The impact of random individual differences in weight change on the measurable objectives of lifestyle weight management services

Running title: Weight change objectives

Greg Atkinson and Alan M. Batterham

Health and Social Care Institute, School of Health and Social Care, Teesside University, Middlesbrough, UK

Corresponding author:

Greg Atkinson (Professor of Health Sciences and Biostatistics Research)

Health and Social Care Institute

Constantine Building

Teesside University

MIDDLESBROUGH

Tees Valley TS1 3BA

UK

Email: greg.atkinson@tees.ac.uk

Tel: +44 (0) 1642 342758

Compliance with Ethical Standards

GA had no financial support and has no conflicts of interest on this paper. AMB had no financial support and no conflicts of interest. This paper was prepared in accordance with the ethics and research governance regulations at Teesside University.

Authorship: Greg Atkinson formulated the idea for this article after reading the Department of Health and NICE guidelines. He was the main writer of the article. Alan M Batterham helped develop the ideas, peer-reviewed drafts of the article and revised it critically for important intellectual content. He was the main writer of the section on coping with missing data.

Abbreviations:

LWMS – weight management services

BMI – Body mass index

SD – Standard deviation

DoH – Department of Health

NICE – National Institute for Health and Care Excellence

BOCF - Baseline Observation Carried Forward

ANCOVA – Analysis of covariance

UK – United Kingdom

Abstract

Obese adults and children can be referred to lifestyle weight management services (LWMS) in which physical activity and/or dietary advice is delivered. Service providers quantify the “weight” change between an initial measurement and follow-up measurement(s), which could be 12-24 months later. A control group is usually absent. We aimed to scrutinise the various LWMS objectives for this weight change that are recommended by UK authorities.

UK guidelines recommend that an adult LWMS should (A) reduce the sample mean body mass of all enrolled adults by at least 3%, and/or (B) reduce the body mass of at least 30% of adults by at least 5%. We highlight the potential for objective B to be met even if no LWMS is implemented, especially over the recommended follow-up periods of 12-24 months. This is due to unavoidable random within-participant fluctuations in weight over such periods of time. A ≥1-kg reduction in mean body mass is also to be expected, even without any LWMS. Therefore, we suggest that objectives A and B are too liberal.

Obesity status in children is indicated by the body mass index (BMI) z-score. Nevertheless, another UK recommendation is for a LWMS to “maintain” or “reduce” the BMI z-score of 80% of the enrolled children. Besides there being no stated thresholds for “maintain” and “reduce”, it is inconceivable to deem a LWMS successful even if 80% of children do not alter their obesity status and even if the remaining 20% of children actually increase their obesity status. Here, we think the BMI z-score has been confused with the body mass z-score.

In conclusion, measurable objectives of UK-based LWMS need to be clarified, and possibly altered to account for typical amounts of random variability in individual weight measurements over the service time period.

Keypoints:

Some objectives for UK lifestyle weight management services (LWMS) focus on individual differences in weight change, but may be compromised by typical random variability in this weight change.

The UK Department of Health (DoH) objective that at least 30% of adults in an LWMS show at least a 5% reduction in weight may be too liberal. We suggest increasing this proportion of people to at least 40%

The UK DoH objective that a LWMS "maintains or reduces" the body mass index (BMI) z score of 80% of children is opaque, and potentially even more liberal than the above adult objective. We suggest that "maintain" and "reduce" are operationally defined, and that the BMI z score should in fact be a body mass z score for the "80% of children" threshold to be robust

Authors of published evaluations of LWMS should consider carefully issues like the relevancy of statistical inference, approaches for dealing with missing data, subgroup comparisons of weight loss and the inherent limitations of an absent comparator group.

Introduction

Generally, a lifestyle weight management service (LWMS) involves the delivery of an intervention that has been designed to reduce body weight and/or prevent weight gain [1]. Interventions focus on physical activity, diet, or both in combination, and can be delivered to individuals on a one-to-one basis or in groups, as well as in clinical or community settings. A distinction has been made between these lifestyle-related services and medical interventions, such as surgery [1]. Lifestyle weight management services are referred to in the UK as “tier 2” and focus more on human behaviour change rather than environmental changes. The majority of tier 2 services are delivered in community settings or a leisure centre, and to a lesser extent in hospitals and general practitioner (general physician) clinics [2].

