The immediate challenge arising from endemically high rates of childhood obesity is to develop, test, and implement effective treatments that reduce its negative consequences on future health and costs. However, there remain no ready “fixes,” as most trials report at most small BMI benefits despite substantial intervention input.

Given the opportunity costs of ineffective interventions, public health providers should rightly query value for money if interventions deliver change that barely exceeds natural history and/or is insufficient for health benefit. However, such judgments require knowledge of the natural history of BMI change in community samples of overweight and obese children participating in research studies. Unfortunately, regression to the mean and biases favoring resolution such as the Hawthorne effect seem no less likely in obesity than with any other chronic disease.

For example, the most recent Cochrane systematic review of childhood obesity treatment included lifestyle trials comparing 2 or more interventions head-to-head, as well as trials comparing interventions with true controls. The former generally concluded that both interventions were equally effective, on the basis of mean reductions occurring within both intervention groups. In contrast, those studies with true controls tended to reveal that the interventions were ineffective, even though average BMI reductions were similar to those observed in the comparative trials. Thus, the meta-analysis of behavioral interventions in trials with true controls revealed only a small favorable between-group effect of 0.06 in BMI z score (95% confidence interval [CI]: −0.12 to −0.01). To define intervention success, one should also understand what BMI change equates to improved health. Although this has no definitive answer, and to some extent depends on the baseline value, it is suggested that BMI z score (considered by some a better marker of fat loss than change in BMI, weight, or weight z score) should fall by at least 0.5 to 0.6 to be confident of reductions in fat and cardiovascular risk in obese children. Because early effects (eg, at 6 months) in many trials are no longer evident by 12 months, many also consider that BMI improvement should be sustained for at least 12 months from baseline for the intervention to be considered effective.

To illustrate the dangers of interpreting group changes in BMI without a control group, we summarize BMI z score changes in community samples.
of overweight and/or obese children in 4 Australian studies, comprising Australia’s largest observational children’s cohort study and our own 3 recent randomized trials. Specifically, we report the change in BMI z scores both continuously and categorically (percentage revealing any, ≥0.25, and ≥0.5 reductions in BMI z score) over a 15-month period from baseline, to ensure 12 full months of follow-up. We hypothesized that children in both study types would show reductions in mean BMI z score in keeping with regression to the mean, with larger reductions in trials due to additional biases (eg, Hawthorne effect, selection bias). All were approved by human ethics committees, and parents provided written informed consent.

We drew on person-level data for the overweight/obese participants in waves 1 to 4 of the kindergarten cohort of the nationally representative Longitudinal Study of Australian Children (LSAC), a broadly focused observational study that recruited 4983 4- to 5-year-olds in 2004 then remeasured their height and weight every 2 years. We also pooled person-level data from our 3 randomized trials (LEAP 1, 2002; LEAP 2, 2005/2006; HopSCOTCH, 2009–2011) of lifestyle interventions nested within BMI surveillance studies in family practice; all had null results, with minimal evidence of differences in average BMI z score reduction between treatment and control groups (Fig 1, top grouping, LEAP 1, LEAP 2, and HopSCOTCH) were approved by The Royal Children’s Hospital Human Ethics Committee (21019, 25006, and 280178) and LSAC by the Australian Institute of Family Studies Ethics committee.

All participants were aged 5 to 10 years, and overweight or obese by International Obesity Task Force cut points at baseline (LSAC, n = 1816; LEAP 1, n = 146; LEAP 2, n = 242; HopSCOTCH, n = 77; retention, 89%-94%). In all studies, children were weighed at baseline and follow-up in light clothing to the nearest 50 g on digital scales and measured to the nearest 0.1 cm with a rigid stadiometer. Follow-up occurred ~24 and 15 months later in LSAC and the trials, respectively. BMI was age- and gender-standardized by using Centers for Disease Control and Prevention reference values. Change in BMI z score was rescaled to the expected change had each child been followed for 15 months. We deemed z score changes of more than 2 SD implausible, resulting in exclusion of 2 children. For the observational data, change in BMI z score was calculated where any 2 consecutive waves of person-level data were available. Trial data were analyzed as a single pooled data set. Analyses used Stata 12.1 (Stata Corp, College Station, TX).

