Patterns in food intake correlate with body mass index

Vipul Periwal and Carson C. Chow

Laboratory of Biological Modeling, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland

Submitted 15 March 2006; accepted in final form 30 May 2006

Periwal, Vipul, and Carson C. Chow. Patterns in food intake correlate with body mass index. Am J Physiol Endocrinol Metab 291: E929–E936, 2006. First published June 13, 2006; doi:10.1152/ajpendo.00122.2006.—Quantifying eating behavior may give clues to both the physiological and behavioral mechanisms behind weight regulation. We analyzed year-long dietary records of 29 stable-weight subjects. The records showed wide daily variations of food intake. We computed the temporal auto-correlation and skewness of food intake mass, energy, carbohydrate, fat, and protein. We also computed the cross-correlation coefficient between intake mass and intake energy. The mass of the food intake exhibits long-term trends that were positively skewed, with wide variability among individuals. The average duration of the trends (P = 0.003) and the skewness (P = 0.006) of the food intake mass were significantly correlated with mean body mass index (BMI). We also found that the lower the correlation coefficient between the energy content and the mass of food intake, the higher the BMI. Our results imply that humans in neutral energy balance eating ad libitum exhibit a long-term positive bias in the food intake that operates partially through the mass of food eaten to defend against eating too little more vigorously than eating too much.

Any features of food intake associated with excess body fat reflect a complex interplay between cognitive decisions, hormonal signals, and metabolic rate on many time scales (9, 12, 14, 16, 30, 39). Weight gain takes place over many years, so finding such patterns may require an examination of food intake over long periods of time. The immediate sensing of intake occurs partly through food intake volume or mass (22–26). Thus, one hypothesis is that long-term positively skewed trends in the energy or mass of food ingested may be indicative of a dysregulation of energy balance, since well-regulated energy balance should exhibit no significant correlations in food intake other than those required for short-term energy balance, and only positively skewed trends will lead to increased adiposity.

Additionally, because the immediate sensing of energy in food intake operates partially through food intake mass, a corollary hypothesis is that a weaker correlation between energy and mass of food would lead to a greater likelihood of overeating and, hence, higher body mass index (BMI). The assumed implication is that if the energy regulatory system is unable to predict the energy content of the food accurately, it will err on the side of eating too much over eating too little.

The present study thus examined the food records of individuals eating ad libitum over the course of 1 yr to search for any eating patterns that would be correlated to excess energy storage. Because direct measurements of body fat were not available, BMI was used as a surrogate for excess body adiposity. However, because BMI is proportional to body weight, any measurement correlated to body weight may also be correlated with BMI. Thus any measurement where the correlation with BMI is stronger than the correlation with body weight is more likely to be indicative of an association with adiposity rather than the size of the individual. This is evident in the fact that it is straightforward to find correlates of body weight in food intake patterns, but finding a direct relationship with BMI, a surrogate for adiposity, is more difficult (see Table 1).

SUBJECTS AND METHODS

Data. Data from the classic Beltsville one-year dietary intake study (20) were used. This unique data set consists of 29 individuals (13 males and 16 females) with daily food intake records for a period of 365 days and weekly weight measurements for the same period. No other body measurements or activity levels were recorded. The subjects maintained a stable body weight throughout the study with a sample mean variation in weight of 1.5 ± 0.6% over the course of the year. This data set has been investigated and validated in prior studies by the original investigators (2, 3, 15, 21) and others (33–35). The coordinators of the study verified the accuracy of the reported intakes by comparing with directly measured intakes on separate occasions.

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
during the year. There are no other data sets available with the characteristics required for our analysis.

Autocorrelation, cross-correlation functions, and skewness computation. All computations were carried out using the Lisp-Stat environment (37).

The temporal correlation functions of the measured dietary components (denoted by v and w) were computed with

$$C(T) = \frac{1}{365 - T} \sum_{s=1}^{365-T} [v(s) - \langle v \rangle][w(s + T) - \langle w \rangle]$$

where the mean values of the time series of v and w are computed over the time intervals, T, in the summation. The variables are first normalized to unit standard deviations. The uncertainties in the correlation functions were computed as the standard deviations from the mean in the summation above. For the autocorrelation (AC) function, consider the case $v = w$. For the cross-correlation coefficient, $T = 0$ is used.