In the UK, overweight and obese adults can be referred by the general medical practitioner to a LWMS if they have body mass indices (BMI) of >25 kg/m^2 and >30 kg/m^2, respectively [3]. Co-morbidities such as diabetes and hypertension are also considered in this referral process. In England, durations of LWMS range from six weeks to 20 months [2], although it is recommended that the effectiveness of all LWMS should be examined at 12 months or more [3]. In one mapping report, 59% of tier 2 services in England were found to adhere to this recommendation [2]. People have their weight measured at baseline and then at one or more follow-up time-points. A primary follow-up time-point is selected and the change in weight relative to a baseline measurement is calculated. It is unusual for LWMS providers to recruit people for a comparator sample who do not receive the service [4, 5]. Therefore, evaluations of services are commonly uncontrolled before-and-after studies (case series).

Although “weight” is actually a force calculated from the product of the acceleration due to gravity and mass (measured in N), in public health contexts “weight” generally encompasses measurements of body mass (kg) and height-normalised body mass index (kg/m^2). For children, these variables can be further expressed as z scores, which represent how many standard deviations a certain child's body mass (or body mass index) is above, or below, age- and sex-matched population reference values. These z scores can also be used to estimate a child's weight centile, e.g. 91st [2].

What are the measurable objectives for a successful service?

The UK's Department of Health (DoH) and National Institute for Health and Care Excellence (NICE) have published guidance for the design and implementation of LWMS [3, 6]. For example, the DoH has published best practice guidance for "tier-2" adult LWMS, and recommended that a service should be designed to meet the following objectives [6], which we refer to as adult objectives A and B:

> **Adult Objective A:** Adults who have attended at least one session of the lifestyle/diet-related intervention achieve a group mean weight loss of at least 3% of their initial weight, at the end of the intervention.
>
> **Adult Objective B:** At least 30% of all adults in the LWMS have achieved a weight loss equal to or greater than 5% of their initial weight at the end of the intervention.

Although the referral process to an adult LWMS is based on BMI [2], it can be seen that objectives A and B refer to changes in body mass (expressed in kg) and then expressed as a percentage of initial body mass. This is because initial obesity status is measured relative to population norms, taking into account the height of a person, whereas an adult LWMS is concerned with change in weight over time within particular adults, whose height does not alter over that same time period. Presumably, expression of weight change as a percentage of initial value is an attempt to normalise any weight loss for inter-individual differences in that initial value; an approach we comment further upon later in this paper.

The above objectives are present in the NICE guidelines for adult LWMS, but are combined within the same sentence to read that successful services, "*are likely to lead to an average weight loss of at least 3%, with at least 30% of participants losing at least 5% of their initial weight*" [3]. The word "*with*" suggests that objectives A and B should *both* be met. Nevertheless, there is evidence from other guidance documents, as well as from published service evaluations, that a joint appraisal of the both objectives is not necessarily made. For example, in the National Obesity Observatory's standard evaluation framework for adult weight management interventions, quantifiable objectives for "success" include the percentage of participants who have reduced their weight (objective B), but there is apparently no mention of a sample mean weight reduction [1]. In the abstract of a recently-published evaluation of a large National Health Service LWMS, Logue et al. [4] also focussed on objective B rather than objective A. Therefore it is apparent that objective B can be considered as an indicator of "success", independently from other measurable objectives, rendering the scrutiny of its robustness particularly important.

Is objective B really suitable?

Objective B is based on the individual differences in the change between a baseline and follow-up measurement; a topic we have reviewed recently [7]. We have been unsuccessful in locating the exact origins of the "30% proportion" component of objective B. The exact rationale for this threshold does not seem to be provided in the relevant guidelines, nor the relevant supporting information and evidence statements [1, 3, 6]. Importantly, there will always be adults who show reductions in weight over a period of 3-24 months, even if nothing is prescribed to them, just as there will always be adults who gain weight [7]. We suspect that this unavoidable amount of random within-participant fluctuation in weight measurements, obtained sometimes months apart, has not been considered in full when objective B was arrived at.

How easy is it to meet objective B?

