and methodological variation. For all except caffeine, the intraindividual variability of 1-day data was larger than the interindividual variability. For vitamin A, almost all of the variance was associated with day-to-day variability. One day data provide a very inadequate estimate of usual intake of individuals. In the design of nutrition studies it is critical that the intended use of dietary data be a major consideration in deciding on methodology. There is no "ideal" dietary method. There may be preferred methods for particular purposes.

KEY WORDS Dietary variability, intake variation, dietary methodology, estimation of nutrient intake

Introduction

In 1979 Beaton et al (1) published the results of an analysis of sources of variance in 24-h dietary data. This analysis encompassed total energy, total protein, total carbohydrate, total fat, saturated fatty acids, monounsaturated fatty acids, polyunsaturated fatty acids, cholesterol, and alcohol intake information. Variance was partitioned into subjects, sequence, interviewer, day of the week, and the residual component including both day to day variation in intake within subjects and methodological error. It was concluded that, except in females, who consumed more of all nutrients on the weekend day than on weekdays and who consumed less of all nutrients than did men, variance could be divided into two major components: interindividual variation and intraindividual variation (residual variation). The paper dis-
cussed some of the implications of the magnitude of the intraindividual variation for study design and for data interpretation. Since that time there has been continuing interest in both estimation of the magnitude of the components of variance (or of the reliability of 1-day data collection and required number of replications to obtain acceptable reliability of the estimated intake) and of the implications for data interpretation.

Recognizing this continuing interest and given the expansion of the National Heart, Lung and Blood Institute (NHLBI) food composition data bank, it was both feasible and desirable to reexamine the same food intake data base and to generate estimates of the partitioning of variance for additional nutrients of interest in nutritional studies. The present paper provides this additional information about the partitioning of variance. The implications of this information have been discussed elsewhere (1–14) and are presented only briefly here.

Methods

Details of the procedures have been presented elsewhere (1) and are summarized here. Thirty adult males and 30 adult females, age 25 to 44, were recruited from among sales and office staff of two Toronto department stores and one public utility. Each subject was interviewed about food consumed during the preceding day on six different occasions. Interviews were conducted by three trained professionals certified by the Lipid Research Clinic program (15). All data were collected between March and August 1977. For individual subjects the minimum time from first to last interview was 18 days; the maximum time was 99 days; the average time was 60 days. Interviews were conducted on Mondays, Wednesdays, and Thursdays or Fridays and related to food intake on Sundays, Tuesdays, and Wednesdays or Thursdays. Scheduling was arranged in accord with a Graeco-Latin square design such that each of the subjects was interviewed twice by each of the three interviewers and twice on each of the three interview days and such that interviewers and days were balanced for each subject and across all subjects. This design facilitated analysis of variance. Food intake records were coded and nutrient intakes were computed by the Nutrition Coding Center using a food composition data base generated for the Lipid Research Clinic and Multiple Risk Factor Intervention Trial programs (15). Food item coding was validated in Toronto (1). Since the time of the original analyses, the food composition data base had been expanded to include more nutrients. The original coded records were entered into the system again in 1981 to provide the nutrient intake data analyzed in the present paper.

As a first step, nutrient intakes were computed for each individual for each day. The computer was programmed to list also the number of food item entries for which a nutrient value was missing. For the nutrients included in the previously reported analyses (1), composition values were available for all food items reported in the intake records. For nutrients included in the present analyses, some data were missing: vitamin A, riboflavin, and niacin (1 entry each); vitamin C, thiamin (3 entries each); iron and calcium (11 and 12 entries); ash (26 entries); and caffeine (37 entries). For these items, no nutrient value was imputed: the level would be treated as 0. The total number of 1-day records affected was 13 of 360. In most cases the missing nutrient data related to specific types of candies, to liqueurs, and to a particular chocolate syrup. There were a few commercial foods for which a nutrient descriptor was missing. It was judged that these missing data would not seriously affect the analyses since the expected contribution of these food items was small. The data for caffeine may be more problematic: 25 of these entries related to candy, eight to liqueurs, and four to ice cream. The NHLBI food composition data base includes additional nutritional variables. These were not used since there were too many missing data. In the particular case of carbohydrate, the NHLBI data base listed “total carbohydrate,” “starch,” “sucrose,” and “other carbohydrate.” Since it is known that the partitioning of sucrose between “sucrose” and “other carbohydrate” was incorrect (if it was not known whether sucrose or other carbohydrate sweeteners were used in commercial products, the “other carbohydrate” entry was used) the two categories were combined as “other carbohydrate” in the present analyses.

The NHLBI nutrient data base had been adjusted to take into account the fatty acid profiles of Canadian foods where these differed from the usual US foods. This was not true with regard to nutrient levels in fortified foods. Thus, it must be recognized that there was an error in the computations to the degree that differences in Canadian and US food fortification policies would have led to a difference in intake. It was believed that this error would not detract seriously from the purpose of the present study. Use of vitamin and mineral supplements was not recorded; only nutrients in foods were computed.

