METHODOLOGICAL ISSUES IN WEIGHT CYCLING[1,2]

Gary Cutter, Ph.D.
AMC Cancer Research Center

Sachiko St. Jeor, Ph.D. and Robert Brunner, Ph.D.
University of Nevada School of Medicine

Pam Wolfe, M.S.
AMC Cancer Research Center

John Foreyt, Ph.D.
Baylor College of Medicine

Alan Dyer, Ph.D.
Northwestern University Medical School

Kelly D. Brownell, Ph.D.
Yale University

ABSTRACT

Recent studies have suggested that weight changes may be related to disease risk independent of weight status. A critical step in testing this assertion is the measurement of weight change and so-called "weight cycling." However intuitive the concept of weight cycling may appear, research in this area is hampered by complex methodological issues. This article discusses various measures of nominal weight cycling, including the standard deviation, coefficient of variation, regression techniques, and cycles. A cycle is a sequence of a gain followed by a loss or vice versa. The various measures are compared in seven hypothetical cases created to illustrate their strengths and weaknesses. Superior performance of the cycles measure over the coefficient of variation, number of fluctuations, and simple regression methods is argued. The linkage of the cycles measure with the statistical theory of runs also provides a basis for testing the significance of weight fluctuations or other variables that may cycle, such as blood lipids, etc. The cycles measure and runs test provide a viable definition for identifying weight cycling and a tool for evaluating the critical amount of weight gained and/or lost in relationship to risk.

(**Ann Behav Med** 1996, *18*(4):280–289)

[1] Preparation of this manuscript was supported in part by NIH/NHLBI Grant #HL34589.

[2] The authors would like to thank two anonymous reviewers for their efforts. Their comments contributed substantially to the revision of this paper.

Reprint Address: G. Cutter, Ph.D., Director, Division of Biostatistics, AMC Cancer Research Center, 1600 Pierce Street, Denver, CO 80114.

INTRODUCTION

Obesity has been recognized as a risk factor for many health problems and is receiving increased attention in the scientific community and popular media. Recent studies suggest that weight change may be related to disease risk independent of weight status (1,2). These results agree with earlier findings from Framingham (3), which also reported that weight change over time was more strongly related to coronary disease risk than was degree of adiposity. Furthermore, inferences about the net effect of weight fluctuations are emerging from both animal and human studies (1–23).

While the focus on risk associated with weight fluctuation has increased, the many pitfalls and failures of dieting have gained increased media attention. From often confusing messages, the problems of "yo-yo dieting," a term commonly used for weight cycling, have drawn special attention. While the idea has intuitive appeal, research in the area of weight cycling leaves many questions unanswered. Recently, in a two-part review of methodological issues in studies of obesity (24,25), Kraemer, Berkowitz, and Hammer pointed out some difficulties in measurement. Notable were problems of reliability and the failure of cross-sectional measurements of weight and body fat to provide insight into the processes or mechanisms underlying obesity. Other studies of weight cycling have also strongly recommended longitudinal designs, which are supported by a convincing rationale (26).

This article should start with a definition of weight cycling; unfortunately, there is no standard. The National Task Force on the Prevention and Treatment of Obesity recently published a review of original reports on weight cycling, yo-yo dieting, and weight fluctuation with the objective of addressing "concerns about the effects of weight cycling and (providing) guidance on the risk-to-benefit ratio of attempts at weight loss, given current scientific knowledge" (4). The definition of weight cycling varied among the 26 articles listed in the review, with most definitions incorporating some form of weight change threshold and some number of gain/loss or loss/gain cycles over a specified time period. One article defined weight cycling based on individual variability about a slope; several used number of diets or number of diets combined with mean lifetime weight loss of a specified magnitude; some used self-report; and three used the coefficient of variation (the standard deviation divided by the mean).

Since it is difficult to assess the risk associated with a factor not clearly defined, it is not surprising that the available evidence for weight cycling as a risk factor is not compelling either from an empirical perspective or for identifying a rudimentary causal model. At this stage in weight cycling research, the major emphasis needs to be on the identification of the important elements of cycling behavior, criteria that will distinguish cyclers from non-cyclers, and methods to assess the associated health risks. Resolving these issues requires an operational definition of a weight cycler.

The terms weight cycling and yo-yo dieting invoke a clear mental image and are likely to implicate a variety of potential biobehavioral risk factors. All persons fluctuate, although most short-term changes are small. The challenge for a summary measure of weight fluctuations is to provide one that captures a pattern or patterns which can be evaluated to assess even the potential of elevated risk. This is not a simple task. Intuitively, one visualizes a pattern of gains and losses but not necessarily the magnitude or duration of the fluctuations or the time period they span. It is the temporal ordering and duration that introduce a new level of complexity for defining weight fluctuation. Measurements taken daily, weekly, monthly, or annually would likely produce different values. For large intervals of time between weight measurements, we observe only smoothed net weight changes which result from complex biopsychosocial changes rather than isolated short-term changes. Longer intervals miss short-term fluctuations that may be relevant as the impact of interval lengths during which weight change occurs is not yet known. If one person loses and regains 25 pounds in four months and another does the same over four years, one suspects these patterns have different biological effects. Measurement issues are critical to attempts to identify levels of fluctuation that may alter risk levels for some outcome. The final measure must also separate effects from their covariance with body weight level as such.

CRITERIA

This paper is not aimed at the validity of the weight cycling phenomenon, but rather the issues surrounding measurement. Further, while weight cycling is used as the focus, the measurement issues apply to any phenomenon that may cycle.

We have focused primarily on the development of a working definition, including a measure with tractable statistical qualities that captures the intuitive meaning of cycling. Working with a small hypothetical data set, we compared our proposed measure with others that might be used to represent weight fluctuations. Also, we considered sensitivity to overall change as well as ability to separate time (frequency), magnitude (amplitude), direction, and ordering of weight changes. Finally, we comment on clinical interpretability.

Sensitivity to change is the *sine qua non* for any measure of fluctuation. Measures that are insensitive to change when it occurs will be of no use. But change has several dimensions.