Assuming that the distribution of percentage changes in weight is approximately Gaussian (Normal), we can derive the proportion of adults in any LWMS sample that is expected to show reductions in weight ≥ 5%. The main factors affecting this expected proportion of adults are the group mean percentage change in weight (i.e., objective A) and the standard deviation of these percentage changes, which is essentially the indicator of how large within-participant fluctuation in weight is (Figure 1). We stress that this SD of weight change is not necessarily dependent on the short-term test-retest reproducibility of weight measurements [7]. Rather, this SD represents the random within-participant variation in weight over the many months between initial and final measurements of weight.

It can be seen in Figures 1 and 2 that it is possible for ≥ 30% of a group of people to show a weight reduction ≥ 5%, even if the mean weight change for the group is zero. This situation is equivalent to no LWMS being implemented at all to the obese people. It naturally becomes easier to realise objective B, as the group mean reduction in weight increases towards fulfilment of objective A (≥3% mean loss). We also highlight that the sample mean change in weight is not just influenced solely by any systematic effects of the LWMS. In an important recent study, Johns *et al.* [8] concluded that evaluators of weight loss programmes should expect that, even in the absence of an intervention, the mean weight of their participants would be up to a kilogram less 1-year after their baseline measurements. These authors also reported that the variation in control group weight loss was marked between studies, with 2-3 kg of mean weight loss being observed at times without any weight loss programme being implemented at all. If this mean change is not accounted for, both objectives A and B would be more readily realised.

It also becomes easier to realise objective B, as the SD of percentage weight changes, i.e. the magnitude of inter-individual random fluctuation in weight, increases (Figure 1). If this SD is 10% or more, then objective B may be met even if the group mean change is 0%. We suspect that an SD of this magnitude is feasible, especially with the longer-term follow-up time-points that are considered the ideal for a LWMS [3]. For example, Jennings et al. [9] evaluated a LWMS for adults with morbid obesity, or obesity and comorbidities. Standard deviations of percentage changes in weight were reported to be 8-9% at follow-up times of 18-24 months. Similarly, for the UK's Counterweight programme, the mean weight change at 24 months was reported in the abstract to be -2.3 kg (95% confidence limits: -3.2 to -1.4 kg) [5]. The SD of change can be estimated from this confidence interval of 1.8 kg [10]. For the stated sample size of 357, the SD of change is approximately 9 kg. The mean initial weight of all the people enrolled in the programme was 101 kg so this SD translates to an SD of percentage change of approximately 9% of initial weight. All this raises the possibility that objective B can be met when, in reality, the LWMS had no effect on weight change over and above normal expected random within-participant variation over the same time period. Therefore, there would be concerns that the amount of random within-participant variation, between baseline and follow-up measurements made two years apart, is high enough for objective B to be realised even if no weight management programme is prescribed at all.

Recommended objectives for a children's LWMS

Children with a BMI z score corresponding to ≥91st centile, can be referred to a LWMS [2]. The UK DoH "best practice" guidelines also include measurable objectives for a children's LWMS [6]. While there does not appear to be a target involving a change in the sample mean weight (similar to the objective A set for adults) in the DoH guidelines, it is recommended that "*80% of children completing the programme [the LWMS] maintain or reduce their BMI (body mass index) z-score*" [6]. As mentioned earlier, body mass index normalises body mass to height-squared. A sex and age-specific z-score is then calculated to quantify the child's population BMI percentile, which is the metric used to diagnose childhood obesity in the first place. The greater the BMI z-score, the greater is the child's obesity status relative to age and sex-matched population norms. According to this particular DoH guideline, a LWMS would be deemed successful even if the BMI z-score (i.e., the obesity status itself) of 80% of the children was "maintained" and even if the BMI z-score (i.e. the obesity status itself) of the remaining 20% of children actually increased, which seems too liberal to inform a decision on the effectiveness of any childhood obesity management.

We maintain that body mass (or body mass z-score) rather than BMI z-score should have been referred to in the DoH guideline. Body mass z scores are certainly the focus in other relevant guidelines, such as those from NICE [11]. According to the DoH guidelines, a LWMS would, in theory, be successful if the body mass per se of 80% of the obese children was "maintained or reduced" during a normal growth period in which their height was also increasing. In this situation the BMI z-score (obesity status) for 80% of children would logically be reduced. It would be preferable for the DoH and NICE guidelines to be consistent, and for the "maintain or reduce" component to be operationally defined.

