Published in final edited form as:

Int J Obes (Lond). 2016 June ; 40(6): 895–898. doi:10.1038/ijo.2015.212.

Rigorous control conditions diminish treatment effects in weight loss randomized controlled trials

John A. Dawson^{a,*}, Kathryn A. Kaiser^{b,c}, Olivia Affuso^{c,d}, Gary R. Cutter^e, and David B. Allison^{b,c}

^aNutritional Sciences, College of Human Sciences, Texas Tech University, Lubbock, TX

^bOffice of Energetics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL

^cNutrition Obesity Research Center, University of Alabama at Birmingham, Birmingham, AL

^dDepartment of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL

^eSection on Research Methods and Clinical Trials, Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL

Abstract

Background—It has not been established whether control conditions with large weight losses (WLs) diminish expected treatment effects in WL or prevention of weight gain (PWG) randomized controlled trials (RCTs).

Subjects/Methods—We performed a meta-analysis of 239 WL/PWG RCTs that include a control group and at least one treatment group. A maximum likelihood meta-analysis framework is used in order to model and understand the relationship between treatment effects and control group outcomes.

Results—Under the informed model, an increase in control group WL of one kilogram corresponds with an expected shrinkage of the treatment effect by 0.309 kg [95% CI (−0.480, −0.138), $p = 0.00081$]; this result is robust against violations of the model assumptions.

*Correspondence may be addressed to: John A. Dawson, 1301 Akron Ave, Human Sciences Building 301F, Lubbock, TX 79409. john.a.dawson@ttu.edu.

Disclosures of Conflicts of Interest

J. A. Dawson and K. A. Kaiser have no conflicts of interest to declare.

O. Affuso has received consulting fees and grants from organizations interested in obesity interventions.

G. Cutter has had the following consulting or DSMB commitments with in the past twelve months: Apotek, Biogen-Idec, Cleveland Clinic, Glaxo Smith Klein Pharmaceuticals, Gilead Pharmaceuticals, Modigenetech/Prolor, Merck/Ono Pharmaceuticals, Merck, Neuren, PCT Bio, Revalesio, Sanofi-Aventis, Teva, Vivus, NHLBI (Protocol Review Committee), NINDS, NMSS, NICHD (OPRU oversight committee). Additionally, G. Cutter has consulted, received speaking fees or acted as part of an advisory boards for the following: Alexion, Allozyme, Bayer, Celgene, Coronado Biosciences, Consortium of MS Centers (grant), Diogenix, Klein-Buendel Incorporated, Medimmune, Novartis, Nuron Biotech, Receptos, Spiniflex Pharmaceuticals, Teva pharmaceuticals. Dr. Cutter is employed by the University of Alabama at Birmingham and is President of Pythagoras, Inc., a private consulting company located in Birmingham AL.

D. B. Allison has received consulting fees and his university has received gifts, grants, and donations from multiple non-profit and for-profit organizations with interests in obesity trials including food and pharmaceutical companies.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NHLBI, NIDDK or the National Institutes of Health.

Conclusions—We find that control conditions with large weight losses diminish expected treatment effects. Our investigation may be helpful to clinicians as they design future WL/PWG studies.

Keywords

Obesity; Biostatistics; Meta-analysis; Maximum likelihood; Weight loss intervention; Study design

Introduction

Investigators planning an obesity randomized controlled trial (RCT) of drugs, dietary supplements, or other treatments must ask themselves: “How intense should we make the intervention provided to the control group and the treatment group in addition to the treatment to be tested?” This is related to but distinct from the debate about placebo versus active controls¹ and concerns the *background intervention*, the intervention that both the control group and the treatment group receive in addition to the treatment to be tested. On the one hand, some individuals advocate using the minimum background intervention that is ethically acceptable, such as providing a pamphlet on good weight loss (WL) practices. The rationale is two-fold: (A) It is suggested that this approach better emulates the real-life conditions under which people will receive the treatment; and (B) some investigators may wish to observe the maximum *change* the treatment can achieve and are concerned that giving too vigorous a background intervention will cause the estimated treatment effect (the mean difference in outcome, such as WL or prevention of weight gain (PWG), between the control and treatment groups) to be diminished. On the other hand, others advocate using the maximum background intervention that is practical, such as providing a vigorous program of diet and exercise supported by intensive counseling. The rationale for this is also several-fold: (C) it is suggested that this approach may be more clinically sound because it assesses the incremental effect of a (presumably) riskier second-line intervention on top of a (presumably) more accepted and benign first-line intervention; (D) Those that hope to subsequently see the studied treatment be widely adopted, in addition to wishing to report a large treatment effect, wish to report the largest possible mean WL in the treatment group. That is, the statement “Patients receiving X treatment lost an average of Y kilograms” is more compelling when Y is large; and (E) clinicians, investigators, and regulators often wish to evaluate the best performance that the treatment can achieve and it is presumed that by coupling it with the most vigorous background intervention possible, this will produce the greatest overall effect.

