International Journal of Obesity (2009) 1–4
 0307-0565/09 $32.00
www.nature.com/ijo

COMMENTARY

Prevention of excess gain

J Stevens[1,2], KP Truesdale[1], C-H Wang[1] and J Cai[3]

[1]*Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA;* [2]*Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA and* [3]*Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA*

Obesity prevention trials are designed to promote healthy weight. The success of these trials is often assessed using one of three metrics—means, incidence or prevalence. In this study, we point out conceptual shortcomings of these metrics and introduce an alternative that we call 'excess gain'. A mathematical demonstration using simulated data shows a scenario in which the statistical power of excess gain compares favorably with that of incidence and prevalence. Prevention of excess gain communicates an easily understood public health message that is applicable to all individuals regardless of weight status.
International Journal of Obesity advance online publication, 11 August 2009; doi:10.1038/ijo.2009.158

Keywords: methods; prevention; body weight; incidence; power

One of the many choices that have to be made by investigators testing an obesity prevention intervention is the selection of the primary outcome. This is an important decision regardless of study design, but takes on added importance when a randomized trial is planned because for a trial it is preferable to pre-specify a single primary outcome to preserve the validity of the probability associated with the test of effect. The size of the sample needed for the trial is dependent on the hypothesized effect size and the variance characteristics of the proposed primary outcome, as well as on the certainty specified. Although a large number of secondary outcomes may be studied in a randomized trial, it is the primary outcome that is the major focus.

For projects testing the impact of an intervention on the prevention of obesity, the outcome usually involves a measure of body mass or body fatness, such as BMI, BMI *z*-score or percent body fat. Feasibility constraints prevent large-scale studies from measuring fatness through precise measures such as DEXA, forcing investigators instead to estimate percent body fat using equations that include measurements such as height, weight, skin-fold thickness, bioelectric impedance and other measures that can be collected in the field. We have commented previously on strengths and weaknesses of the available measures in field trials of obesity interventions and recommended that investigators use a gender, age and race appropriate measure of percent body fat estimated from an equation that has an $R^2 \geqslant 0.8$ as assessed against a criterion method.[1] In that paper, we addressed the measurement to be used as primary outcome in an obesity prevention trial, but did not examine the mathematical expression or metric of the outcome.

The most prominent metrics used in obesity trials are means, incidence and (less often) prevalence. Below we examine the strengths and weaknesses of these currently popular metrics and then suggest a new metric: excess gain. This new metric is compared with the traditional metrics in an example scenario of a randomized controlled trial in which differences between a control and intervention group are tested at the end of the study period.

Strengths and weaknesses of means

Comparison of mean changes in a body measurement is a common outcome in obesity prevention trials. A strength of this approach is that, compared with the common alternatives, it yields a statistically powerful analysis. One practical weakness of the comparison of means in obesity prevention trials is that the mean change can appear small in magnitude, and might be interpreted as trivial, even when there is a substantial intervention effect. This limits the usefulness of the mean to convey messages to the public. Another limitation is that the mean does not discriminate between the changes that occur in participants with different

Correspondence: Dr J Stevens, Chair, Department of Nutrition, 245 Roseanu Hall, CB 7461, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27599, USA.
E-mail: June_Stevens@unc.edu
Received 18 June 2009; accepted 24 June 2009

body sizes. Obesity prevention interventions are not created with the intent that they will have an equivalent impact across the full range of body sizes that occur in the population. Investigators hypothesize that the interventions they test will promote healthy eating and/or activity patterns that will prevent or reduce obesity. Implicit in this hypothesis is the belief that participants who gain unhealthy weight are more likely to have unhealthy diet and activity patterns than those who do not. An intervention that prevents obesity would change those with unhealthy habits, and the expectation is that shifts in the BMI or % fat distribution will not be uniform, but rather focused on the right hand side of the distribution. In addition, a 5 percentage point change in percent body fat in an overweight versus an underweight child would be viewed very differently in terms of health impact. Nevertheless, when mean change is the outcome, these changes can contribute to the mean in an identical way. This raises the question of how well the mean summarizes the differences that are of most interest to an investigator studying obesity prevention.