Other issues

Analysis of change based on individual percentages is not optimal [12]. The act of calculating percentage changes can itself lead to skewed data, even if the baseline, follow-up and changes in weight (expressed in kg) are all normally distributed [12]. We would encourage all LWMS evaluators to explore the distribution of the percentage change data they are analysing. Like any ratio statistic, the use of a percentage change is associated with the assumption that any weight change is directly proportional to initial weight status, which might not be upheld. In essence, a percentage change may be under- or over-representing weight change for different people with different initial weight status [12].

Evaluators of LWMS probably need to consider carefully whether statistical inference is actually relevant to their analyses. If the goal is to compare the outcomes of a specific LWMS to the governmental objectives, then inference (from sample to population) may not be relevant. Nevertheless, many published evaluations of LWMS *do* include indicators of inference like standard errors, confidence limits and P-values [5]. If statistical inference is deemed relevant, then a confidence interval can be reported for the mean change in weight. Moreover, a confidence interval can also be reported for the proportion of the sample who has met a certain weight reduction threshold, e.g., adult objective B or the “80% proportion” for children. For example, if 30% of a sample of 50 adults attending a LWMS were found to reduce their weight by at least 5%, then the confidence interval for this proportion of people would be approximately 17% to 43%. This confidence interval would be approximately 24% to 36% for a sample size of 200 people. There is a useful confidence interval calculator for proportions at: http://www.surveysystem.com/sscalc.htm

The methods for dealing with missing data in evaluations of LWMS recommended by the Department of Health and NICE are also not optimum [13]. A common approach is to impute the missing follow-up weight for participants who withdraw after baseline using “Baseline Observation Carried Forward” (BOCF), a variant of Last Observation Carried Forward in designs with a single follow-up time-point. This method assumes that the missing measurement is equal to the weight measurement made at baseline. Nevertheless, the BOCF is not a principled method for dealing with missing data, and generally leads to biased and speciously precise estimates even when the data are missing at random. A perhaps more reasonable assumption is that participants withdrawing from a LWMS are non-responders or negative responders to the intervention, such that the data are not missing at random. Evaluators of LWMSs should adopt a principled approach to dealing with missing data, after careful consideration of potential missingness mechanisms. In simple designs with a single baseline and follow-up measure an analysis of complete cases only is a more valid approach under a missing at random assumption than single value imputation methods. Sensitivity to departures from the missing at random assumption can be explored using multiple imputation methods [14].

Some authors of service evaluations have focussed on sub-group comparisons of weight changes [4, 15]. For example, weight changes over 12 weeks were compared between men and women who attended a behaviour change programme supported with motivational interviewing [15]. The authors concluded that men lost more weight than women in response to the programme. Nevertheless, the mean baseline weight for the men was 23.5 kg more than that for the women, and this difference was not adjusted for in the analysis of change with an appropriate analysis of covariance (ANCOVA) model [12].

What are the solutions?

Ultimately, individual differences in the response to a LWMS intervention should be quantified while taking into account the random between-participant variation over a similar time period. This information can be derived from an appropriate comparator group. Nevertheless, it is unusual for service providers to have been able to recruit comparator participants, so it is likely that objective B needs changing to account properly for random within-participant changes in weight.

If the UK DoH objective B is changed to be consistent with the characteristics of objective A (mean reduction ≥ 3%), then the proportion of cases showing a reduction ≥ 5% should be approximately 40%, rather than 30% for typical magnitudes of SD of change (Figure 1). Therefore, it seems sensible to change the proportion of cases threshold in objective B from 30% to at least 40%. Nevertheless, there may need to be further consideration if the distribution of percentage changes is skewed. It should also be borne in mind that at least 1kg of the change in sample mean weight might be due to factors other than the intervention itself [8], so it could be that objective A is also too liberal in an uncontrolled evaluation of a LWMS.

The recommended UK DoH objective for a children's LWMS needs reviewing as soon as possible. It is inconceivable for a LWMS to be deemed successful just because 80% of the children do not alter their BMI z-scores, even though 20% of the children could have increased their BMI z-scores. Besides clarifying whether it is actually weight z-score that should have been referred to in this guideline, as well as rationalising the 80% threshold, the definitions of "maintain" and "reduce" also need to be provided to make this particular guideline useful to evaluators.