The distribution of raw data (intakes per day) and of logarithmically transformed data were examined using the Statistical Analysis System test of the hypothesis that the data come from a normal distribution. Most distributions were found to have characteristics lying between those of the normal and log normal distribution. Most observed distributions were not significantly different

4 Appreciation is expressed to Dr Victor Grambsch and to staff at the Nutrition Coding Center, University of Minnesota, Minneapolis, for computing the nutrient vectors from the original dietary records using the NHLBI systems (15).

5 “Other carbohydrate” would include mono- and disaccharides and also some nonstarch polysaccharides. Since dietary fiber was not measured in the food composition data, part of the “other carbohydrate” would be dietary fiber. “Total carbohydrate,” calculated by difference, includes crude fiber and, of course, dietary fiber.
from a log normal distribution, while all but those for starch (females) and for iron (females) were significantly different from a normal distribution. The distributions for other carbohydrates (females), for vitamin A (both sexes), for vitamin C (both sexes), and for caffeine (both sexes) were significantly different from both the normal and log normal distributions; but only that for vitamin A demonstrated a marked departure. Given the magnitude of the data base of observations the test applied was seen as a very stringent one and departures, unless major, did not necessarily imply that an assumption of normality was unreasonable. The goodness of fit has importance not so much for the analysis of variance as for the interpretation of confidence limits based upon estimated standard deviation of intraindividual variation (1). It was decided to present the analyses with both untransformed and logarithmically transformed data.

Initial analysis of variance was conducted as described earlier (1). After establishing that interaction terms were not significant, the analysis was repeated with variance being attributed to subjects, to sequence of interview, to interviewer, to day of the week, and to the residual using the SAS Maximum Likelihood Variance Component Estimation Procedure. This was done for the 30 \( \times \) 6 observations for individuals of each sex. Tables 1 and 2, using original and logarithmically transformed data, were derived from this analysis. Statistical significance tests were based on the F ratio. Finally, the analysis of variance was repeated, using the same method, to partition variance into two components attributable to subject differences (termed interindividual variation) and residual (incorporating true day-to-day variation in intakes of individuals and methodological variations, and termed intraindividual variation (Table 3). From these variance estimates, the coefficients of variation (CV) presented later were calculated. This last step was performed only for the untransformed data.

Results

Table 1 presents the results of analysis of variance performed with untransformed data. In agreement with earlier analyses (1), it would appear that in these subjects there was little evidence to suggest a training effect of repeated interviews (note that these were not on adjacent days) nor was there evidence of an effect attributable to interviewer differences (note that these were trained and standardized interviewers). In males, there was no day of the week effect while in females, for a number of the nutrients, there was a systematic difference in intake across days. These effects achieved statistical significance for all nutrients except vitamin A, vitamin C, riboflavin, calcium, and crude fiber.

When intakes were expressed in relation to energy (eg, carbohydrate as percentage energy intake or thiamin intake per 1000 kcal), the day of the week effect was generally diminished (Table 1B). Interestingly, although no day of the week effect was evident for calcium intake per day, a significant effect became apparent for calcium intake/1000 kcal among women. These effects were somewhat reduced by logarithmic transformation of the data (Tables 2A and B). Mean intakes per day by day of the week and by sex are displayed in Table 3. Both calcium and caffeine behaved differently from most nutrients. In most cases, the mean daily intake of females increased on the weekend day in keeping with the rise in total energy intake. This accounted for the demonstration of a day of the week effect in total intake of the nutrient and disappearance of this effect when the intake is expressed in relation to energy. However, in the case of calcium, the intake failed to rise on the weekend, implying that the calcium/energy ratio must have fallen. In the case of caffeine, the total intake fell on the weekend; the ratio of caffeine to energy fell to an even greater degree.

Comparison of Tables 1 and 2 suggests that logarithmic conversion of the data had relatively little effect on the apparent partitioning of variance in males but in females tended to increase the proportion of variance attributed to subjects and to decrease the other components of variance. The most marked effects of the transformations were evident for vitamin A and for caffeine. For vitamin A in both males and females there was an appreciable increase in the amount of variance attributable to subjects with logarithmic transformation. In the case of caffeine, the effects of the transformation on partitioning of variance were opposite in males and females. In males, the proportion of variance attributed to subjects was higher with transformed data than with untransformed data. In females, the proportion of variance attribu-

In these analyses the 6 \( \times \) 30 observations for each sex were treated as if they were independent. For intakes of starch, other carbohydrate, vitamin C, thiamin, riboflavin, niacin, calcium, iron, and caffeine, the skewness values ranged between 0.60 and 2.39 for males and between 0.45 and 2.41 for females. Kurtosis values ranged between -0.41 and 11.08 for males and between 0.60 and 2.39 for females. Kurtosis values ranged between -0.41 and 11.08 for males and between 0.60 and 2.39 for females. Kurtosis values ranged between -0.41 and 11.08 for males and between 0.60 and 2.39 for females. In contrast, for vitamin A, the skewness values were 6.86 and 3.46 and the kurtosis values were 62.27 and 15.09 for males and females, respectively.
### TABLE 1A
Relative contributions of sources of variance; original data (%); A. Intakes per day