The overweight and obese children in the observational and trial samples had similar baseline BMI z scores (Table 1). In all 6 continuous analyses (overweight, obese, and overweight/obese groups for the observational and intervention studies separately), mean BMI z score reduced slightly over time. Mean changes were smaller in the observational (−0.05 to −0.10) than the intervention (−0.12 to −0.14) studies, but individuals varied markedly. In the observational study, ~62% showed any reduction in BMI z score, whereas 16% fell by at least 0.25 and 4% by at least 0.5. The equivalent figures for the intervention trials were 68% (any reduction), 30% (≥0.25), and 11% (≥0.5). Strengths of our analyses include the large harmonized samples of overweight and obese children. Factors mitigating against participation bias in LSAC include its breadth and absence of obesity focus. However, we acknowledge lack of body composition data regarding actual adiposity, and were not able to consider other important health outcomes such as fitness and self-esteem.

Thus, in both the intervention and observational studies, there was invariably a reduction in mean BMI. This would be expected by regression to the mean, given selection on the basis of high BMI, with additional factors such as the Hawthorne effect likely influencing the slightly larger reductions in the trials. These largely artifactual changes in BMI should be expected but appear often to be overlooked. These data raise interesting questions about the efficacy reported by current tertiary obesity clinics. For example, a pooled sample from 175 German-speaking tertiary pediatric obesity clinics achieved a mean reduction of −0.15 BMI z score after a median of 1.2 years (interquartile range, 0.9–2.2; see lower part of Fig 1). The natural tendency for high BMI z scores to fall somewhat could even be heightened in tertiary clinics because one might expect (and our analyses confirm) larger natural improvements in treatment than...
observational studies, and in obese than overweight children.

We conclude that accounting for the natural tendency for BMI in groups selected on the basis of overweight to fall a little could prevent implementation of interventions that could never hope to deflect the course of the obesity epidemic. Although unlikely to be harmful to individuals, scaling even low-intensity interventions to the population is not only costly but may subsequently impede more effective interventions because of competing workforce demands and/or the challenges of decommissioning services that may erroneously be perceived as beneficial. These are not trivial issues.

Three recommendations arise. Firstly, it seems not only ethical but necessary to adequately control community-based intervention trials. We have not found this to be particularly costly, or to be unpalatable to participants. In the absence of controlled comparisons, we suggest researchers and policy makers should use other means to estimate attributable fractions and actual health gain over and above that expected without intervention.

Secondly, for services where traditional controlled trials may be impossible, designs could be used such as multiple baselines (providing a good mechanism for assessing the extent of regression to the mean) and stepped wedge (incorporating controls into service roll-out). If clinical services could be resourced to conduct continuous networked research, then incremental treatment improvements could be measured over long time horizons (an exceptionally successful approach in other conditions such as childhood cancer). Finally, a proportion of obese children in every study naturally achieve substantial improvement that would be expected to confer important health benefit; research could focus on how body set-point change is achieved in this unusual subgroup.

ACKNOWLEDGMENTS

We thank all the parents and children who took part in all 4 studies.

REFERENCES


FINANCIAL DISCLOSURE: Dr Wake was supported by National Health and Medical Research Council (NHMRC) Population Health Career Development award 546405 and Senior Research Fellowship 1046518; Dr Sabin was supported by NHMRC Postdoctoral Training Fellowship 1012201; the other authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: LEAP1 was funded by a grant from the Australian Health Ministers’ Advisory Council for Primary Driven Research (AHMAC PDR 2991/15), LEAP2 by NHMRC project grant 334309, and HopSCOTCH by NHMRC project grant 491212. This article uses unit record data from Growing Up in Australia: The Longitudinal Study of Australian Children, which is conducted in partnership between the Department of Social Services, the Australian Institute of Family Studies, and the Australian Bureau of Statistics. Research at the Murdoch Childrens Research Institute is supported by the Victorian Government’s Operational Infrastructure Support Program. The findings and views reported are those of the authors and should not be attributed to the Department of Social Services, the Australian Institute of Family Studies, or the Australian Bureau of Statistics. Funding bodies had no influence on the design and conduct of the study, collection, management, analysis, and interpretation of the data, nor the preparation, review, or approval of the manuscript.