The focus of this review, however, is not on the relative merits of specific types of background interventions (e.g., providing only a weight-loss pamphlet) but rather on the sequelae from choosing a relatively weak or strong background intervention, namely the amount of WL in the control group, how much this differs from that of the intervention group (i.e., the treatment effect) and their relationship. Information about this relationship is needed if informed decisions are to be made about the relative merits of strong versus weak background interventions, as well as considerations in power analyses and sample size calculations. Yet, after a careful search of the literature and discussion with colleagues in the

field, we have been unable to identify any published or unpublished analyses that provide information about the relation between the true mean change in control group outcome and treatment effect, when considering the broader population of WL/PWG RCTs. It is tempting to try to answer whether there is an association between these values using standard statistical approaches such as ordinary linear regression or correlation coefficients. However, as explained by Schmid *et al.*² and McIntosh³, such simple approaches are likely biased and a different approach is required. Herein, we address this question by performing a meta-analysis of WL/PWG RCTs and discuss the implications of its results with respect to the qualitative choice of the background intervention and its potential consequences.

Materials and Methods

This meta-analysis was performed as a secondary analysis to another investigation⁴; as it uses publically available summary data, these projects received a certificate of exemption from the University of Alabama at Birmingham Institutional Review Board in accordance with the Common Rule. The review criteria here are slightly changed from that previous investigation, so we provide the relevant details below:

Article Search Criteria and Retrieval

Articles, abstracts and doctoral dissertations were retrieved using searches performed on electronic databases: PubMed, Cochrane Library, ISI Web of Science, PsycINFO, CINAHL (Cumulative Index to Nursing and Allied Health Literature) and Dissertation Abstracts. PubMed was searched with and without MeSH headings to identify publications for inclusion, using the following limits: publication date, RCTs and human studies.

Article Selection Criteria

All studies were evaluated according to the following inclusion criteria:

1. The data were from human studies.
2. The study was a randomized controlled trial (RCT).
3. The study had a total sample size of at least 30 participants at enrollment.
4. The study protocol included an intervention period of at least 8 weeks.
5. Weight loss and/or weight gain prevention was a primary or secondary outcome variable.
6. The publication was available in the English language.
7. The study was published between January 1, 2007 and July 1, 2009.

The above provided an initial pool of 317 WL/PWG RCTs in adults for meta-analysis. However, this analysis also has the following exclusion criterion:

1. The study could not be a purely head-to-head comparison of two or more experimental treatments (i.e., no real control group)

This criterion excluded 78 studies from the initial pool, so that the 239 remaining studies are all experiments that compare a control group of some kind with one or more treatment

groups; information regarding these studies may be found in the Supplementary Materials. The nature of the control groups varied in scope from those that were not given a treatment (e.g., the control group consisted of patients on a wait list) to those that received the standard of care alternative to some experimental drug. Of the 239 trials, 80 had more than a single treatment arm; within these the number of study arms ranged from three to eight.

Statistical Model

In all of the studies under consideration, the observed control group and treatment group WLs (and hence the observed treatment effect) have some error associated with them, as they are based on random samples and therefore we should not naively treat them as substitutes for the true values of those quantities, either within studies or across the population of studies. Thus, in order to interrogate the relationship between namely the amount of WL in the control group and how much this differs from that of the intervention group, we apply the model provided by McIntosh³, which assumes that true (population-level) WL in the i^{th} study's control arm (θ_{C_i}) and the treatment effect (Δ_i) are related thusly

$$\begin{aligned} \theta_{C_i} &\sim N(\mu_C, \nu_C) \\ \Delta_i | \theta_{C_i} &\sim N[(\mu_E - \mu_C) + \beta^* \times (\theta_{C_i} - \mu_C), \nu_E] \end{aligned} \quad (\text{EQ1})$$

where β^* is one less than the 'structural slope'. To clarify this point, the original model provided by McIntosh is equivalent to the one given by (EQ1) but the original model is parameterized in terms of the control WL and the treatment WL, rather than the control WL and the treatment effect; since we are interested in the latter, this parameterization with β^* is more informative.