<table>
<thead>
<tr>
<th>Component of variance</th>
<th>Energy</th>
<th>Total CHO*</th>
<th>Starch</th>
<th>Other CHO*</th>
<th>Vitamin A</th>
<th>Vitamin C</th>
<th>Thiamin</th>
<th>Riboflavin</th>
<th>Niacin</th>
<th>Calcium</th>
<th>Iron</th>
<th>Total ash</th>
<th>Caffeine</th>
<th>Crude fiber</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjects</td>
<td>48.0</td>
<td>38.1</td>
<td>26.5</td>
<td>34.0</td>
<td>0</td>
<td>22.5</td>
<td>28.6</td>
<td>29.8</td>
<td>39.1</td>
<td>32.0</td>
<td>38.2</td>
<td>38.8</td>
<td>50.5</td>
<td>28.2</td>
</tr>
<tr>
<td>Sequence</td>
<td>1.96</td>
<td>0.60</td>
<td>1.32</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.58</td>
<td>1.53</td>
<td>0.14</td>
<td>0.23</td>
<td>0.23</td>
<td>0.44</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>Interviewer</td>
<td>0.50</td>
<td>0.95</td>
<td>0.85</td>
<td>0.09</td>
<td>0</td>
<td>0</td>
<td>0.44</td>
<td>0</td>
<td>0</td>
<td>1.14</td>
<td>0.17</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Day of wk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.81</td>
<td>0.27</td>
<td>0.53</td>
<td>0</td>
<td>0.18</td>
<td>0.21</td>
<td>0.21</td>
<td>4.81‡</td>
<td>4.81‡</td>
<td>1.19</td>
</tr>
<tr>
<td>Residual</td>
<td>49.6</td>
<td>60.4</td>
<td>72.2</td>
<td>65.2</td>
<td>99.9</td>
<td>76.6</td>
<td>70.6</td>
<td>67.7</td>
<td>60.8</td>
<td>66.3</td>
<td>61.7</td>
<td>57.4</td>
<td>44.7</td>
<td>70.4</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjects</td>
<td>42.1</td>
<td>42.1</td>
<td>40.6</td>
<td>37.0</td>
<td>4.74</td>
<td>34.2</td>
<td>20.6</td>
<td>31.2</td>
<td>21.4</td>
<td>52.6</td>
<td>30.8</td>
<td>42.8</td>
<td>58.5</td>
<td>29.1</td>
</tr>
<tr>
<td>Sequence</td>
<td>0</td>
<td>1.08</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.81</td>
<td>3.35†</td>
<td>0</td>
<td>1.06</td>
<td>0.14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Interviewer</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.63</td>
<td>0</td>
<td>0.92</td>
<td>0</td>
<td>0.59</td>
<td>1.05</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Day of wk</td>
<td>7.19‡</td>
<td>3.78§</td>
<td>3.61‡</td>
<td>1.80†</td>
<td>0.52</td>
<td>1.52</td>
<td>5.73‡</td>
<td>0.46</td>
<td>4.13§</td>
<td>8.11‡</td>
<td>3.60§</td>
<td>1.84†</td>
<td>1.84†</td>
<td>0</td>
</tr>
<tr>
<td>Residual</td>
<td>50.7</td>
<td>54.1</td>
<td>54.7</td>
<td>61.2</td>
<td>93.1</td>
<td>62.5</td>
<td>69.3</td>
<td>68.3</td>
<td>72.8</td>
<td>46.2</td>
<td>61.4</td>
<td>53.7</td>
<td>39.6</td>
<td>70.9</td>
</tr>
</tbody>
</table>

* Carbohydrate. † p ≤ 0.05. ‡ p ≤ 0.01.

### TABLE 1B
Relative Contribution of Sources of Variance. Original Data (%); B. Intake relative to energy or total carbohydrate intake

<table>
<thead>
<tr>
<th>Component of variance</th>
<th>Relative to energy</th>
<th>Relative to total CHO*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total CHO*</td>
<td>Starch</td>
</tr>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjects</td>
<td>29.6</td>
<td>31.8</td>
</tr>
<tr>
<td>Sequence</td>
<td>0</td>
<td>0.46</td>
</tr>
<tr>
<td>Interviewer</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Day of wk</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Residual</td>
<td>70.4</td>
<td>67.8</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjects</td>
<td>37.0</td>
<td>36.7</td>
</tr>
<tr>
<td>Sequence</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Interviewer</td>
<td>0.50</td>
<td>0</td>
</tr>
<tr>
<td>Day of wk</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Residual</td>
<td>62.5</td>
<td>63.3</td>
</tr>
</tbody>
</table>

* Carbohydrate. † p ≤ 0.05. ‡ p ≤ 0.01.
### TABLE 2A
Relative contributions of sources of variance, logarithmically transformed data (%); A. Intakes per day