Strengths and weaknesses of incidence and prevalence

Incidence and prevalence of obesity are heavily focused on quantification of movement around a single value, which is (almost always) located on the right hand side of the sample distribution. Usually, a single cut point is chosen, and the incidence and prevalence are determined based on whether an individual's body mass or body fat measurement is greater or less than the cut point value. Different from the mean, changes, regardless of magnitude, do not impact these estimates if they do not involve a change in status relative to the designated cut point. Although incidence and prevalence address the literal meaning of 'obesity prevention', they do not successfully capture the full intent of obesity intervention researchers. For example, if compared with control an intervention successfully prevented weight gain in children who were obese at baseline, this outcome would be very important, yet it would not be captured in a usual obesity incidence or prevalence analyses.

An alternative metric: excess gain

Given the weaknesses in comparisons of means, incidence and prevalence, we propose an alternative metric, which we call 'excess gain'. To be classified as a case of excess gain, a participant would need to meet the following two criteria at the end of a trial: (1) a body measurement (for example, BMI, BMI *z*-score and percent body fat) that is greater than the predesignated cut point (for example, a level consistent with obesity) and (2) a gain of $>3\%$ in the body measurement compared with baseline (in adults) or with the expected maintenance value (in youth). As growth and changes in body composition are expected in children, normal increases in size and fatness need to be considered in the calculation, whereas in adults this type of adjustment is usually not needed. The intention of the second criteria is to identify participants who have gained an amount more than the maintenance value in the selected body size measurement. The 3% minimum gain was chosen based on our previous work[2] in which we suggested that changes in body measurements of $< \pm 3\%$ are consistent with weight maintenance in adults. Changes smaller than 3% could be the result of measurement error, fluctuations in fluid balance or other diurnal variations, and are unlikely to be clinically important.

Table 1 and Figure 1 contrast excess gain with incidence and prevalence. Table 1 shows six weight gain scenarios, and an example of each of these scenarios is shown in the figure. Studies examining incidence exclude from the analysis scenarios E and F in which participants were obese at baseline, and classify as cases the participants fitting scenarios C and D. Studies examining prevalence would classify four of the scenarios (C, D, E and F) as cases. Note that even though C shows a change in BMI *z*-score that is within the bounds of weight maintenance, participants with this change would be classified as incident and prevalent cases. Large versus negligible changes in BMI *z*-score are not systematically treated differently for either incidence or prevalence. The excess gain identification of cases has the advantage that participants who just eek over the cut point with a very small change are not counted as cases. In addition, participants classified as obese at baseline remain in the analysis, and a substantial weight gain in an obese participant constitutes a case.

Table 1 Classification of incident, prevalent and excess gain cases for six weight change scenarios

Scenario	*Weight change scenario*			*Incidence case*	*Prevalent case*	*Excess gain case*
	Baseline weight status	*Follow-up weight status*	*Magnitude of weight change (%)*			
A	Not obese	Not obese	$\leqslant 3$	No	No	No
B	Not obese	Not obese	>3	No	No	No
C	Not obese	Obese	$\leqslant 3$	Yes	Yes	No
D	Not obese	Obese	>3	Yes	Yes	Yes
E	Obese	Obese	$\leqslant 3$	No[a]	Yes	No
F	Obese	Obese	>3	No[a]	Yes	Yes

[a]Excluded from the analysis.

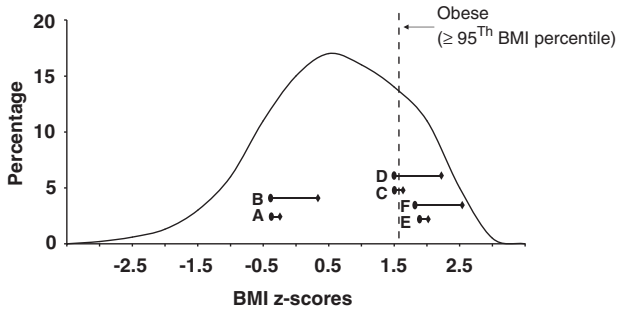

Figure 1 Six weight change scenarios (A–E) illustrated using the distribution of BMI *z*-scores in sixth grade girls in the TAAG study. Scenarios A, C and E indicate 3% changes and are not shown precisely to scale.