This model allows for the within-study treatment effect Δ_i to deviate from the population-level treatment effect ($\mu_E - \mu_C$, where μ_E is the average treatment group WL across studies and μ_C is the average control group WL across studies) in response to a divergence of the control arm WL θ_{C_i} from μ_C , and the nature of that deviation is determined by β^* ; see Figure 1. For a concise summary of the possible clinical interpretations of β^* , depending on its value, see Table 1; by estimating β^* , we estimate the relationship between the intensity of the background intervention and the expected treatment effect. Estimates for quantities of interest were derived via an Expectation-Maximization (EM) algorithm⁵ as described by McIntosh³ (see pp. 1718–9) with standard errors for the EM estimates obtained via a supplemented EM (SEM^{6,7}).

Statistical Adjustments and Considerations

It is the case that 80 of our studies have more than two study arms. However, the model presented in the previous section assumes that all studies have exactly two arms: a treatment group and a control group. Additionally, the model assumes that the group WL variances are known quantities. We address these discrepancies by treating a K-arm study as (K-1) two-arm studies, with the same control group values and using the study-derived estimates of the

group WL variances as the true values. Both of these choices are violations of the model's inherent assumptions and we present sensitivity analyses that assess their impact.

Additionally, missing data exist for both means and standard deviations of weight change within study arms⁴. Thus, we performed multiple imputation using SPSS version 20 (ref. 8) and a total of 30 imputations. Correspondingly, the point estimates and measures of their uncertainty will reflect a pooling over these imputations⁷ (see pp. 522–3).

Code Versioning and Availability

All computations, including the implementation of the EM and SEM algorithms, were performed in R version 2.15.2 (ref. 9); relevant code are available from the corresponding author by request.

Results

Meta-analysis of our WL/PWG studies indicates that the maximum likelihood estimate for β^* is -0.309 and the corresponding 95% confidence interval is $(-0.480, -0.138)$. The probability of observing an estimate of β^* at least this far from 0 when β^* is zero is 0.00081. Therefore, we conclude that β^* falls within $(-1, 0)$ and hence increases in control group WL are associated with corresponding decreases in treatment effect. This and other computed quantities, such as the population variance of the control groups across the population of studies, may be found in Table 2. Our finding for β^* has several clinical ramifications, which we will discuss presently after providing some sensitivity analyses in the next section in order to assess the robustness of our results in light of known model violations.

Sensitivity Analyses

We conducted three sensitivity analyses to assess the robustness of our results against known violations of the model assumptions. First, our treatment of studies with multiple treatment arms violates the assumption of independence and identical distribution in EQ1. In order to examine its potential effects, we reran the analysis one thousand times, using only one treatment arm chosen at random per study. In none of the thousand iterations of this procedure does the 95% confidence interval for β^* cross either -1 (all lower bounds greater than -0.56) or 0 (all upper bounds lower than -0.09).

Second, in order to gauge the potential effects of using sample variances as true variances, we reran the analysis using moderated variances: sample variances were replaced by weighted averages of the sample estimate and the median of all variances in that arm type (control or treatment), with weights by sample size compared with the median sample size. After doing so, the maximum likelihood estimate for β^* is -0.343 and the corresponding 95% confidence interval is $(-0.504, -0.181)$.

Lastly, we checked to see if the presence of a few very large studies unduly influenced our results by removing any of our two-arm studies with sample size greater than 600; 600 roughly corresponds to the upper 4% of the observed distribution of sample sizes and was chosen because it resides in the largest gap between observed sample sizes on the high end (neighbored by 566 and 707); 14 studies are 'large' according to this criterion. Under this

restriction, the maximum likelihood estimate for β^* is -0.319 and the corresponding 95% confidence interval is $(-0.486, -0.151)$.

We conclude that our analysis is robust against the violations of model assumptions incurred by our specific approach.

Discussion

While all values of β^* in $(-1, 0)$ can be qualitatively described in the same manner, the degree to which increases in control group WL shrink the expected treatment effect depend on the magnitude of β^* . Under the informed model, for every kilogram by which the control group population mean weight loss increased, the treatment *effect* is on average reduced by about 0.3 kg. So, for example, if there are two WL RCTs that have control group WLs equal to 1 kg and 5 kg, respectively, our meta-analysis predicts expected treatment group WLs of 3.1 kg and 5.9 kg, which correspond to expected treatment effects of 2.1 and 0.9 kg, respectively.