<table>
<thead>
<tr>
<th>Component of variance</th>
<th>Energy</th>
<th>Total CHO*</th>
<th>Starch</th>
<th>Other CHO*</th>
<th>Vitamin A</th>
<th>Vitamin C</th>
<th>Thiamin</th>
<th>Riboflavin</th>
<th>Niacin</th>
<th>Calcium</th>
<th>Iron</th>
<th>Total ash</th>
<th>Caffeine</th>
<th>Crude fiber</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjects</td>
<td>44.2</td>
<td>35.2</td>
<td>24.4</td>
<td>35.8</td>
<td>17.2</td>
<td>19.0</td>
<td>30.1</td>
<td>34.7</td>
<td>31.0</td>
<td>27.6</td>
<td>31.4</td>
<td>36.0</td>
<td>82.8</td>
<td>32.9</td>
</tr>
<tr>
<td>Sequence</td>
<td>2.24†</td>
<td>0.12</td>
<td>0.73</td>
<td>0.10</td>
<td>0.64</td>
<td>0.21</td>
<td>0.65</td>
<td>0</td>
<td>2.68‡</td>
<td>0.11</td>
<td>3.39†</td>
<td>0.02</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Interviewer</td>
<td>0.89</td>
<td>1.54</td>
<td>0.185†</td>
<td>0.61</td>
<td>0.30</td>
<td>1.58</td>
<td>0.24</td>
<td>0.96</td>
<td>0.88</td>
<td>0</td>
<td>1.44</td>
<td>0.05</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>Day of wk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.63‡</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>52.7</td>
<td>63.2</td>
<td>74.9</td>
<td>62.4</td>
<td>81.8</td>
<td>79.7</td>
<td>69.4</td>
<td>63.1</td>
<td>68.7</td>
<td>68.8</td>
<td>67.7</td>
<td>59.2</td>
<td>15.5</td>
<td>65.3</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjects</td>
<td>56.7</td>
<td>59.1</td>
<td>44.1</td>
<td>60.0</td>
<td>26.8</td>
<td>42.5</td>
<td>33.3</td>
<td>43.2</td>
<td>22.5</td>
<td>53.0</td>
<td>38.0</td>
<td>50.3</td>
<td>44.6</td>
<td>34.7</td>
</tr>
<tr>
<td>Sequence</td>
<td>0.08</td>
<td>0.58</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.24</td>
<td>0</td>
<td>1.77</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Interviewer</td>
<td>0</td>
<td>0.64</td>
<td>0</td>
<td>0.88</td>
<td>0</td>
<td>0.34</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Day of wk</td>
<td>4.30‡</td>
<td>2.08‡</td>
<td>1.74‡</td>
<td>0.17</td>
<td>0</td>
<td>6.57‡</td>
<td>0.70</td>
<td>4.03‡</td>
<td>6.00‡</td>
<td>2.64‡</td>
<td>4.19†</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>38.9</td>
<td>38.8</td>
<td>53.0</td>
<td>39.9</td>
<td>72.3</td>
<td>57.5</td>
<td>57.9</td>
<td>56.1</td>
<td>70.4</td>
<td>46.6</td>
<td>56.0</td>
<td>47.0</td>
<td>51.2</td>
<td>65.3</td>
</tr>
</tbody>
</table>

* Carbohydrate. † p ≤ 0.05. ‡ p ≤ 0.001. § p ≤ 0.001.

### TABLE 2B
Relative contribution of sources of variance, logarithmically transformed data (%); B. Intake relative to energy or total carbohydrate intake

<table>
<thead>
<tr>
<th>Component of variance</th>
<th>Relative to energy</th>
<th>Relative to total CHO*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total CHO*</td>
<td>Starch</td>
</tr>
<tr>
<td><strong>Males</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjects</td>
<td>31.0</td>
<td>28.1</td>
</tr>
<tr>
<td>Sequence</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Interviewer</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Day of wk</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Residual</td>
<td>69.0</td>
<td>71.9</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjects</td>
<td>43.7</td>
<td>52.3</td>
</tr>
<tr>
<td>Sequence</td>
<td>0</td>
<td>0.04</td>
</tr>
<tr>
<td>Interviewer</td>
<td>0.40</td>
<td>0.34</td>
</tr>
<tr>
<td>Day of wk</td>
<td>0</td>
<td>0.40</td>
</tr>
<tr>
<td>Residual</td>
<td>55.9</td>
<td>46.9</td>
</tr>
</tbody>
</table>