It seems unclear how the various proportions of people, e.g. 30%, 80%, as well as the threshold weight reductions, e.g. 5%, 3%, have been arrived at. It is possible that the thresholds for percentage weight reduction have been derived using information from randomised controlled trials of interventions. Even if so, it is also unclear whether the thresholds have actually been formulated after accounting for any "natural" typical change in the control group. A clinically-relevant control-group-adjusted mean weight change may have been arrived at by subtracting the control group mean change from intervention group mean change observed in a randomised controlled trial, e.g. 6% reduction in intervention arm - 1%

reduction in control group arm = 5% change. However, as mentioned above, evaluations of LWMS tend not to involve control groups. Assuming the initial mean body mass of our hypothetical sample was 100 kg, and in order to account for a "natural" mean reduction of 1 kg (1%) even if no LWMS is completed [8], the threshold for LWMS evaluation should, in fact, be 6% rather than 5%. A similar adjustment should be made for the proportion of people thresholds, e.g. 30% of people showing a ≥ 5% reduction in weight. We showed this above by projecting what the proportion of people would be expected to show such a weight reduction in a hypothetical control group who do not receive any LWMS. In conclusion, these proportions and thresholds need to be more clearly-rationalised and the objectives should be written in a less opaque manner, with proper definitions being made for labels such as "maintain" and "reduce".

References

1. Roberts KC, Cavill N, Rutter H. (2009). *Standard Evaluation Framework for weight management interventions.* Oxford: National Obesity Observatory.
2. Public Health England (2015). National Mapping of Weight Management Services. Provision of teir 2 and teir 3 services in England. PHE Publications gateway number 2015521.
3. National Institute for Health and Care Excellence (NICE). (2014). *Weight management: lifestyle services for overweight or obese adults.* Public Health Guideline, http://www.nice.org.uk/guidance/ph53 Published 28th May, 2014.
4. Logue J, Allardice G, Gillies M, et al. (2014). Outcomes of a specialist weight management programme in the UK National Health Service: prospective study of 1838 patients. *BMJ Open* 4, e003747.
5. Counterweight Project Team. (2008) Evaluation of the Counterweight Programme for obesity management in primary care: a starting point for continuous improvement. *Br J Gen Pract* 58, 548-54.

6. Department of Health. (2013). *Developing a specification for lifestyle weight management services. Best practice guidance for tier 2 services*. Obesity and Food Policy Branch, PHD. March 2013. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/142723/Weight_Management_Service_Spec_FINAL_with_IRB.pdf
7. Atkinson G, Batterham AM. (2015). True and false interindividual differences in the physiological response to an intervention. *Exp Physiol* 100, 577-588.
8. Johns DJ, Hartmann-Boyce J, Jebb SA, and Aveyard P (2016. Weight change among people randomized to minimal intervention control groups in weight loss trials. Obesity 24, 772–780.
9. Jennings A, Hughes CA, Kumaravel B, Bachmann MO, Steel N, Capehorn M, et al. (2014). Evaluation of a multidisciplinary Tier 3 weight management service for adults with morbid obesity, or obesity and comorbidities, based in primary care. *Clin Obes* 4, 254–266
10. Higgins JPT, Green S (editors). (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011]. The Cochrane Collaboration, Available from http://bit.ly/2bwksTi
11. National Institute for Health and Care Excellence (NICE). (2013). *Weight management: lifestyle services for overweight or obese children and young people. Public health guideline* Published: 23 October 2013, www.nice.org.uk/guidance/ph47
12. Vickers AJ. (2001). The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Med Res Meth* Suppl 1, 6.
13. Carpenter JR, Kenward MG. (2007). *A critique of common approaches to missing data. In: Missing data in randomised controlled trials— a practical guide.* Birmingham: National Institute for Health Research.
14. Allinson P. (2001). *Missing Data. Sage University Papers Series on Quantitative Applications in the Social Sciences*, 07-136. pp 83–84

15. Boghal, MS, Langford R. (2014). Gender differences in weight loss; evidence from a NHS weight management service. *Public Health* 128, 811-813.