TABLE 1

Annual Weights for Seven Persons

Year	Person A	Person B	Person C	Person D	Person E	Person F	Person G
1	160	160	180	180	160	200	187
2	165	160	174	175	175	195	182
3	171	160	170	170	190	190	189
4	166	160	165	175	180	185	181
5	160	160	160	180	170	180	190
6	165	165	160	174	175	175	180
7	170	170	160	171	180	170	189
8	174	174	160	168	185	180	179
9	178	179	160	163	190	190	185
10	182	182	160	160	195	175	180
11	187	187	160	155	200	160	187

Time and magnitude are obvious quantifiers of change. Time will have different interpretations depending on how we measure fluctuation; the term frequency will replace time and amplitude will replace magnitude when we consider time series analysis. In weight change, the long-term biological effects of order and direction of change are not well understood. Does weight loss followed by weight gain have different biological effects than weight gain followed by weight loss? In the short term, for example, we know that blood pressure and cholesterol level fall with weight loss; but we don't know whether the fall is "on average," and it may or may not occur in a specific individual.

Measures of weight cycling capable of separating each of these patterns of change could be used to assess biological impact of cycle components, if any, and combine cycles that may be apparently but not functionally different over time. In addition, if we are to distinguish cyclers from non-cyclers and further define a group that is extremely stable (i.e. true maintainers of their weight), a measure with straightforward statistical properties will hold some appeal. There are sophisticated statistical time series models that will tease apart the components of a series of data. However, the utility of a general measure, as opposed to an analytical tool, is its clinical interpretability and practical applicability. Setting aside for the moment the choice of an appropriate threshold level for change and time interval between measurements (weigh-in), we define a cycle as a gain followed by a loss or a loss followed by a gain. We will show that this simple intuitive definition meets, directly or indirectly, all criteria cited above. To illustrate our points, we present seven hypothetical subjects whose weight sequences are illustrated in Table 1 and plotted in Figure 1. We arbitrarily use eleven annual measurements, providing ten intervals of change. Subject G is intended to be our classic weight cycler, while the other six appear to be a gainer (Subject B), a loser (Subject C), and four fluctuators (Subjects A, D, E, and F).

Starting with mean weight, we have computed a series of increasingly complex measures for our hypothetical subjects. Our aim is to identify which elements of interest each measure reveals and what important information is lost based on the criteria discussed above. Often the Body Mass Index (BMI) is used in the literature rather than weight alone, and it could be used interchangeably in most of our arguments. For simplicity, we used only weight.

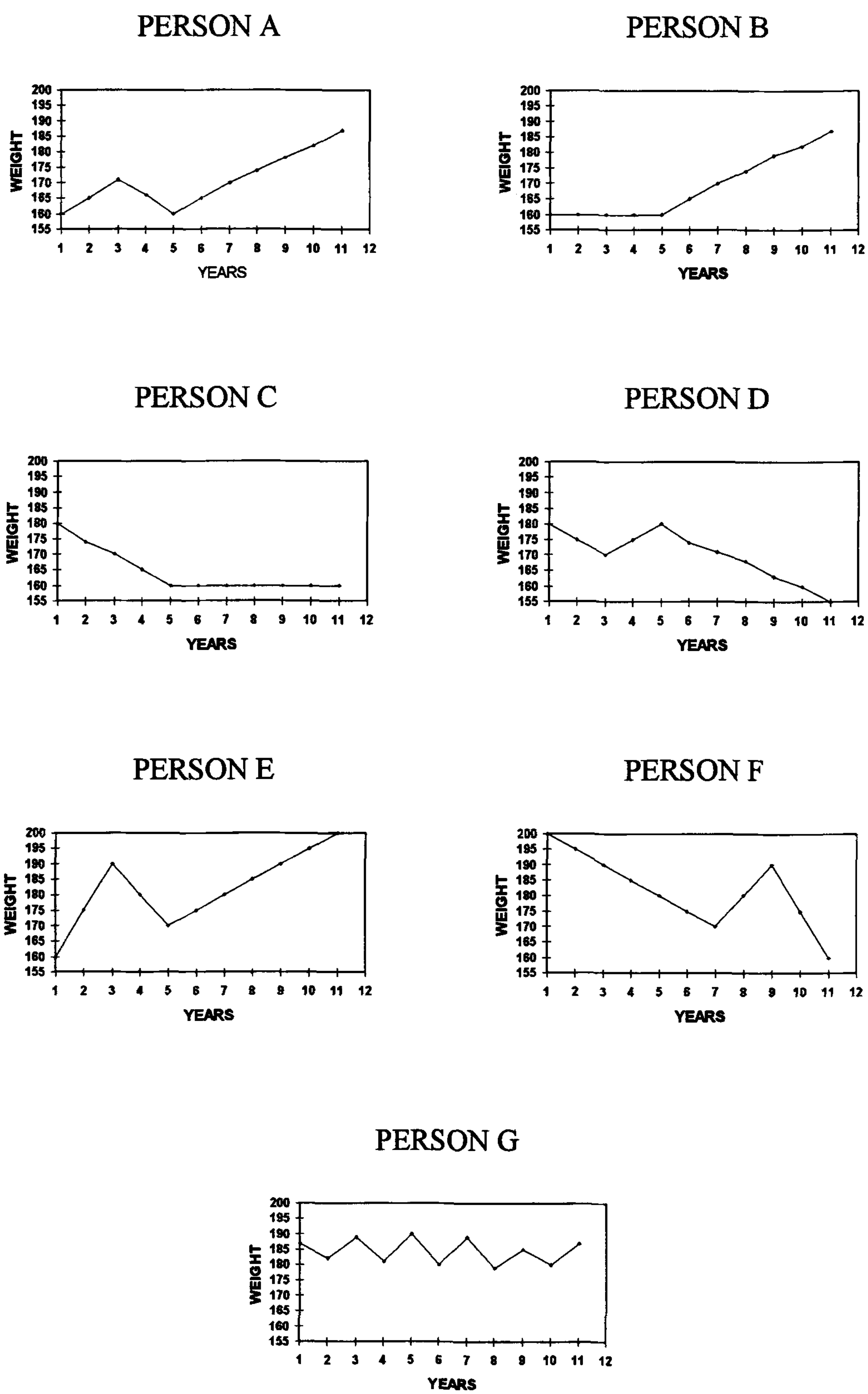

FIGURE 1: Annual weights of seven hypothetical persons.