The skewness was computed using

$$\text{Skewness} = \sum_{T=0}^{90} \frac{1}{365 - T} \sum_{s=1}^{365-T} [(v(s) - \langle v \rangle) [w(s + T) - \langle w \rangle]$$

where v is first normalized to unit standard deviation over the time intervals in the summation of s.

Singular spectrum analysis. The singular spectrum analysis (SSA) of the correlation functions was carried out as described in Refs. 13 and 38. SSA looks for structures in the time series by performing an eigenvector decomposition of the lagged covariance matrix over a given time window. The eigenvectors correspond to temporal struc-

Table 1. Regression $R^2$ values (P values) of DC measurements (yearly mean, zero-crossing, and skewness) vs. yearly mean body weight and BMI

<table>
<thead>
<tr>
<th>DC</th>
<th>Weight vs. Mean DC</th>
<th>BMI vs. Mean DC</th>
<th>Weight vs. DC Zero Crossing</th>
<th>BMI vs. DC Zero Crossing</th>
<th>Weight vs. DC Skewness</th>
<th>BMI vs. DC Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>0.24 (0.005)</td>
<td>0.03 (0.19)</td>
<td>0.19 (0.01)</td>
<td>0.27 (0.003)</td>
<td>0.18 (0.01)</td>
<td>0.23 (0.006)</td>
</tr>
<tr>
<td>Energy</td>
<td>0.43 (&lt;0.001)</td>
<td>0.13 (0.03)</td>
<td>0.00 (1.0)</td>
<td>0.01 (0.31)</td>
<td>0.00 (1.0)</td>
<td>0.01 (0.31)</td>
</tr>
<tr>
<td>Fat</td>
<td>0.44 (&lt;0.001)</td>
<td>0.18 (0.01)</td>
<td>0.02 (0.24)</td>
<td>0.01 (0.31)</td>
<td>0.05 (0.13)</td>
<td>0.06 (0.11)</td>
</tr>
<tr>
<td>Carbohydrate</td>
<td>0.23 (0.006)</td>
<td>0.04 (0.16)</td>
<td>0.00 (1.0)</td>
<td>0.00 (1.0)</td>
<td>0.01 (0.31)</td>
<td>0.00 (1.0)</td>
</tr>
<tr>
<td>Protein</td>
<td>0.59 (&lt;0.001)</td>
<td>0.29 (0.002)</td>
<td>0.07 (0.09)</td>
<td>0.07 (0.09)</td>
<td>0.04 (0.16)</td>
<td>0.06 (0.11)</td>
</tr>
</tbody>
</table>

DC, dietary component; BMI, body mass index. Only the zero-crossing time and skewness of the mass of food intake showed significant correlations with BMI that exceeded the correlations with body weight. A correlation between mean energy intake and body weight is expected from thermodynamic considerations and provides some validation of the data set.

![Fig. 1. Example of time series for food intake mass for 4 of the subjects in the Beltsville one-year dietary intake study. Green and red lines are running averages for 7 and 28 days, respectively.](http://ajpendo.physiology.org/)

AJP-Endocrinol Metab • VOL 291 • NOVEMBER 2006 • www.ajpendo.org
tures in the time series, and they are ranked by their eigenvalues according to the maximum possible amount of autocovariance on the time interval. Eigenvalues in the SSA analysis were tested for statistical significance as follows: 1,000 independent permutations of the AC function were analyzed using SSA. The largest eigenvalues from each permutation were listed in descending order, and only those eigenvalues above the 90th percentile of this distribution were regarded as statistically significant. If no eigenvalue qualified, only the highest eigenvalue was used to reconstruct the smoothed correlation function. The analysis was carried out with different choices of SSA time window parameters ranging from 3 to 15 days. There was no material difference in the results, so the reported results are those for a 15-day window. Uncertainties in the zero-crossing time were obtained by smoothing the uncertainty in the unsmoothed correlation function value and using the slope at the zero crossing to propagate the uncertainty in the correlation value to an uncertainty in the zero-crossing time. The uncertainty-weighted regression of the zero crossing with BMI did not differ materially from the regression with uniform weighting.