Figure Legends

Figure 1

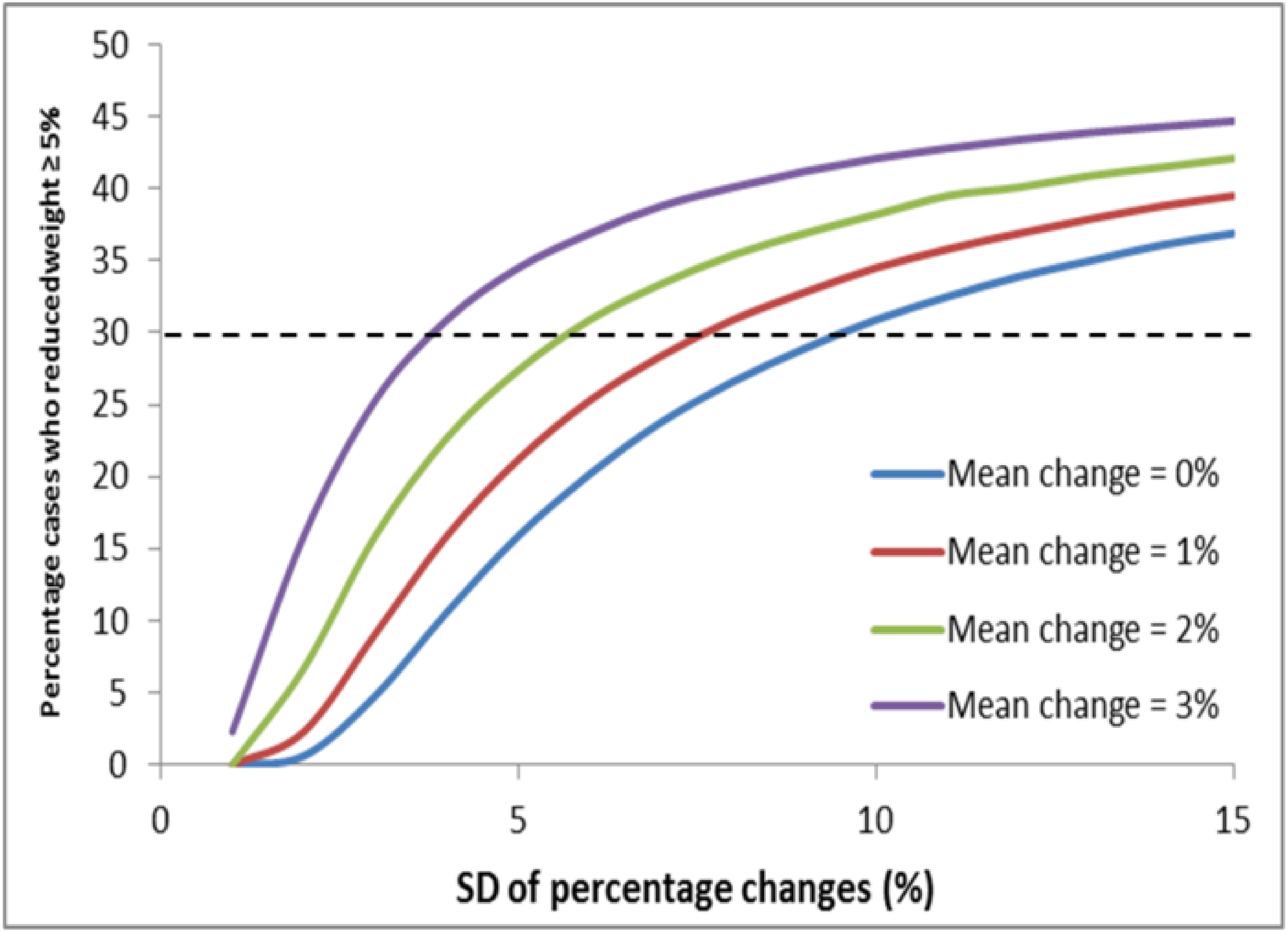
The proportion of people who would apparently reduce their weight by 5% or more for various magnitudes of group mean percentage change and standard deviations of percentage change. The proportions are calculated from the fundamental properties of the Normal curve. The group mean change of zero represents the situation where no weight management service has been prescribed at all. It can be seen that the proportion of people showing a reduction ≥5% can exceed 30% in this situation, especially with the greater SDs of change expected for longer (12-24 month) follow-up time-points.