Simple Measures Characterizing Weight Over Time

Average or mean weight over a specified time period of measurement is a simple way of integrating a series of measures taken over time. Alternatively, one might use the median weight to reduce the influence of a few observations that may be unreliable. However, neither of these two measures taken alone reveals change. For example, without a record of the eleven weights, Subject G (Table 2A) is indistinguishable from one who weighed 185 pounds at each weigh-in. More importantly, neither the mean nor the median is sensitive to time at any

TABLE 2A
Measures of Weight Fluctuation Based on Ten Annual Weights

Weight Fluctuation Measure	A	B	C	D	E	F	G
(1) Mean	170.7	168.8	164.5	170.1	181.8	181.8	184.5
(2) Median	168	160	160	170	180	180	190
(3) Overall Weight Change	27	27	−20	−25	40	−40	0
(4) Abs Weight Change	49	27	20	45	80	80	76
(5) % Weight Change (3)/base	0.2%	0.2%	−0.1%	−0.1%	0.3%	−0.2%	0%
(6) Time Weighted Weight Change	2.7	3.0	0.7	−2.5	3.0	−3.6	0.4
(7) Variance	77.4	104.0	50.3	64.5	136.4	136.4	17.3
(8) Standard Deviation	8.799	10.196	7.090	8.031	11.677	11.677	4.156
(9) Coefficient of Variation	5.15%	6.04%	4.31%	4.72%	6.42%	6.42%	2.25%
(10) MGL-Baseline	G	G	L	L	G	L	M
(11) MGL-Interval	GGLLGGM MMG	MMMMGGM GMG	LMLLMMM MMM	LLGGLMM LML	GGLLGGG GGG	LLLLLLGG LGG	LGLGLGL GLG
(12) # of Fluctuations	3	1	2	3	4	4	4
(13) Crossing (abs)	7	4	3	5	16	10	12
(14) Crossing-Score	3	4	−3	−3	8	−8	−2
(15) Crossing (abs-tm)	0.5	0.4	0.1	0.5	1.0	1.6	1.0
(16) Crossing (time weighted)	0.3	0.4	−0.1	−0.3	0.6	−0.7	−0.1
(17) Cycle/ Number of Cycles	2	0	0	2	2	2	9
Number of Runs	5	1	1	5	5	5	1

particular weight or to the direction or order of changes in the weights. A person who loses 55 pounds over eleven years would have the same mean weight as a person who gains 55 pounds assuming the highest and lowest weights are the same (but the direction is reversed). Similarly, a person with the same weight for ten years who gains 55 pounds in the last year would have the same mean weight as a person who gains 5 pounds per year; and though the medians would differ, the order in which the weights occurred would require a separate statistic. Thus, both the mean and the median are insensitive to frequency of weight changes.

The deficiencies associated with the simple estimates of mean or median are often overlooked because the clinical interpretation is readily obvious. These measures suggest a total body burden in the form of an integrated risk as it is measured over time. Coupling this with the individual's weight history and, most importantly, with the most recent measure of weight, a clinician can give standard advice to a patient about his or her weight status.

Although considerable effort goes into recording interim weights in longitudinal studies, much of the information is not utilized. Quite often the data are summarized simply as overall weight change (i.e. the last measurement minus the first). This measure is sensitive to the difference between the endpoint weight measures but totally insensitive to the interim weight changes; in particular, it is insensitive to time or duration at a particular weight or any ordering of how one arrived at the final weight. As this measure depends on only two points, a hundred pounds gained in the last year is not distinguishable from a hundred pounds gained in the first year and maintained thereafter or from 20 pounds gained at each of five consecutive years. The index, which contrasts the initial and final weight, is only sensitive to the magnitude and the direction of that overall change. Again, this measure fails to use any information from interim measurements and does not reflect time, order, frequency, or individual magnitudes of multiple weight changes. Such an overall measure is useful clinically as it is based on the notion that current weight is paramount, and such a simple change indicator provides a history which, while limited, may be helpful to the clinician's expectations about the future path and its modification.

In order to improve on the deficiency of measures that do not utilize interim information, investigators often try to characterize the weight fluctuations. Thus, if a person is measured once a year for eleven years, these eleven measurements will yield ten change scores or weight fluctuations. In general, if there are N observations of weight, there are $N-1$ interval changes. To summarize these changes, the average weight change would seem reasonable. Unfortunately, this summary degenerates into the net weight change defined by the last weight observation minus the first weight observation divided by the number of changes. That is, it provides the same information as overall weight change, scaled by a factor of 1/10. To salvage the interim change information, we consider absolute weight change, the sum over the absolute values of the interim change scores. This measure fails to capture the order, frequency, and direction of changes but summarizes clearly the cumulative magnitude of fluctuations. Absolute weight change treats all changes as biologically equivalent regardless of direction. Taken alone, this measure is not useful clinically until the body burden of such change is shown to be a risk factor. Further, for this measure, people gaining and losing weight can have the same score (compare Subjects E and F in Table 2A).

Unlike absolute weight change, the percent of weight change combines overall weight change with initial weight to give an indication of relative change. Keying weight change to the starting weight adds to the available information, since a 10-pound change in a 300-pound person may not have the same effect as a 10-pound change in a 100-pound person.

Another alternative to the absolute weight change measure is the time weighted average weight change which can be constructed to give more or less importance to more recent measures, depending on how elapsed time affects the risk associated with a change in weight. Many choices of weighting coefficients can be used. For example, for eleven annual weights, one might weight inversely with time on the assumption that recent changes are more important for risk modification. Conversely, if bouts of weight gain and loss insult the vascular system and contribute to an atherosclerotic process, larger weights should be applied to earlier observations. For Table 2A, we chose the first assumption and used the formula:

Time weighted average weight change

$$= TWWC = \frac{(wt12) + 2(wt23) + \ldots + 10(wt10,11)}{55} \quad (1)$$

where:

wt12 = weight difference @ time 1-2

wt23 = weight difference @ time 2-3

.