Prima facie, the importance of this result that more rigorous background interventions do ‘steal’ some efficacy from the studied treatment might not be obvious. However, it has several direct impacts in the planning and execution of WL/PWG RCTs. First, use of strong background interventions are associated with smaller-than-expected treatment effects; even taking the suppositions of rationales (C), (D) and (E) from the introduction as truth, governmental regulatory bodies are first and foremost interested in seeing statistically significant and clinically meaningful treatment *effects*, as opposed to merely large WLs among subjects in the intervention arm; indeed, according to the FDA guidance for weight loss RCTs¹⁰, the only two primary efficacy endpoints that merit consideration are “[t]he difference in mean percent loss of baseline body weight in the active-product vs placebo-treated group” or “[t]he proportion of subjects who lose at least 5 percent of baseline body weight in the active-product vs placebo-treated group”. If it cannot be shown that one of these primary endpoints has a statistically significant and clinically meaningful treatment effect, the investigational product is simply nonviable.

Second, providing an estimate of the expected reduction in treatment effect given the amount of control WL allows researchers to better *a priori* estimate the treatment effects that they might see in their WL/PWG RCTs, so that these studies will have sufficient sample size in order to be adequately powered. Considering that only modest weight losses were observed in these *published* RCTs (as the reader may see in Table 2, the average treatment effect was only a little more than 2 kg), regardless of the observed differences in rigor of control interventions, researchers simply cannot rely on large effect sizes to power studies. As the reader may also note from Table 2, there is considerable variation about the estimates of the treatment effects given the control WLs, so our guideposts should perhaps be treated only as a conservative measure of what kind of attenuations of treatment effect might be seen in practice, just to be safe. Third and lastly, knowing what kinds of treatment effects to expect will help provide meaningful informed consent among subjects undergoing WL/PWG RCTs, considering that roughly half of subjects would not countenance even a single

percentage point chance of a serious adverse event when expecting only modest (5 – 10%) weight loss¹¹.

Some of the strengths of this analysis include the very large sample size of RCTs utilized to derive the estimates. Additionally, very rigorous methods were used to select studies, extract the data and assure completeness and accuracy of the dataset; for example, three rounds of inter-rater reliability were performed when assessing study characteristics and data, to ensure agreement at or above 80% (ref. 4). While some differences in adjusted weight loss may be observed between different types of interventions (e.g., diet plus exercise typically having a greater effect on weight loss, versus exercise alone), the range of effects outside of bariatric surgery (not included in this analysis) is not large. Some limitations of the present analysis are due to potential issues in the studies that inform it and include combined gender analyses, the inclusion of diabetic and insulin-resistant patient samples and the possibility that measures of baseline body composition such as BMI and percent body fat were not comparable across control and treatment groups. Related in part to this last point, the quantity β^* reflects the association between the treatment effect and divergences of the observed control mean from the average control mean across studies. While we feel it is reasonable to assume that these divergences are primarily being driven in differences in the background intervention across studies, it is also possible that the values of the divergences may be affected by differences in population-specific risk factors which may (such as age or gender) or may not (such as BMI or some ‘unknown unknown’) have been recorded. However, since those differences might strengthen or weaken the divergences across studies in a manner independent of the background intervention, it is our belief that they should only add to the variability of the control means marginally without biasing them.

Analysis of more homogeneous sub-populations may show different estimates. However, these types of participants and targeted populations are representative of the present day status of overweight and obese persons that are typically recruited for obesity RCTs. One of the major barriers to our sample size was the degree of data not reported and/or not available in public repositories or from author contacts. We employed multiple imputation and sensitivity analyses to address these limitations.

In conclusion, our results confirm the intuition of some investigators that more rigorous background interventions do ‘steal’ some efficacy from the studied treatment. For every kilogram by which the control group population mean weight loss increased, the treatment *effect* is reduced by about 0.3 kg. This recognition allows researchers to obtain more accurate estimates of treatment effects *a priori*, in order to design adequately powered investigations that will have a greater chance of gaining regulatory approval.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

DBA conceived of the analysis. GAH advised on study design and data collection/coding approaches. KAK and OA collected and coded data. KAK performed multiple imputation of the missing data. JAD coded and performed the

analyses. All authors were involved in writing the paper and had final approval of the submitted and published versions.