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

TABLE 1 Person-Level BMI Z Score Change Over 15 Months by Baseline BMI Status in Our Intervention and Observational Studies

<table>
<thead>
<tr>
<th>Intervention trials</th>
<th>Mean (SD) Baseline BMI</th>
<th>Z Score Mean (SD)</th>
<th>Range</th>
<th>Percent With BMI Z Score Reduction (95% CI)</th>
<th>Any</th>
<th>≥ 0.25</th>
<th>≥ 0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overweight (n = 314)</td>
<td>19.62 (1.53) 1.54 (0.25)</td>
<td>−0.12 (0.31)</td>
<td>−1.24 to 0.71</td>
<td>64.7 (58.3 to 70.0) 32.2 (27.0 to 37.4)</td>
<td>11.8 (8.2 to 15.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obese (n = 151)</td>
<td>23.23 (2.58) 2.14 (0.21)</td>
<td>−0.14 (0.24)</td>
<td>−1.12 to 0.49</td>
<td>73.5 (66.4 to 80.6) 24.5 (17.8 to 31.4)</td>
<td>8.0 (3.6 to 12.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overweight and obese (n = 465)</td>
<td>20.79 (2.57) 1.73 (0.37)</td>
<td>−0.13 (0.29)</td>
<td>−1.24 to 0.71</td>
<td>67.5 (63.3 to 71.8) 29.7 (25.5 to 33.8)</td>
<td>10.5 (7.7 to 13.3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observational study</th>
<th>Mean (SD) Baseline BMI</th>
<th>Z Score Mean (SD)</th>
<th>Range</th>
<th>Percent With BMI Z Score Reduction (95% CI)</th>
<th>Any</th>
<th>≥ 0.25</th>
<th>≥ 0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overweight (n = 1318)</td>
<td>19.49 (1.27) 1.48 (0.24)</td>
<td>−0.05 (0.26)</td>
<td>−1.97 to 1.15</td>
<td>57.2 (54.5 to 59.9) 16.7 (14.7 to 18.7)</td>
<td>3.9 (2.8 to 4.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obese (n = 496)</td>
<td>25.58 (2.44) 2.25 (0.28)</td>
<td>−0.10 (0.21)</td>
<td>−1.78 to 0.58</td>
<td>75.8 (72.0 to 79.6) 13.1 (10.1 to 16.1)</td>
<td>4.2 (2.5 to 6.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overweight and obese (n = 1814)</td>
<td>20.61 (2.48) 1.89 (0.42)</td>
<td>−0.07 (0.25)</td>
<td>−1.97 to 1.15</td>
<td>62.3 (60.1 to 64.5) 15.7 (14.0 to 17.4)</td>
<td>4.0 (3.1 to 4.9)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BMI z score change and percent reduction values scaled to the expected change over 15 mo. Outliers (n = 2) with change >2 SDs were excluded.


Natural BMI Reductions and Overestimation of Obesity Trial Effectiveness
Melissa Wake, Susan Clifford, Kate Lycett, Kim Jachno, Matthew A. Sabin, Sarah Baldwin and John Carlin

Pediatrics 2015;135;e292; originally published online January 12, 2015;
DOI: 10.1542/peds.2014-1832
Natural BMI Reductions and Overestimation of Obesity Trial Effectiveness
Melissa Wake, Susan Clifford, Kate Lycett, Kim Jachno, Matthew A. Sabin, Sarah Baldwin and John Carlin
Pediatrics 2015;135:e292; originally published online January 12, 2015;
DOI: 10.1542/peds.2014-1832

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/135/2/e292.full.html