* Carbohydrate. † p ≤ 0.05. ‡ p ≤ 0.01. § p ≤ 0.001.
## VARIANCE IN DIETARY DATA

**TABLE 3**

Effect of day of the wk on nutrient intake

<table>
<thead>
<tr>
<th>Nutrient vector</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sunday*</td>
<td>Tuesday</td>
</tr>
<tr>
<td>Total energy (kcal/day)</td>
<td>2656</td>
<td>2608</td>
</tr>
<tr>
<td>Total CHO† (g/day)</td>
<td>261.2</td>
<td>265.4</td>
</tr>
<tr>
<td>Starch (g/day)</td>
<td>113.8</td>
<td>119.7</td>
</tr>
<tr>
<td>Other CHO (g/day)</td>
<td>143.6</td>
<td>141.8</td>
</tr>
<tr>
<td>Vitamin A (IU/day)</td>
<td>5423</td>
<td>6515</td>
</tr>
<tr>
<td>Vitamin C (mg/day)</td>
<td>101.2</td>
<td>106.8</td>
</tr>
<tr>
<td>Thiamin (mg/day)</td>
<td>1.81</td>
<td>1.65</td>
</tr>
<tr>
<td>Riboflavin (mg/day)</td>
<td>2.24</td>
<td>2.37</td>
</tr>
<tr>
<td>Niacin (equivalents/day)</td>
<td>24.9</td>
<td>25.4</td>
</tr>
<tr>
<td>Calcium (mg/day)</td>
<td>861.2</td>
<td>908.7</td>
</tr>
<tr>
<td>Iron (mg/day)</td>
<td>15.6</td>
<td>15.3</td>
</tr>
<tr>
<td>Total ash (g/day)</td>
<td>17.7</td>
<td>17.3</td>
</tr>
<tr>
<td>Caffeine (mg/day)</td>
<td>456.4</td>
<td>433.8</td>
</tr>
<tr>
<td>Crude fiber (g/day)</td>
<td>3.53</td>
<td>3.59</td>
</tr>
</tbody>
</table>

* Day refers to day on which food was consumed. Sample size is 30 x 2 for each day and each sex.
† Carbohydrate

The variance in dietary data was due to subjects was reduced by transformation. Clearly, the analyses of variance must be interpreted cautiously for these nutrient vectors. For the other nutrients, the effects of transformation were not of so great a magnitude that the analyses are jeopardized.

Table 4 presents a summarized analysis of nontransformed data displaying only two components of variance, consistent with the display and discussion in the earlier publication (1). The variance is represented in Table 4 by the coefficient of variation, and the proportionality of intra- and interindividual variability is represented by the ratios of the coefficients of variation to be consistent with the earlier discussion and formulas developed at that time (1). To derive the more customary ratio of variances the ratios shown in Table 4 must be squared.

In adopting the convention of presenting the coefficient of variation there is an implicit assumption of a linear relationship between the mean and the SD. This was tested by fitting linear regressions of the means and SDs for the sets of six observations for each individual, separately for the two sexes. With untransformed data, the intercepts did not differ significantly from 0 except for starch as percentage energy for both sexes. The linear relationships were significant for all variables when expressed as intakes per day and, when expressed in relationship to energy intake, for all variables except total carbohydrates (females), vitamin A (females), and total ash (males). The values of R² for these regressions are shown in Table 5. It would appear that for most nutrients, but not all, there is a linear relationship between the mean and the standard deviation and that the coefficient of variation is an appropriate descriptor of variability.

### Discussion

In the earlier analysis (1) it was found that women consumed more total food energy on Sundays than on either of the weekdays. For most nutrient vectors, a day of the week effect was apparent for the women; but this effect disappeared when intakes were expressed per 1000 kcal. It was suggested that this implied that the quantity of food, rather than the nature of the mix of foods, was changing on the weekend. The observations of intake by day of the week (Table 3) imply a clear difference in eating patterns of the women on weekend and working days, not just in terms of the total quantity of food consumed but also in the selection of foods. To the degree that riboflavin behaved like calcium, it is tempting to suggest that the use of dairy products may have been fairly constant in these women and that this was a major determinant of both calcium and riboflavin intake. To the degree that caffeine intake was deter-
TABLE 4
Estimated coefficients of variation (CV) for total, interindividual, and intraindividual* variability in 1-day intake†: original data