.

.

wt10,11 = weight difference @ time 10-11

The denominator is the sum of the weights (1 + 2 + ... + 10). Note that this measure assigns similar values to Subjects B and E and to Subjects C and G, masking the distinct differences in the patterns of fluctuation. Our choice of weighting factors drives the outcome and is yet another area for research.

Variability

None of the measures discussed so far captures information about individual cycles. Overall weight change, percent weight change, and time weighted weight change are signed; so net direction of change is apparent, but interim patterns are not. Mean and median are static and absolute weight change captures only overall magnitude. Frequently, only simple measures of variability have been used to define fluctuators or yo-yo dieters (1,16,21).

A statistic that often accompanies the mean is the variance or its square root, the standard deviation. The within-person variance and standard deviation provide individual measures of the dispersion of multiple measurements of weight. These are independent of both mean weight and the ordering of the sequence of the measures.

In some applications, a measure of variance is exactly what is sought: a summary statistic that gives only information about the average squared deviation from the mean but is unaffected by the mean level. Similarly, the dispersion among all observations, especially when there is no natural ordering of the observations, is a valid summary of weight fluctuation.

However, neither the variance nor the standard deviation is sensitive to the direction of weight change. The same values are obtained for a person gaining 55 pounds over eleven years in 5-pound increments, a person losing 55 pounds over the same interval, or a person whose weight was the same at endpoint as baseline but who had sufficient fluctuations between measurements. Subjects E and F have the highest variance, but the variance fails to distinguish Subjects C, D, and A.

Variance is difficult to explain in a clinical context because patients are unfamiliar with the concept and have difficulty in contemplating squared units. However, the within-person standard deviation provides a useful measure in this regard by simply providing a number in pounds for how far an individual's weight, on average, has deviated from its mean. This is an improvement over variance but provides no measure of magnitude relative to initial body size. This is often remedied by using the coefficient of variation (CV), which is the within-person standard deviation divided by the mean weight, sometimes taken to be body weight at the start of the observation period or the average weight taken over time. The coefficient of variation yields a measure that can be interpreted as an average percentage variation in body weight.

The coefficient of variation for Subject G is the smallest at 2.25%. Subject B gains 27 pounds with the majority of the weight gained in the last five years. There is no yo-yoing, but the CV is higher (6.04%). Subject C has a net weight loss over the first four years of 20 pounds and is then stable for the following six years with a lower CV of 4.31%. Subjects E and F have the same means, standard deviations, and largest CVs and exhibit extensive yo-yoing, but Subject E has gained 40 pounds over the eleven years whereas Subject F has lost 40 pounds. The equality in variability of Subjects E and F may be reasonable, but there could be substantial differences in the effects on cardiovascular disease risk factors, such as blood pressure or cholesterol.

The coefficient of variation is a reasonable starting point as a summary statistic, but it has deficiencies which are in the time, order, and direction of changes. If these characteristics are not important in defining risk or predicting outcome, this may not be a problem, but one needs a measure that can test their importance before ruling them out. For our hypothetical subjects, the coefficient of variation yields about the same information as variance since their mean weights are similar.

ORDER, DIRECTION, INTERNAL PATTERNS

The mental image conjured up by the label yo-yo dieting is one of oscillating changes in weight that occur with some regularity. It is the oscillations, their magnitude, and changing directions with respect to time that are crucial to the mental construct. We turn now to measures that emphasize these attributes of change.

Categorical Measures of Fluctuation (Maintainers, Gainers, and Losers)

Several additional indices can be devised to correct the deficiencies of the standard approaches. Foremost among these is a classification into categories called maintainers, gainers, and losers (M, G, and L) which are easy to compute and interpret individually. Each category is formed by a weight change cutoff, arbitrarily set at five pounds for this discussion. Two measures are considered. The first measure simply subtracts the baseline weight from the final weight and is denoted as MGL-baseline. This measure represents only the difference between beginning and ending weights and does not account for time, frequency, or order. It carries less information than overall weight change (i.e. we see direction but magnitude is lost). The advantage is that this measure is easy to interpret and compute.

A modification of this MGL classification is to characterize an individual successively over each interval of measurement, denoted as MGL-interval. This measure classifies the change in weight from year to year providing a sequence of MGL categories, marking the transitions on an interval by interval basis. This measure must be augmented by MGL-baseline or similar measure of net change, because with a five-pound threshold, a person could gain four pounds per year over eleven years for a total weight gain of 44 pounds and yet would be considered a maintainer for each of the ten intervals. While this could be a problem if we were concerned with weight gain or obesity, the labelling correctly reports no cycling. Of course, cycles with magnitude less than five pounds will be missed, but this relates to the setting of the threshold.

While the measure now captures the interim visit information, its characterization of frequency and magnitude are limited by the predetermined intervals of weighting and the preset threshold level. It may be sensible to choose a threshold that is based on a percentage of body weight, because the same weight change may have different biological effects and clinical meaning depending on an individual's initial weight. We will return to the issues of threshold levels and weigh-in intervals later.

MGL-interval is sensitive to time, direction, and order within the constraints mentioned above. Magnitude of change is lost; when the threshold is exceeded, changes of 50 pounds are indistinguishable from changes of 10 pounds. However, the clinical interpretation of the patterns of Ms, Gs, or Ls seems simple to present and understand.

Change Scoring of Weight Fluctuations

The MGL-interval sequence presents analytical difficulties at the individual level. It is unclear how to relate various individual sequences to other variables as the number of observations increases. It is tempting to group individuals with like patterns and compare changes in risk factors, but the number of distinct patterns could be quite large relative to the sample size as the number of weights recorded increases. An alternative measure of fluctuation, number of fluctuations, counts the number of changes that exceed a threshold. One can simply count the number of Gs and Ls in the sequence as a summary measure. This measure is sensitive to frequency of change but is insensitive to the order, magnitude, and direction of change. It has the same clinical meaning as MGL-interval. A minor modification replaces the M with a zero value, the G with a $+1$, and the L with a -1 and sums over the intervals. A deficiency of this measure is that it groups the fluctuators with the maintainers (a total score close to zero), defeating the purpose of the interval change measures. It is a simple and powerful measure of net gain or loss but misses the mark entirely for fluctuations.