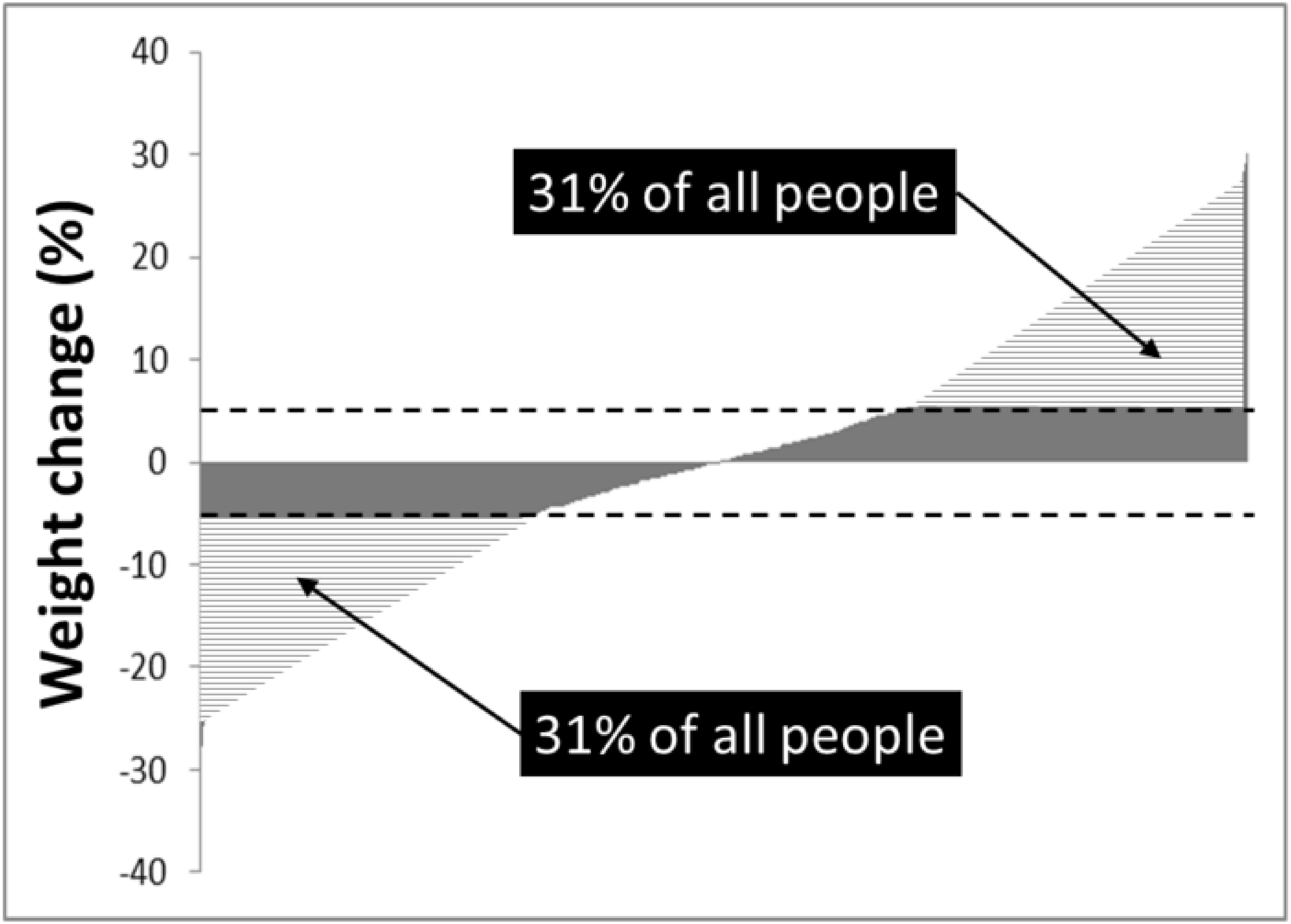
Figure 2

The changes in weight for a hypothetical sample of 1000 people when the mean change in weight is zero and the SD of percentage changes in weight is 10%. Data are assumed to be Gaussian distributed. It can be seen that over 30% of people are expected to show a reduction in weight of 5% or more, even when no weight management service has been implemented between baseline and follow-up measurements of weight. This is due to within-participant random variation in weight during the baseline-to-follow-up period. This random variation is unavoidable.

Figure 1

Click here to download Figure Figure 1.tif




Weight change (%)
40
30
20
10
0
-10
-20
-30
-40
31% of all people
31% of all people