<table>
<thead>
<tr>
<th>Nutrient vector</th>
<th>Males</th>
<th></th>
<th></th>
<th></th>
<th>Females</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>CV</td>
<td>Mean</td>
<td>CV</td>
<td></td>
<td>Mean</td>
<td>CV</td>
<td>Ratio of CV's (intra/inter)</td>
</tr>
<tr>
<td></td>
<td>Total (CV)</td>
<td>Inter (CV)</td>
<td>Intra (CV)</td>
<td>Ratio of CV's (intra/inter)</td>
<td>Total (CV)</td>
<td>Inter (CV)</td>
<td>Intra (CV)</td>
<td>Ratio of CV's (intra/inter)</td>
</tr>
<tr>
<td>Energy (kcal/day)</td>
<td>2637</td>
<td>35.5</td>
<td>24.5</td>
<td>25.7</td>
<td>1.1</td>
<td>1792</td>
<td>39.9</td>
<td>25.4</td>
</tr>
<tr>
<td>Total CHO (g/day)</td>
<td>264.4</td>
<td>37.4</td>
<td>23.0</td>
<td>29.5</td>
<td>1.3</td>
<td>180.6</td>
<td>46.6</td>
<td>29.9</td>
</tr>
<tr>
<td>Starch (g/day)</td>
<td>120.2</td>
<td>51.1</td>
<td>26.1</td>
<td>43.9</td>
<td>1.7</td>
<td>78.3</td>
<td>53.5</td>
<td>33.6</td>
</tr>
<tr>
<td>Other CHO (g/day)</td>
<td>140.5</td>
<td>43.2</td>
<td>25.1</td>
<td>35.2</td>
<td>1.4</td>
<td>98.9</td>
<td>56.8</td>
<td>35.4</td>
</tr>
<tr>
<td>Vitamin A (IU/day)</td>
<td>6369</td>
<td>146.6</td>
<td>0</td>
<td>146.6</td>
<td>1</td>
<td>5188</td>
<td>114.1</td>
<td>22.7</td>
</tr>
<tr>
<td>Vitamin C (mg/day)</td>
<td>111.5</td>
<td>76.0</td>
<td>35.9</td>
<td>67.0</td>
<td>1.9</td>
<td>105.9</td>
<td>79.9</td>
<td>46.2</td>
</tr>
<tr>
<td>Thiamin (mg/day)</td>
<td>1.75</td>
<td>51.2</td>
<td>27.3</td>
<td>43.4</td>
<td>1.6</td>
<td>1.29</td>
<td>58.1</td>
<td>25.0</td>
</tr>
<tr>
<td>Riboflavin (mg/day)</td>
<td>2.38</td>
<td>51.2</td>
<td>27.3</td>
<td>43.1</td>
<td>1.6</td>
<td>1.71</td>
<td>51.4</td>
<td>28.7</td>
</tr>
<tr>
<td>Niacin (equivalents/day)</td>
<td>25.3</td>
<td>46.0</td>
<td>28.8</td>
<td>35.9</td>
<td>1.3</td>
<td>16.7</td>
<td>40.2</td>
<td>17.9</td>
</tr>
<tr>
<td>Calcium (mg/day)</td>
<td>902.0</td>
<td>49.6</td>
<td>27.9</td>
<td>41.0</td>
<td>1.5</td>
<td>671.3</td>
<td>63.2</td>
<td>45.8</td>
</tr>
<tr>
<td>Iron (mg/day)</td>
<td>15.7</td>
<td>43.8</td>
<td>26.9</td>
<td>34.6</td>
<td>1.3</td>
<td>11.3</td>
<td>36.8</td>
<td>19.6</td>
</tr>
<tr>
<td>Total ash (g/day)</td>
<td>17.5</td>
<td>38.6</td>
<td>23.9</td>
<td>30.3</td>
<td>1.3</td>
<td>12.4</td>
<td>35.5</td>
<td>23.0</td>
</tr>
<tr>
<td>Caffeine (mg/day)</td>
<td>557.0</td>
<td>61.2</td>
<td>43.7</td>
<td>41.1</td>
<td>0.9</td>
<td>509.8</td>
<td>67.6</td>
<td>51.8</td>
</tr>
<tr>
<td>Crude fiber (g/day)</td>
<td>3.77</td>
<td>58.5</td>
<td>30.8</td>
<td>49.7</td>
<td>1.6</td>
<td>3.48</td>
<td>58.7</td>
<td>31.7</td>
</tr>
<tr>
<td>Starch (% total CHO)</td>
<td>45.3</td>
<td>28.7</td>
<td>13.6</td>
<td>25.2</td>
<td>1.9</td>
<td>43.6</td>
<td>36.3</td>
<td>24.9</td>
</tr>
<tr>
<td>Other CHO (% total CHO)</td>
<td>53.3</td>
<td>24.4</td>
<td>11.6</td>
<td>21.5</td>
<td>1.8</td>
<td>54.1</td>
<td>28.3</td>
<td>19.2</td>
</tr>
<tr>
<td>Total CHO (% energy)</td>
<td>40.6</td>
<td>20.5</td>
<td>11.2</td>
<td>17.2</td>
<td>1.5</td>
<td>40.2</td>
<td>27.1</td>
<td>16.5</td>
</tr>
<tr>
<td>Starch (% energy)</td>
<td>18.4</td>
<td>35.7</td>
<td>20.1</td>
<td>29.5</td>
<td>1.5</td>
<td>17.6</td>
<td>42.0</td>
<td>25.5</td>
</tr>
<tr>
<td>Other CHO (% energy)</td>
<td>21.7</td>
<td>33.6</td>
<td>15.0</td>
<td>30.1</td>
<td>2.0</td>
<td>21.7</td>
<td>41.1</td>
<td>25.2</td>
</tr>
<tr>
<td>Vitamin A (IU/1000 kcal)</td>
<td>2608</td>
<td>172.5</td>
<td>0</td>
<td>172.5</td>
<td>1</td>
<td>3506</td>
<td>199.6</td>
<td>28.5</td>
</tr>
<tr>
<td>Vitamin C (mg/1000 kcal)</td>
<td>42.4</td>
<td>71.5</td>
<td>31.7</td>
<td>64.1</td>
<td>2.0</td>
<td>63.4</td>
<td>85.5</td>
<td>47.4</td>
</tr>
<tr>
<td>Thiamin (mg/1000 kcal)</td>
<td>0.674</td>
<td>39.9</td>
<td>16.9</td>
<td>36.2</td>
<td>2.1</td>
<td>0.747</td>
<td>47.9</td>
<td>12.2</td>
</tr>
<tr>
<td>Riboflavin (mg/1000 kcal)</td>
<td>0.929</td>
<td>48.5</td>
<td>18.7</td>
<td>44.8</td>
<td>2.4</td>
<td>1.009</td>
<td>48.4</td>
<td>19.9</td>
</tr>
<tr>
<td>Niacin (equivalents/1000 kcal)</td>
<td>9.64</td>
<td>30.0</td>
<td>6.7</td>
<td>29.2</td>
<td>4.4</td>
<td>10.18</td>
<td>46.2</td>
<td>24.3</td>
</tr>
<tr>
<td>Calcium (mg/1000 kcal)</td>
<td>356.1</td>
<td>49.7</td>
<td>26.4</td>
<td>42.1</td>
<td>1.6</td>
<td>388.8</td>
<td>57.7</td>
<td>32.7</td>
</tr>
<tr>
<td>Iron (mg/1000 kcal)</td>
<td>5.99</td>
<td>27.6</td>
<td>12.7</td>
<td>24.5</td>
<td>1.9</td>
<td>6.72</td>
<td>34.2</td>
<td>18.2</td>
</tr>
<tr>
<td>Total ash (g/1000 kcal)</td>
<td>6.70</td>
<td>21.3</td>
<td>8.1</td>
<td>19.6</td>
<td>2.4</td>
<td>7.41</td>
<td>33.6</td>
<td>14.1</td>
</tr>
<tr>
<td>Caffeine (mg/1000 kcal)</td>
<td>236.0</td>
<td>70.8</td>
<td>47.3</td>
<td>51.0</td>
<td>1.1</td>
<td>390.0</td>
<td>141.2</td>
<td>103.9</td>
</tr>
<tr>
<td>Crude Fiber (g/1000 kcal)</td>
<td>1.43</td>
<td>50.5</td>
<td>23.6</td>
<td>44.7</td>
<td>1.9</td>
<td>2.09</td>
<td>65.1</td>
<td>34.7</td>
</tr>
</tbody>
</table>