Illustration to compare metrics

As one form of evaluation of excess gain as an outcome, we analyzed the statistical power required to detect differences in excess gain compared with differences in the mean, prevalence and incidence. For this illustration, we constructed a fictitious example that mimicked what might be observed in actual data. As a basis for our calculations we used data from the Trial of Activity in Adolescent Girls (TAAG) study, a multicenter study of sixth and eighth grade girls. The primary hypothesis of the TAAG study was that an intervention that links schools and community partners will result in higher levels of moderate-to-vigorous physical activity in middle school girls.[3] In the TAAG longitudinal cohort, percent body fat was measured at sixth and eighth grades and the intervention did not impact the change in percent body fat.[4,5]

For this demonstration, the 'control group' was assigned the percent body fat levels of the control girls in sixth and eighth grade in TAAG ($n=582$). Data were simulated for the 'intervention group' that showed a reduction in mean percent body fat of the magnitude suggested in the meta-analysis by Katz *et al*.[6] (−0.29 standardized mean difference) of school-based interventions for obesity. In TAAG, the −0.29 standardized mean difference was equivalent to −2.7 change in percent body fat.

Part 1 of the excess gain definition requires a predesignated cut point chosen to indicate an unhealthy level. In this study, we chose 32% body fat.[7] For part 2 of the excess gain definition, we needed to account for the fact that it is normal for girls in the age range studied here to increase in percent body fat. We arbitrarily defined the expected maintenance level to be the median amount of gain in fatness among girls who were between the 40th and 60th percentiles of BMI *z*-score in both sixth and eighth grades. Thus, for this demonstration, we estimated the maintenance value in eighth grade to be an increase of 3.4 percentage points above the baseline percent body fat (sixth grade). Girls who increased in percent body fat by 3% more than the expected maintenance value were judged to have had a meaningful increase in body fatness and to meet criteria 2, as described above in the definition of excess gain.

For each simulation iteration, we selected 582 girls from the control data set randomly with replacement. Rather than assume that the mean change in percent body fat was the same in all girls, we arbitrarily set the intervention effect to be three times larger in girls with a baseline percent body fat that was above compared with that below the mean, and assumed that the intervention effect had a linear association with baseline percent body fat. As a girl could be selected more than once per iteration, we added a random amount of error to the eighth grade value. The error ranged from −2.5 to 2.0 percentage points, which is consistent with the amount of measurement error for percent body fat calculated from an equation derived from work in girls similar to those in TAAG.[8] The simulation was repeated 500 times and the average of the means, variances and percentages were used.

Table 2 shows the means and frequencies for control and 'intervention' girls for the metrics of interest. To calculate incidence, participants who were not 'at risk' must be excluded. In this study, approximately one-third of the sixth grade girls were over our designated cut point, and therefore were excluded from the incidence calculations. As would be expected, the prevalence was higher than the incidence in both the control and intervention groups, but the differences were similar at −11.2 and −10.2%. The percentage of girls with excess gain in the control and intervention groups was lower than the prevalence and incidence; however, the difference between the groups was similar to that seen for incidence and prevalence at −11.0%.

Given the values found in Table 2, all four metrics had over 90% power to detect significant differences ($P<0.05$) between the groups given a sample of 582 per group. If the sample was reduced to 150 girls per group (Table 3), there was 99.8% power to detect a difference in mean change between the groups and 88.2% power for excess gain. The power was much lower for incidence and prevalence. The mean change required the smallest study size for 90% power (63 participants per group), and was decidedly the most powerful of the outcomes examined. However, among the dichotomous metrics, excess gain provided the most power and required about half the number of subjects needed for incidence and prevalence for 90% power.

Conclusions

Obesity prevention trials compare changes in the mean, incidence or prevalence of the chosen outcome in the treatment and control arms, and these three metrics are supported by decades of precedence. In fact, the precedence of these metrics is so strong that their weaknesses are rarely fully considered. Nevertheless, they are not without flaws and do not completely capture the intent of investigators studying obesity interventions. In this work, we have suggested an alternative metric, excess gain, which addresses some of the shortcomings of the currently used metrics. In a

Table 2 Simulated results from an obesity prevention intervention in girls in sixth grade at baseline and eighth grade at follow-up

Metric	*Control*	*Intervention*	*Difference*
Sixth grade % body fat	27.9 (9.3)	28.0 (9.2)	0.1
Eighth grade % body fat	31.3 (8.4)	28.6 (7.2)	−2.7
Change in % body fat	3.4 (4.7)	0.7 (5.6)	−2.7
Prevalence of % body fat ⩾32.0%	45.4%	34.2%	−11.2%
Incidence of % body fat ⩾32.0%	22.8%	12.6%	−10.2%
Excess gain	18.7%	7.7%	−11.0%