Counting Measures of Threshold Changes

Since one expects weight to fluctuate naturally to some extent, the concept of a significant fluctuation may be best characterized as deviation above a threshold. This threshold represents the point beyond which weight change may have a biological effect for a person at a given weight. One biological consequence that has been the object of speculation suggests that the body recalibrates or resets metabolic function in the face of changes greater than some threshold value. That is, weight fluctuations within a yet-to-be-determined range will not increase one's risk independent of weight gain. This threshold approach is in keeping with the maintainer, gainer, and loser categorization. A person who is a maintainer across ten intervals and gains as much as 40 pounds is at increased risk, but it is from the weight gain and not the weight fluctuations. Until research identifies threshold amounts of weight (if they exist at all), the choice is arbitrary.

Another measure that attempts to enumerate the number of times a person changes weight by this threshold amount can be called crossings. Crossings(abs) is a measure of fluctuations; it counts changes of a specified amount, independent of directional shifts. From this counting process, one records the longitudinal weight history or stability of weight from a reference time point (a baseline or interim weight) to a future time point. Crossings(abs) represents a total count of threshold changes, similar to the absolute weight changes (they are equal when the threshold value is set at one pound). This measure provides an overall assessment that is sensitive to the frequency and magnitude of changes and has limited clinical interpretation. Still, it lacks information on the timing of the weight changes, the order, and the direction of the changes.

A simple modification of the crossings measure is to create a so-called score statistic or crossing score. This score counts the number of times the threshold limit, say five pounds, is exceeded or crossed and further includes the sign of the direction of the change. This would provide a measure sensitive to overall change, similar to overall weight change, while losing information on interim patterns since the positive and negative components cancel. A person with a 40-pound weight gain in four increments of ten pounds (yearly component $+2$ crossings of five-pound limit) for each of four years (crossing score $= 2 + 2 + 2 + 2 = 8$) would not be a weight cycler but has the same crossing score as one who gained 40, lost 15, gained 25, and lost 10 pounds (crossing score $= 8 - 3 + 5 - 2 = 8$). Although the number of unsigned crossings is different (8 versus 18), the lack of difference in the crossing score indicates its lack of sensitivity to frequency, magnitude, and order of change.

A further modification of this approach that helps compensate for the lack of order information and the recency of the results uses time weighted crossings and time weighted crossing scores. We used the weighting scheme described above for time weighted weight change. Crossings and the MGL-interval are very similar, except that crossings provide a refinement by counting the number of thresholds crossed rather than a simple gain or loss of the threshold amount. Time weighting improves time and order sensitivity, but both approaches still share the deficiencies of the non-weighted versions. The time weighted counting version (crossings) records overall fluctuations allowing emphasis on recent threshold crossings, but it ignores the direction of change. On the other hand, the time weighted crossing score provides emphasis on the net number of thresholds crossed (gains and losses still tend to cancel each other) with greater emphasis placed on more recent events.

CYCLES

While many of the above measures have been utilized in discussions of weight fluctuation, all are somewhat deficient in estimating vital aspects of cycling. The course of weight over time as a series of maintainer, gainer, and loser categories captures most closely these elements. A simple mechanism is needed to summarize them.

A mechanism to overcome this problem is based on rethinking the question we are asking. The issue of weight cycling can be thought of as a question of whether the changes in weight observed over successive intervals are random or arise from some other process. Statistically, this problem has been solved using a technique called the runs test. A run is defined as a succession of like values (GGG or LLL), and the number of runs in a sequence is one plus the number of juxtapositions of unlike neighbors. For example, GLLGGLGGGL consists of six runs. Calculating the probability that a given sequence was generated by a random process or that several sequences resulted from the same underlying process is straightforward (27).

The presence of M forces a further adjustment; if we define GL and LG to be a cycle (C) and GG, LL, MG, GM, ML, and LM all to be not-a-cycle (N), then the sequences derived for MGL-intervals can be converted into cycles. From this simple procedure we define our proposed measure of weight fluctuation: the number of cycles. This measure is derived from a coding of the information that has the statistical properties required for the runs test (28).

Not only does this measure seem obvious, but it provides a mechanism to assess whether a subject is experiencing more fluctuations than would be expected by chance or whether there are too few (i.e. true maintenance). Sample size and power analyses identifying the number of intervals necessary to adequately determine weight fluctuation can be derived to aid the design of epidemiological studies and intervention strategies. Although the effectiveness of the measure depends on the choice of threshold levels and weigh-in intervals, the technique may be useful in the search for the appropriate values, since it separates weight changes that are not sufficiently large to allow us to differentiate between random variation in the majority of cases and weight changes that surely identify individuals involved in some active feedback process, showing too much or even too little periodicity.

The minimum sample size required for the statistical test exceeds ten measurements (nine intervals) (28), but the cycles information can still be used as a summary measure if fewer observations are present. Taking a longer series for our first example, suppose we have three years of monthly weights for one subject and we find that of the 36 monthly change measures there are 18 intervals of a five pound or more gain followed by a five pound or more loss (a cycle, C). Given 18 cycles, there must be at least 25 runs [25 changes from C to N, a non-cycle (GG, LL MG, etc.), or from N to C] or fewer than 13 runs to declare an individual a statistically significant cycler or a maintainer. In the first case (where there are more than 25 runs), the Cs alternate more regularly with the Ns than we would expect by chance; in the second case, the Cs are grouped together in a way that also suggests we are not looking at a random process. To make the point more intuitive, one would suspect that neither CNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCN nor NNNNNNNNNNNNNNNNNNCCCCCCCCCCCCCCCCCC was generated by a random process (28, pp. 252, 253).

The cycles measure for Subject E is computed as follows: the subject's MGL pattern is GGLLGGGGGG and from this the cycles are NCNCNNNNN. With two Ns and seven Cs, tables F_1 and F_2 (28) show that this pattern is not statistically different from a random sequence of Cs and Ns. Subject G, on the other hand, has MGL of LGLGLGLGLG which yields cycles CCCCCCCCC. This is clearly not a random pattern. Subjects B and C have no cycles using this measure. Each has one run, but it is all Ns; they are not cycling.