* Intraindividual variability is taken as the residual in the analysis of variance and includes both true intraindividual variation and variability of the methodology.† Analysis based upon 30 × 6 observations for each sex.‡ Carbohydrate.
VARIANCE IN DIETARY DATA

TABLE 5
Relationship between mean and SD in linear regressions

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Unit</th>
<th>R'</th>
<th>R'</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
<td>Males</td>
</tr>
<tr>
<td>Energy (kcal)</td>
<td>per day</td>
<td>0.37</td>
<td>0.39</td>
</tr>
<tr>
<td>Total CHO (g)</td>
<td>per day</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>Starch (g)</td>
<td>per day</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>Other CHO (g)</td>
<td>per day</td>
<td>0.23</td>
<td>0.31</td>
</tr>
<tr>
<td>Vitamin A (IU)</td>
<td>per day</td>
<td>0.73</td>
<td>0.82</td>
</tr>
<tr>
<td>Vitamin C (mg)</td>
<td>per day</td>
<td>0.29</td>
<td>0.71</td>
</tr>
<tr>
<td>Thiamin (mg)</td>
<td>per day</td>
<td>0.38</td>
<td>0.59</td>
</tr>
<tr>
<td>Riboflavin (mg)</td>
<td>per day</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>Niacin (equiv)</td>
<td>per day</td>
<td>0.54</td>
<td>0.35</td>
</tr>
<tr>
<td>Calcium (mg)</td>
<td>per day</td>
<td>0.13</td>
<td>0.24</td>
</tr>
<tr>
<td>Iron (mg)</td>
<td>per day</td>
<td>0.47</td>
<td>0.35</td>
</tr>
<tr>
<td>Total ash (g)</td>
<td>per day</td>
<td>0.30</td>
<td>0.11</td>
</tr>
<tr>
<td>Caffeine (mg)</td>
<td>per day</td>
<td>0.53</td>
<td>0.40</td>
</tr>
<tr>
<td>Crude fiber (g)</td>
<td>per day</td>
<td>0.23</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Mined by the intake of caffeine-containing beverages, the data may suggest that the intake of such beverages decreases on non-working days. Earlier (1) it was shown that the intake of alcoholic beverages increased on the weekend day. These observations give emphasis to the need to sample days of the week in the conduct of dietary studies. McGee et al (14) analyzing data from Hawaiian men, and Todd et al (6) working with seminarians in California, have both reported significant weekend differences in food intake. It must be assumed that such differences would be the norm in both males and females.

As in the earlier analyses (1), it appeared that there were only minor differences among the three interviewers and that there was no "training effect" in the subjects. These observations are important from a methodological standpoint since they confirm that the NHLBI interviewer training and standardization procedures, when implemented with well-motivated professionals, can yield reproducible performance. A potential major source of methodological error has been controlled.

In 1979 it was reported that the ratio of the CV's of inter- to intraindividual variation ranged from a low of about unity for energy to about two for cholesterol. These ratios increased when intakes were expressed in relation to energy (the interindividual variation was reduced while the intraindividual variation, the variability of food selection, remained about the same). The display of data in Table 4 is quite consistent with the pattern of distribution of variability reported earlier (1) and again gives emphasis to the conclusion that the ratio of variabilities is nutrient specific. For the vitamins and minerals, expressing intakes in relation to energy had less effect than with energy sources.