Statistical validation. Food records are well known to suffer from errors in reporting. Constant misreporting will have no effect on the AC functions or skewness measure. Random misreporting could decrease temporal correlations in the AC functions. However, as long as the misreporting is not programmatically generated to exhibit the long time correlations that we find, then they should only contribute as added random noise. As a check of the robustness of our analysis, the time series was permuted with the same permutations applied to all subjects over 1,000 permutations. The AC functions were then computed, and SSA smoothing was applied. A histogram of resulting zero-crossing times was computed and the distribution of zero-crossing times associated with the permuted time series compared with the actual zero-crossing times. The mean permuted time series zero-crossing time was 2 ± 2 days, which was much shorter than the zero-crossing times observed in the original time series. The AC function over staggered subintervals of the entire time series was also computed, and it was found that the zero-crossing times computed using these subintervals were highly correlated further, indicating that the results were not spurious. One-tailed tests for $P$ values were used as appropriate for our directional hypothesis that larger values of the zero-crossing time, skewness, or cross-correlation coefficient are associated with larger values of BMI.

Outlier detection. Regressions were computed with a leave-one-out iteration over all subjects. Cluster analysis (29) of the resulting values of $r^2$ picked out a single cluster, and two outlier subjects (common to all the regressions) whose removal from the regression led to $r^2$ values much larger than the mean values in the cluster. Upon removing these two subjects from the subject data sets, there were no more outliers in the distributions of leave-one-out $r^2$ values. The two excluded subjects exhibited implausible characteristics, with high body weight and very low reported energy intake, which may indicate severe misreporting.

BMI as a measure of adiposity. The BMI was computed by taking the average weight (measured in kilograms) of each subject over 365 days and dividing by the height (measured in meters) squared. BMI is an imperfect indicator of adiposity. However, for the population at large, the BMI is more highly correlated with body fat than any other indicator of height and weight (8).

RESULTS

The yearly means of the food intake energy, food intake mass, and mass of the components of the intake were first computed. It was found that the yearly means of the mass and...
dietary components did not correlate as strongly with BMI as with body weight (see Table 1). For example, yearly mean intake energy correlated significantly with body weight \( r^2 = 0.43, P < 0.001 \) but much more weakly with BMI \( r^2 = 0.13, P = 0.03 \). Figures 1 and 2 show examples of the time series for the mass and energy of food consumed each day for four of the subjects, representative of the diversity of the data set. Note that the mass and energy of food intake for an individual is highly variable (33). These variations are reduced when the intake is averaged over time scales of 1 wk or 1 mo, as seen in the running averages. The large diversity indicates that cross-sectional averaging of food intake time series may mask any patterns that are in the data.

We used the temporal AC function to detect any possible long-term trends in eating behavior. The AC function gave correlations in the intake over a range of time intervals. It could detect the existence of long-term trends in intake but did not discern whether these trends corresponded to overeating or undereating compared with the yearly mean. Hence, the skewness was also computed to discern this difference. Positive skewness associated with these long-term trends would imply the existence of longer episodes of mild overeating balanced by shorter episodes of more intense undereating. Our hypothesis was that the longer positively skewed eating trends persisted, the stronger the bias would be toward overeating and a higher BMI.

Examples of the AC functions of the mass and energy of food ingested for each individual subject are shown in Figs. 3 and 4 for the same four subjects in Figs. 1 and 2. It was found that the length of time for which correlated trends lasted varied widely from subject to subject for both mass intake and energy intake. Tarasuk and Beaton (34) found positive short-term temporal correlations that are consistent with our results. In some subjects, a strong weekly variation was seen (see Figs. 3D and 4D), as noted by Tarasuk and Beaton (34, 35).