Consider again the information in Table 2A, where the measures discussed above are computed for the seven hypothetical individuals whose weight data are shown in Table 1. Subject G best fits our idealized view of a yo-yo dieter. The variances, standard deviations, and coefficient of variation yield similar information concerning fluctuations. Subjects E and F are the most variable followed by B and then A. Subjects C and D are not distinguishable. These measures capture well the magnitude of change but entirely miss the pattern of gain and loss that is at the center of our notion of cycling. Absolute weight change separates Subject G from Subjects A, B, C, and D, but the distinction is an artifact of the data. That is, a person who gained (or lost) seven pounds each year would also have an absolute weight change of 77. The time weighted change provides another picture similar to the above measures, while emphasizing more recent information. The MGL-baseline measure, the overall weight change, and the percent weight change fail to distinguish Subject G from a person whose weight remains stable and in general carry no information about interim patterns. The MGL-interval measure provides a nice categorical summary of the path of weight change.

Crossings (abs) provides about the same information as absolute weight change (scaled differently), while the crossing score is analogous to overall weight change. The absolute value time weighted crossing score carries similar information to crossings (abs), while the time weighted crossing score provides similar distinctions to those seen in the crossing score. Both are scaled and weighted to emphasize recent changes.

The final measure, number of cycles, shows Subject G with nine cycles, the maximum number possible with ten intervals. Subject G is followed by Subjects A, D, E, and F with two cycles each, and this measure shows no cycles for Subjects B and C. It is our choice of threshold that puts Subject A in the same category with Subjects D, E, and F. Although the measure loses the magnitude of the changes (as do all threshold measures except absolute crossings), it captures frequency of change. Number of cycles masks the order of change, but the preliminary recording of successive changes into MGL-intervals required to count the cycles shows the order of change. Cycles and number of cycles lead to a very easy and clear interpretation and allow the efficient combination of like groups [i.e. GLGL contains the same number of cycles (three cycles) as LGLG]. Clearly, the choice of a threshold and the interval of data collection are crucial to the performance and interpretation of this measure. We turn now to these issues.

TABLE 2B

Measures of Weight Fluctuation Based on Ten Annual Weights

Weight Fluctuation Measure	A	B	C	D	E	F	G
(18) Time Series–Linear trend $Beta_0$ baseline	159.36	154.36	173.36	180.50	168.18	195.45	185.59
$Beta_1$ trend	2.27	2.89	−1.78	−2.08	2.73	−2.73	−0.23
Residual standard deviation	4.78	3.66	4.13	4.32	7.78	7.78	4.31
Coefficient of Variation	2.80	2.17	2.51	2.54	4.28	4.28	2.34
R-squared	0.73	0.88	0.69	0.74	0.60	0.60	0.03
(19) Time Series–Cosinor Model							
$Beta_0$ baseline	158.28	159.32	179.15	181.48	166.40	195.25	185.25
$Beta_1$ trend	2.35	2.89	−1.78	−2.15	2.85	−2.85	−0.23
$Beta_2$ (abs. value) amplitude	4.56	8.63	10.08	4.11	7.48	7.48	3.72
Period	8 yr	20 yr	20 yr	8 yr	8 yr	8 yr	2 yr
f(t)	$\sin(2\pi t/8)$	$\sin(2\pi t/20)$	$\sin(2\pi t/20)$	$\sin(2\pi t/8)$	$\sin(2\pi t/8)$	$\cos(2\pi t/8)$	$\cos(4\pi t/4)$
Residual standard deviation	3.49	1.50	1.30	3.17	5.63	5.63	1.43
Coefficient of Variation	2.05	0.89	0.79	1.86	3.10	3.10	0.78
R-squared	0.87	0.98	0.97	0.88	0.81	0.81	0.91

Continuous or Regression Measures of Weight Fluctuation

We claimed in our introduction that the choice of weigh-in interval would affect any conclusions that can be drawn from the data. Clearly, changes occurring at a higher frequency than the weights are observed will go undetected. Preliminary research must be conducted in which weights are recorded frequently enough to capture all changes that could be of scientific interest. Such data could be used in statistical analyses with the goal of developing methods to evaluate the effect on known biological risk factors of the frequency and amplitude of weight changes. These two dimensions of weight fluctuation are at the center of our notion of cycling. We now offer a brief sketch of the direction this analysis might take.

A Road Map for Analysis to Establish Threshold Values and the Appropriate Intervals of Weight Measurements

The purpose of this paper is to describe measures without getting embroiled in the issues surrounding psychobiological process and whether weight cycling is or is not a risk factor for morbidity or mortality. However, we felt it useful to include a road map of how research should proceed to identify the key factors of frequency and amplitude. We see these techniques as necessary accompaniments to the research needed to establish the appropriate threshold and measurement times to make the cycles definition valid and reliable, whether for weight cycling or other biological periodicities.

Linear regression is a powerful technique that can simultaneously provide a summary of the baseline weight, the trend or average weight gain or loss, and an overall measure of weight fluctuation. This would not be the method of choice for longitudinal data because of the special characteristics of data collected over time, but it is one technique listed in the collection of reports reviewed by the National Task Force on the Prevention and Treatment of Obesity (4). The simple version uses weight as the dependent variable and time as the independent variable. The measure of weight fluctuation is the residual variability about the time trend line (the square root of the mean square error from the regression model). Also, a coefficient of variation can be computed by dividing the standard deviation about the regression line by the mean of the dependent variable. Note that the use of baseline weight in place of mean weight can lead to the confounding of the relative variability with the direction of change overall. For example, Subjects E and F exhibit identical patterns of change, one the mirror image of the other. The CVs reported in Table 2B are identical, but if initial weights were substituted, the CVs would be 4.86% and 3.89%, respectively. Table 2B shows regression coefficients for the seven persons reported in Table 1.

While this model does capture overall weight change across the time interval ($beta_1$) and variability (about a trend line instead of about the mean), it does not provide information on order, duration, frequency, or direction of the variability. In fact, the residual variability could indicate lack of a linear trend, which is not necessarily an indication of weight cycling.