The authors would like to thank Christopher Schmid for comments and feedback that helped improve the manuscript. The authors would also like to thank Tiffany Carson, Katherine Ingram and Firas Abbas for assistance in data collection and coding.

This work was primarily funded by Grant Number R01DK078826, with other support from T32HL072757, P30DK056336, T32HL007457, T32HL079888 and T32DK062710.

References

1. Leon AC, Solomon DA. Toward rapprochement in the placebo control debate. *Eval Health Prof.* 2003; 26(4):404–414. [PubMed: 14631611]
2. Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Stat Med.* 1998; 17:1923–1942. [PubMed: 9777687]
3. McIntosh MW. The population risk as an explanatory variable in research synthesis of clinical trials. *Stat Med.* 1996; 15:1713–1728. [PubMed: 8870154]
4. Affuso O, Kaiser KA, Carson TL, Ingram KH, Schwiers M, Robertson H, et al. Associations of run-in periods with weight loss in obesity randomized controlled trials. *Obes Rev.* 2013; 15 (1):68–73. [PubMed: 24118736]
5. Dempster AP, Laird NM, Rubin DB. Maximum likelihood estimation from incomplete data. *JRSSB.* 1977; 39:1–38.
6. Meng X-L, Rubin DB. Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *JASA.* 1991; 86(416):899–909.
7. Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. *Bayesian data analysis.* 2. Chapman and Hall/CRC; Boca Raton: 2004.
8. IBM Corporation. *SPSS version 20.* IBM; Armonk, NY: 2012.
9. R Development Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing; Vienna, Austria: 2009.
10. US Food and Drug Administration. [Last accessed May 14, 2015] Developing products for weight management: draft guidance. <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm064981.htm>
11. Allison DB, Elobeid MA, Cope MB, Brock DW, Faith MS, Sargent S, et al. Sample size in obesity trials: Patient perspective vs current practice. *Med Decis Making.* 2010; 30 (1):68–75. [PubMed: 19675318]

Figure 1.

An illustration of the relationship between β^* and expected treatment effect using data simulated under the structural model. In the figure, simulated treatment group means are plotted on the Y-axis against control group means on the X-axis, using $\beta^* = -0.309$; γ represents an arbitrary offset. Since β^* is in $(-1,0)$, larger control group means are associated with smaller treatment effects μ_i (represented by the lengths of the vertical segments) on average.

Table 1

Interpretations of the possible states of β^*

Possible states	Interpretation
$\beta^* < -1$	Increasing control group weight loss (WL) θ_{C_i} is associated with shrinking treatment group WL θ_{E_i} . This scenario is unlikely in our setting as both WL groups receive a common background intervention in every study and hence are expected to be positively correlated.
$\beta^* = -1$	θ_{C_i} and θ_{E_i} are uncorrelated. This scenario is also unlikely due to a common background intervention in every study.
β^* is in $(-1, 0)$	Increasing θ_{C_i} is associated with increasing θ_{E_i} , but at a slower rate than θ_{C_i} ; thus, increasing θ_{C_i} is associated with shrinking treatment effect τ_i . See Figure 1 for an illustration of this scenario.
$\beta^* = 0$	θ_{C_i} and τ_i are uncorrelated. Hence the expected treatment effect is constant and equal to τ for all values of θ_{C_i} .
$\beta^* > 0$	Increasing θ_{C_i} is associated with increasing τ_i .

Table 2

Structural model parameters as estimated in the meta-analysis

Parameter	Estimate	Within-Imputation Variance	Between-Imputation Variance	Total Variance	Approximate d.f.	95% C.I. of the Estimate
μ_E	3.136	0.028	0.0018	0.030	65.48	(2.79, 3.48)
μ_C	1.043	0.011	0.0036	0.015	35.06	(0.797, 1.289)
β^*	-0.309	0.0049	0.0020	0.0071	33.70	(-4.480, -0.138)
v_E	8.369	0.375	0.116	0.495	35.35	(6.941, 9.797)
v_C	5.473	0.169	0.080	0.252	33.11	(4.452, 6.494)

This table contains quantities related to the parameters of the structural model. μ_E and μ_C correspond to the true average weight loss across treatment and control arms, respectively, in kilograms. v_C is the population-level variance for true weight loss means in the control arms across studies and v_E is the residual variance; see EQ1. β^* is one less than to the structural slope; see Table 1 for its interpretation.