We estimated the duration of trends from the AC functions (see SUBJECTS AND METHODS) and compared the results to the BMI of the subjects. The trend duration is given by the temporal interval, for which the AC function is positive (indicating a duration of positively correlated deviations from the mean). We estimated this time interval with the time the AC function first crossed zero. To obtain a good estimate of this first zero-crossing time, the AC function was smoothed using SSA (see SUBJECTS AND METHODS) (13, 38). The computed zero-crossing times for intake mass ranged from 1 to 149 days, with a mean zero-crossing time of 48 days (SD = 39 days). An analysis of a permuted surrogate data set showed that these long zero-crossing times were significant (see SUBJECTS AND METHODS). Figure 5 shows the histogram of the zero-crossing times for the food intake mass and energy correlations. There was a wide spread in the zero-crossing times, with some individuals possessing zero-crossing times as long as 150 days. These long-term trends in eating cannot be completely explained by a seasonal effect, as first observed by Tarasuk and Beaton (34, 35). We computed zero-crossing times for sliding 200-day intervals for all subjects and found that the phases and durations of the long-term trends differed significantly from subject to subject.
The zero-crossing time was linearly regressed to BMI (computed from the weight averaged over the year; see SUBJECTS AND METHODS). As seen in Fig. 7A, zero-crossing time of food intake mass was correlated to BMI with an $r^2 = 0.27$, $P = 0.003$ [BMI (kg/m$^2$) = 21($\pm$0.7) + 0.035($\pm$0.01) zero-crossing (days)]. In contrast, the zero-crossing time was less correlated with body weight ($r^2 = 0.19$, $P = 0.01$). This suggested that the relationship of zero-crossing time to BMI was operating more through adiposity than body weight. However, the zero-crossing times of energy and other dietary components either were not correlated well with BMI or the correlation with BMI was weaker than the correlation with body weight (see Table 1).

The AC function did not discern the sign of the intake mass trends. A positive value for the AC function over a given time interval could be the result of episodes of overeating or undereating with respect to the yearly mean. To test for a bias towards overeating, the integrated skewness was computed. The histograms of skewness for the mass and energy of the intake are shown in Fig. 6, showing a bias towards positive skewness. Figure 7B shows a plot of BMI vs. the average

Fig. 4. Example of autocorrelation functions of food intake energy for the same 4 subjects as Fig. 2. Darker lines are the SSA smoothed curves from which zero-crossing times are computed.

Fig. 5. Histogram of zero-crossing times for intake mass (A) and intake energy (B).
skewness. There is a statistically significant correlation of skewness with BMI with $r^2 = 0.23$, $P = 0.006$ [BMI (kg/m$^2$) = 22(±0.6) + 0.45(±0.16) skewness (days)]. However, no correlation of the skewness of the energy distribution or any other dietary component with BMI was found.

The intake mass zero-crossing time and skewness were correlated with $r^2 = 0.36$, $P < 0.001$ [zero-crossing (days) = 32.3(±7.8) + 8.3(±2.2) skewness (days)]. This indicated that the long-term trends tended to be more positive rather than negative, and the longer the trends lasted the more positive they tended to be. This was further confirmed because the double regression of BMI with these two quantities yielded an $r^2 = 0.313$, $P = 0.001$ [BMI (kg/m$^2$) = 21(±0.7) + 0.024(±0.01) zero-crossing (days) + 0.25(±0.2) skewness (days)], which was significantly less than the sum of the individual $r^2$ values. Our skewness measure was integrated over 90 days, so positive skewness implied that the long-term trends mostly involved overeating compared with the yearly mean. Because the area above the mean must be balanced by that below the mean, if the long-term trends were mostly positive, then shorter but more intense negative trends must also exist, resulting in long episodes of mild overeating balanced by shorter but larger amplitude episodes of undereating.

The cross-correlation coefficient between the intake energy and intake mass was also considered. The results are shown in Fig. 8. It was found that the cross-correlation of energy and mass was negatively correlated with BMI with $r^2 = 0.19$, $P = 0.01$ [BMI (kg/m$^2$) = 29(±2) - 11(±4) cross-correlation coefficient]. This implied that the less intake mass was correlated to intake energy, the higher the BMI.