Table 3 Estimated power given a sample size of 150 girls per group and the estimated number of girls (per group) needed for 90% power for four metrics

Metric	*Estimated power given 150 girls per group (%)*	*Estimated study size for 90% power*
Change in % body fat	99.8	63
Prevalence of % body fat ⩾32.0%	63.3	325
Incidence of % body fat ⩾32.0%[a]	59.8	357
Excess gain	88.2	160

[a]Based on data from TAAG, two-thirds of the girls would be eligible for the incidence analysis.

study using BMI and standard obesity cut points, excess gain would prevent identifying as cases participants who actually maintained their BMI and instead include as cases obese individuals who gained weight. These attributes seem in keeping with the intent of trials that promote healthy weight. Further, prevention of excess gain communicates an easily understood public health message that is applicable to both normal weight and obese individuals.

The most important attribute of a study endpoint is not its statistical power, but its ability to capture the conceptual intent of the investigator. For instance, incidence is often preferred over prevalence because of its capacity to indicate change regardless of any differences in power. Nevertheless, a comparison of the statistical power associated with different outcomes is of practical interest because it impacts study size and therefore study costs. In one example in which we compared the statistical power of several metrics, excess gain performed surprisingly well. However, we are acutely aware of the shortcomings of the example presented here and readily admit that the results of this exercise were entirely dependent on the assumptions that were made. Nevertheless, this work has inspired us to recommend that future work be conducted to compare the performance of excess gain with conventional metrics using actual data from obesity prevention trials. An extension of the excess gain concept would be to compare the mean amounts of excess gain rather than to examine the counts of those with excess gain. Other metrics could also be proposed and studied. We urge investigators to carefully consider the weaknesses of the currently used metrics in obesity prevention research and remain open to new options.

Conflict of interest

Dr Stevens is the recipient of a Distinguished Professorship awarded by the American Institute for Cancer Research. She has led or been a co-investigator on research projects funded by the NIH, the Centers for Disease Control, the American Heart Association, Nestle Waters, Sanofi-Aventis and Gatorade. Dr Truesdale has received funding from NIH and Sanofi-Aventis. Dr Cai has received funding from NIH, Lance Armstrong foundation, CDC and the American Heart Association. Dr Wang declares no potential conflict of interest (NIH funding only).

References

1 Stevens J, Taber DR, Murray DM, Ward DS. Advances and controversies in the design of obesity prevention trials. *Obesity* 2007; **15**: 2163–2170.
2 Stevens J, Truesdale KP, McClain JE, Cai J. The definition of weight maintenance. *Int J Obes (Lond)* 2006; **30**: 391–399.
3 Stevens J, Murray DM, Catellier DJ, Hannan PJ, Lytle LA, Elder JP *et al*. Design of the Trial of Activity in Adolescent Girls (TAAG). *Contemp Clin Trials* 2005; **26**: 223–233.
4 Webber LS, Catellier DJ, Lytle LA, Murray DM, Pratt CA, Young DR *et al*. Promoting physical activity in middle school girls: Trial of Activity for Adolescent Girls. *Am J Prev Med* 2008; **34**: 173–184.
5 Stevens J, Murray DM, Baggett CD, Elder JP, Lohman TG, Lytle LA *et al*. Objectively assessed associations between physical activity and body composition in middle-school girls: the Trial of Activity for Adolescent Girls. *Am J Epidemiol* 2007; **166**: 1298–1305.
6 Katz DL, O'Connell M, Njike VY, Yeh MC, Nawaz H. Strategies for the prevention and control of obesity in the school setting: systematic review and meta-analysis. *Int J Obes (Lond)* 2008; **32**: 1780–1789.
7 Williams DP, Going SB, Lohman TG, Harsha DW, Srinivasan SR, Webber LS *et al*. Body fatness and risk for elevated blood pressure, total cholesterol, and serum lipoprotein ratios in children and adolescents. *Am J Public Health* 1992; **82**: 358–363.
8 Going S, Nichols J, Loftin M, Stewart D, Lohman T, Tuuri G *et al*. Validation of bioelectrical impedance analysis (BIA) for estimation of body composition in Black, White and Hispanic adolescent girls. *Int J Body Compos Res* 2006; **4**: 161–167.