As an example, compare the results for Subjects C and G. One visualizes Subject G as more consistent with our image of a yo-yo dieter, but the residual standard deviations are similar. A plot of the data shows that Subject C exhibits a non-linear downward trend; it is the non-linearity that is driving the residual error, not cycling behavior. The model gives interpretable coefficients for Subjects E and F, insofar as the trends are in opposite directions and the CVs match at 4.28, but the explanatory power is low ($R^2 = 0.60$), even with only eleven data points.

A refinement of this simple model is to take a time series approach to the data. There are techniques for handling the complications arising from the serial correlation we expect to find in longitudinal data. Since the data we use here represent one-year intervals, the problem of serial correlation is likely lessened. However, if the data were gathered weekly or monthly, we would expect the measurements taken on a particular subject to be much more highly correlated over time. Recent developments in statistical theory allow efficient utilization of longitudinal information and provide a methodology to analyze correlated data (29,30). This approach can provide a summary path for each individual as well as a measure of that person's deviation from the overall population model. The technique may be exceptionally useful in teasing apart the components of risk potentially associated with weight change and those solely related to the yo-yoing. A very readable description of how some of these techniques differ from least squares regression was published recently in the *American Statistician* (31).

An alternative analytical approach, initially somewhat simpler, is to specify a structural time series model that will capture an overall trend (if a trend is present), while separating explicitly the frequency and amplitude of the cycle (32,33). The model will estimate an intercept term, which indicates the estimated baseline weight; a slope term, which describes the overall weight gain or loss over the entire period; and a coefficient that estimates the magnitude of the weight fluctuations for our idealized pattern of yo-yoing. For each individual the frequency is embedded in the design matrix. The model can be written:

$Wt(t) = beta_0 + beta_1(t) + beta_2f(t) + \epsilon$

Where (t) indicates the time in equally spaced intervals; and $f(t) = Sin(2\pi t/p + C)$ or $Cos(2\pi t/p + C)$; a sine (or cosine) curve based on measurements at time points t, with origin at C (the phase coefficient), period 2π, p measurements per period, and amplitude $beta_2$. The error term, ϵ, is not restricted to normality.

To illustrate how this model works, once again consider the data in Table 1. The eleven weights were used in the above regression model to estimate the coefficients by subject. The values for (t) are 0, 1, 2, . . . 10 indicating baseline and the number of years since baseline. If we assume for simplicity the origin is at zero (C = 0), the period is eight years, and the amplitude is the only parameter to be estimated, then f(t) = 0, 1, 0, −1, . . . for a sine function or f(t) = 1, 0, −1, 0, . . . for a cosine function. Subject G appears to complete a cycle every two years, so f(t) = 1, −1, 1, −1, . . . a cosine function tailored to capture the higher frequency. Subjects B and C are on a longer cycle (if they are cycling at all) so f(t) = sin(2pi*t/20) = 0, 0.31, 0.59 . . . . The estimated coefficients $beta_0$, $beta_1$, and $beta_2$ are shown for each of these seven individuals in the last four rows of Table 2B. Also shown are the several specifications of f(t). The R-squared indicates the percent of variation in weight explained by the model.

If we are to find the appropriate parameters for the curve designed to model the sinusoidal fluctuations, we must let the data guide us. The model summarizes the pertinent components of the process fairly well, and this modeling procedure will yield values for frequency and amplitude that should be useful in establishing whether cycling is associated with biological risk factors.

Working with richer models and adequate amounts of data should allow us to quantify biologically meaningful changes, separate the amplitude and the frequency, and assess each against known risk factors. Research in this area will enrich the development of the cycles approach to defining weight cyclers, as well as any measure that utilizes the threshold concept. Standardizing the interval and the threshold will bring us closer to a single, interpretable definition of weight cycling, which can be evaluated for associations with risk factors.

SUMMARY AND CONCLUSIONS

The basic goal of this paper has been to compare discrete and continuous measures of weight change, fluctuations, and patterns applied to retrospective and prospective measures of weight or other variables where their pattern of change is of interest. There are various criteria and purposes for which one might use such indices; there is no single best one for the broad array of aims to which these techniques must provide information. In the process, we have suggested a qualitative definition for weight cycling as well as a research path that could provide quantitative information to improve the definition.

Variability has been at the center of many analyses for some time. The initial approach was to utilize the standard measures from the literature (within-person standard deviation and coefficient of variation). For relevance to the literature, these have been computed and utilized for comparison but are deficient in describing the phenomena we are trying to address. As noted above, neither of these measures is sensitive to weight gain versus weight loss. Furthermore, we do not believe these effectively evaluate biological risks that may be associated with weight cycling because they misclassify many other patterns of weight change into the same group. The review of the literature on variability suggested several other measures, but most suffered from similar limitations.

To study the behavior of various indices, we examined patterns of responses and created various timed weighted measures of the values according to temporal schema. This does not eliminate the gain versus loss problem directly but rather facilitates the study of the impact of various patterns on the risk factors. Several modifications were attempted and a more integrated approach was set forth. The cycles measure is clinically interpretable and statistically tractable. It incorporates a definition of a cycle: a gain followed by a loss or a loss followed by a gain. The appropriate threshold and intervals for measuring weight are implicit in the definition and should be addressed explicitly using a time series or longitudinal approach.

Research should continue to establish or refute the existence of excess risks of weight fluctuations and to define the threshold of weight changes if such risks exist. The search for a biologically plausible mechanism should continue, as this may help the modelling and development of appropriate summary statistics. Given the importance of these findings, if indeed there is excess risk associated with fluctuations, a well-understood measure of the concept is essential. The concept of fluctuations is a simple one, understood by scientists as well as non-scientists, and thus demands that when we suggest there are risks associated with fluctuations, we must be sure fluctuations are clearly defined. The dangers of surrogate measurement or indirect correlates of weight gain, which often confound measures of fluctuation, could lead to greater impediments to weight management. Until better measures are devised, the number of cycles appears to be the simplest and most appropriate measure of weight cycling.

REFERENCES

(1) Blair SN: Long-term benefits and adverse effects of weight loss: Data from two prospective epidemiological studies. NIH Obesity Technology Conference. 1992.