**DISCUSSION**

The data showed that the mean values of the dietary components were more correlated with the weight of the subjects (Table 1) rather than with their BMI (1, 4–6, 23, 31). Of the five dietary components considered (food intake mass, energy, fat, carbohydrate, and protein), only the zero-crossing time and skewness of food intake mass exhibited significant correlations with mean BMI. Importantly, the correlation of the zero-crossing time and skewness of food intake mass was greater than that with mean body weight, supporting the notion that these measures were indicative of increased adiposity rather than just increased weight.

The subjects in the present data set demonstrated that a lack of correlation between mass and energy of food intake could lead to a higher BMI. In other words, over long periods of time, it was the lack of consistency in energy density, rather than energy density itself, that seemed to be most important for BMI. The long-term control of energy intake, therefore, par-
tially operated through the mass of the ingested food. It had been reported in short-term controlled studies that the volume or mass of food might be a more crucial component for satiety than the metabolically relevant energy content, so it was suggested that lower energy density of foods could lead to lower energy intake (22–26). The present work finds that this connection did not exist in the long-term ad libitum diets of the data set analyzed. However, cross-correlation coefficient of intake energy and intake mass being negatively correlated with BMI was consistent with the fact that an increase in variety in meals may lead to increased food intake (27, 28). A possible explanation for this negative correlation is: if satiety is governed partly through the mass (or volume) of food consumed, then the less that energy content is predictable by the mass (manifested as a lower cross-correlation between the two) and the less precisely energy can be controlled. Eating foods with expected energy content consistent with mass ingested might allow the body to make a more accurate prediction of the energy content. Conversely, frequently changing the energy content of foods relative to ingested mass may confuse the energy control system. If the system is biased towards defending against eating too little more vigorously than against eating too much (32), then when confused the body may elect to err on the positive, leading to eventual weight gain.

This behavior was reflected in the positively skewed trends observed in the data. The higher BMI subjects showed an eating pattern that consisted of temporally correlated episodes of slight overeating interspersed with shorter but more intense periods of undereating to compensate. These overeating episodes ranged over weeks or even months.

We note that these mild overeating episodes are masked by the wide daily variations in food intake and would never be detectable without measuring the AC function. The mechanisms behind the observed eating habits are not known. They likely involve a complex interplay between environmental circumstances, behavioral patterns, and molecular cues. It should be pointed out that, for most of the subjects, intake fell during the four 1-wk balance periods where the subjects were required to collect duplicates of the food they consumed. It is possible that these periods may have contributed to some of these long-term trends, although we found no systematic cross-sectional effect of these periods.

There were some limitations to the present analysis. One was that it used BMI as a correlate of adiposity. A more accurate measure was not available. A better measure of body composition might have provided a finer picture of the relative importance of different dietary components. Although the number of subjects (n = 27, excluding outliers) was sufficient to observe statistically significant results, it was not large enough to consider male and female subjects separately. There may also be a concern that misreporting of food intake could affect the results, but the measures we computed seemed robust against such systematic errors. Because the daily variations greatly dominated any long-term trends, it would require programmatically generated misreporting to generate the long-term trends we observed. The fact that the addition of perturbations through randomly permuting the data set did not generate any long-term zero-crossing times further supported this claim.

The results presented here may suggest strategies for containing weight gain. First, it may be important to maintain consistency in the energy content in food consumed relative to the mass of the food ingested. Second, it may be more important to monitor the total amount of food ingested over periods of 1 wk to 1 mo rather than on a day-to-day basis. A useful future investigation would be to test in a long-term study whether the existence of long-term correlated mild overeating or a weak correlation between mass and energy of food intake could be used to identify subjects at risk for obesity.

ACKNOWLEDGMENTS

We thank W. Rumpler, D. Paul, K. Hall, A. Sherman, Y. Chen, and P. Tataranni for helpful discussions and useful comments. We additionally thank W. Rumpler for providing the data.

GRANTS

This research was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases.

REFERENCES

5. Bell EA, Roe LS, and Rolls BJ. Sensory-specific satiety is affected more by volume than by energy content of a liquid food. Physiol Behav 78: 593–600, 2003.