There are major implications of these observations for nutritional studies. One day intake data, no matter how accurate, are a very poor descriptor of an individual's usual intake. The estimates of the magnitude of intraindividual variation may be used as a guideline in deciding how many replications of one day observations are needed before a sufficiently reliable estimate of usual intake is obtained (1, 5, 6, 12-14).

The SD of the estimate of the usual intake of an individual (SDU) with replicated observations of one day intakes may be predicted by the equation

$$SD_U = \frac{\text{Mean} \times CV_{\text{intra}}}{\sqrt{n} \times 100}$$

where M and CV_{intra} may be taken from Table 4 and n is the number of replicated days of observation. The "error" of the estimate of usual intake can be of critical importance in nutritional analyses in which intake and other variables are compared at the level of the individual. The presence of major intraindividual variability in the independent variable may mask correlations or bias regressions toward 0 (1, 2, 5). The magnitude of this effect can be predicted from a knowledge of the partitioning of variance in the dietary data (1, 2). The effect can be diminished by
increasing the number of replicate observations.

A similar problem arises if one uses dietary data to classify individuals (e.g., as "high/low" or "adequate/inadequate" intake). If it is usual intake that is of interest, serious errors of classification can occur when one day intakes are used. In the probability approach to the assessment of nutrient intake (13, 16, 17) failure to take into account the presence of intraindividual variability (failure to estimate usual intake) is likely to give rise to serious misinterpretation of expected adequacy of the intake (9).

As noted earlier, care must be taken in interpreting the analyses of variance for vitamin A and for caffeine. In both of these cases logarithmic transformation of the data resulted in substantive changes in the partitioning of the variance. In neither case did the observed distribution fit either a normal or log normal pattern. Nevertheless it can be safely concluded that the estimation of "usual" vitamin A intake in a population like this will prove very difficult at the level of the individual. The levels of vitamin A among individual foods varies extremely; as a result there is very wide variation in intake in the same individuals from day-to-day, and many replications would be needed to obtain a reliable estimate of the usual intake of an individual or to reduce the attenuation of correlation and regression in groups (1, 2, 9). This is apparent in table 4 from the very high proportion of variability assigned to the intraindividual component. However, because of the nature of the distribution of intake, these analyses may not provide good estimates of the number of replications needed.

The observations in Table 4 take on particular significance because of current interest in vitamin A (or perhaps β-carotene) intake in relation to cancer. In epidemiological studies there may be serious errors if 1-day intakes are used to describe individuals (if group means are the focus of interest the problem would be decreased). For vitamin A in particular, it might be preferable to use dietary techniques (such as history or food frequency approach) that provide at least a crude estimate of the usual intake of an individual (12).

The findings reported herein and in the earlier publication (1) are unlikely to be artifacts attributable to the particular method of data collection, the 24-h recall. Very recently Todd et al (6) have reported a 30-day study in which male subjects maintained taped records of weighed or measured intakes of foods or written records of estimated intakes of foods during alternate 5-day periods. In addition, at least one 24-h recall was conducted for each subject. The estimates of the coefficient of intraindividual variation in one day intakes of energy and protein were comparable to those reported by Beaton et al (1) with 24-h recalls. The weighed records showed slightly smaller variability. These observations are particularly interesting since Todd et al (6) concluded from direct comparison of recalled and recorded intake that there were major errors in the 24-h recalls; yet there was no difference in the group mean. If this picture holds more generally, then one must assume that the true errors of the three methods examined by Todd et al (6) must tend to balance out over time. At least, well-trained interviewers in Toronto solicit information with similar day-to-day variability as was seen in intakes recorded by two different techniques in California.

Although the specific estimates of the partitioning of variance reported in these studies refer only to the populations studied (generally free-living presumably healthy young to middle-age adults), the principles and implications elaborated in this and other papers apply in the general situation. If planned data analyses call for the use of intake data at the level of the individual, then preliminary studies should be conducted with the population of interest to obtain appropriate estimates of the intraindividual variation. The final decision on dietary methodology and number of replications (or duration of observation) should take this into account. Clearly the final methodology must take into account day of the week effects (1, 6, 14) and, depending on the purpose of the study, perhaps also seasonal differences. Importantly, the choice of methodology may vary also with the particular nutrient(s) under study (1, 6, 14).

Conclusions

A knowledge of the approximate partitioning of variance in food intake data is essential to both the design and interpretation of nu-
tritional studies including epidemiological studies that have a dietary component. Because a large component of intrapersonal variation in daily nutrient intake can attenuate analyzed relationships between diet and other variables, investigators must beware of the false negative conclusion. It is essential that in the design of nutritional studies the intended use of dietary data be considered, the requisite reliability of the data and the precise nature of the information needed be determined and then the decision be taken on the method of dietary data collection and the requisite number of replications. It is now clear that there is no ideal method of collecting dietary information. Rather, there may be “preferred” methods for particular uses (12, 14). Those who use dietary data for secondary analyses must be conscious of the limitations inherent in the dietary data as a result of the original methodological decision.

References