(2) Sonne-Holm S, Sorensen TI, Jensen G, Schnohr P: Independent effects of weight change and attained body weight on prevalence

of arterial hypertension in obese and non-obese men. *British Medical Journal.* 1989, *299*(6702):767–770.

(3) Ashley FW, Kannel WB: Relation of weight change to changes in atherogenic traits: The Framingham Study. *Journal of Chronic Diseases.* 1974, *27:*103.

(4) National Task Force on the Prevention and Treatment of Obesity: Weight cycling. *Journal of the American Medical Association.* 1994, *272*(15):1196–1202.

(5) Archambault CM, Czyzewski D, Cordua y Cruz GD, Forety JP, Mariotto MJ: Effects of weight cycling in female rats. *Physiology and Behavior.* 1989, *46*(3):417–421.

(6) Bene CR, Klesges RC, Burks M, O'Neil PM: The impact of weight cycling on energy balance and response to treatment. *International Journal of Obesity.* 1991, *15*(S3):45.

(7) Brownell KD, Greenwood MRC, Stellar E, Shrager EE: The effects of repeated cycles of weight loss and regain in rats. *Physiology and Behavior.* 1986, *38:*459–464.

(8) Blackburn GL, Wilson GT, Kanders BS, et al: Weight cycling: The experience of human dieters. *American Journal of Clinical Nutrition.* 1989, *49*(S5):1105–1109.

(9) Blanchard FA, Frost RO: Two factors of restraint: Concern for dieting and weight fluctuation. *Behaviour Research and Therapy.* 1893, *21:*259.

(10) VanDale D, Saris WHM: Repetitive weight loss and weight regain: Effects on weight reduction, resting metabolic rate, and lipolytic activity before and after exercise and/or diet treatment. *American Journal of Clinical Nutrition.* 1989, *49*(3):409–416.

(11) Graham B, Chang S, Lin D, Yakubu F, Hill JO: Effect of weight cycling on susceptibility to dietary obesity. *American Journal of Physiology.* 1990, *259*(6, Pt. 2):R1096–1102.

(12) Gray DS, Fisler JS, Bray GA: Effects of repeated weight loss and regain on body composition in obese rats. *American Journal of Clinical Nutrition.* 1988, *47*(3):393–399.

(13) Hill JO, Thacker S, Newby D, Sykes MN, Digirolamo M: Influence of food restriction coupled with weight cycling on carcass energy restoration during ad-libitum refeeding. *International Journal of Obesity.* 1988, *12*(6):547–555.

(14) Holbrook TL, Barrett-Connor E, Wingard DL: The association of lifetime weight and weight control patterns with diabetes among men and women in an adult community. *International Journal of Obesity.* 1989, *13:*723–729.

(15) Jebb SA, Goldberg GR, Coward WA, Murgatroyd PR, Prentice AM: Effects of weight cycling caused by intermittent dieting on metabolic rate and body composition in obese women. *International Journal of Obesity.* 1991, *15*(5):367–374.

(16) Lissner L, Odell PM, D'Agostino RB, et al: Variability of body weight and health outcomes in the Framingham population. *New England Journal of Medicine.* 1991, *324*(26):1839–1844.

(17) Lissner L, Andres R, Muller DC, Shimokata H: Body weight variability in men: Metabolic rate, health, and longevity. *International Journal of Obesity.* 1990, *14*(4):373–383.

(18) Manore MM, Berry TE, Skinner JS, Carroll SS: Energy expenditure at rest and during exercise in non-obese female cyclical dieters and in non-dieting control subjects. *American Journal of Clinical Nutrition.* 1991, *54:*41–46.

(19) Reed DR, Contreras RJ, Maggio C, Greenwood MR, Rodin J: Weight cycling in female rats increases dietary fat selection and adiposity. *Physiology and Behavior.* 1988, *42*(4):389–395.

(20) Rodin J, Radke-Sharpe N, Rebuffe-Scrive M, Greenwood MRC: Weight cycling and fat distribution. *International Journal of Obesity.* 1990, *14*(4):303–310.

(21) Saris WHM: Physiological aspects of exercise in weight cycling. *American Journal of Clinical Nutrition.* 1989, *49:*1099.

(22) Schotte DE, Cohen E, Singh SP: Effects of weight cycling on metabolic control in male outpatients with non-insulin-dependent diabetes mellitus. *Health Psychology.* 1990, *9*(5):599–605.

(23) Williamson PS, Levy BT: Long-term body weight fluctuation in an overweight population. *International Journal of Obesity.* 1988, *12:*579–583.

(24) Kraemer HC, Berkowitz RI, Hammer LD: Methodological difficulties in studies of obesity. I. Measurement issues. *Annals of Behavioral Medicine.* 1990, *12*(3):112–118.

(25) Kraemer HC, Berkowitz RI, Hammer LD: Methodological difficulties in studies of obesity. II. Design and analysis issues. *Annals of Behavioral Medicine.* 1990, *12*(3):119–124.

(26) Kramer FM, Jeffery RW, Forster JL, Snell MK: Long-term follow-up of behavioral treatment for obesity: Patterns of weight regain among men and women. *International Journal of Obesity.* 1989, *13*(2):123–136.

(27) Mood AM, Alexander M, Graybill FA: *Introduction to the Theory of Statistics.* New York: McGraw-Hill, 1963.

(28) Siegel S: *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill, 1956.

(29) Liang KY, Zeger SL: *Longitudinal Data Analysis Using Generalized Linear Models.* New York: Chapman Hall, 1986.

(30) Zeger SL, Liang KY: Longitudinal data analysis for discrete and continuous outcomes. *Biometrics.* 1986, *42:*121–130.

(31) Dunlop DD: Regression for longitudinal data: A bridge from least squares regression. *American Statistician.* 1994, *48*(4):299–303.

(32) Harvey AC: *Forecasting, Structural Time Series, and the Kalman Filter.* Cambridge, England: Cambridge University Press, 1989.

(33) Jones RH: *Longitudinal Data with Serial Correlation: A State Space Approach.* New York: Chapman Hall, 1993.