Published in final edited form as:

Clin Obes. 2015 February ; 5(1): 38–41. doi:10.1111/cob.12083.

Randomized Controlled Trial Examining Expectancy Effects on the Accuracy of Weight Measurement

Gareth R. Dutton, PhD, Kevin R. Fontaine, PhD, Amy S. Alcorn, MPH, RD, John Dawson, PhD, Patrice L. Capers, PhD, and David B. Allison, PhD

University of Alabama at Birmingham

Abstract

BACKGROUND—Researchers’ and participants’ expectations can influence treatment response. Less is known about the effects of researchers’ expectations on the accuracy of data collection in the context of a weight loss trial.

METHODS—Student raters (N=58; age=20.1 ± 2.3 years) were recruited to weigh individuals who they thought were completing a 12-month weight loss trial, although these ‘participants’ were actually standardized patients (SPs) playing these roles. Prior to data collection, student raters were provided information suggesting that the tested treatment had been effective. Each student rater received a list of 9–10 ‘participants’ to weigh. While the list identified each person as ‘treatment’ or ‘control’, this assignment was at random, which allowed us to examine the effects of non-blinding and expectancy manipulation on weight measurement accuracy. We hypothesized that raters would record the weights of ‘treatment participants’ as lower than those of ‘control participants’. This study is registered at ClinicalTrials.gov as COB12083.

RESULTS—Contrary to our hypothesis, raters recorded weights that were 0.293 kilograms heavier when weighing ‘treatment’ versus ‘control’ SPs, though this difference was not significant (p=0.175).

CONCLUSIONS—This pilot study found no evidence that manipulating expectancies about treatment efficacy or not blinding raters biased measurements. Future work should examine other bias which may be created by not blinding research staff who implement weight loss trials as well as the participants in those trials.

Keywords

weight; measurement; expectancy effect; randomized controlled trial; bias

Introduction

Double-blind randomized controlled trials (RCT) represent the “gold standard” for evaluating the efficacy of weight loss interventions.¹ In this design, participants and investigators are naïve to randomized assignment, and participants are treated identically

Corresponding Author: David B. Allison, Ph.D., Ryals Public Health Building, Room 140J, University of Alabama at Birmingham, 1665 University Boulevard, Birmingham, Alabama 35294, dallison@uab.edu.

Conflict of Interest: The authors declare no conflict of interest.

except for the treatment received. Double-blinding may be particularly important, since researchers can otherwise convey information verbally and/or non-verbally about the study protocol that may significantly influence participants' behaviors and attitudes.²⁻⁸ Additionally, participants' beliefs and expectancies about treatment can substantially affect their response as well.⁹⁻¹³

However, less is known about researchers' beliefs and expectancies about treatment and the potential influence on study procedures and the accuracy of data collection. This is especially relevant, as it is not always practical (or even possible) to initiate and maintain blinding of both researchers and participants to treatment assignment when it involves behavioral or lifestyle interventions. The purpose of this study was to examine the effects of non-blinding and manipulating expectations about treatment efficacy among individuals tasked with measuring and recording weights of supposed participants in a weight loss trial. It was hypothesized that non-blinded individuals who were told that the weight loss treatment protocol was very effective would tend to underestimate recorded weights of 'treatment participants' and/or overestimate recorded weights of 'control participants'.

Materials and Methods

Participants

Undergraduate and graduate students (N=59) were recruited via flyers placed in residence halls and other campus buildings and were initially informed that researchers were seeking individuals to assist with data collection (i.e., weighing and recording weights) for the final assessment of a large, 12-month weight loss trial being conducted at the academic medical center.

Procedures

Students were scheduled to arrive to the research clinic, where they were provided with a brief overview of the 'weight loss study' and their roles for data collection. Prior to data collection, the same research staff member met with each participant and described the study as a 2-arm weight loss trial in which participants were randomized to either a lifestyle intervention or no-treatment control. During this initial orientation, the staff member indicated that the intervention appeared to be very effective in producing weight loss and underscored the importance of the weights to be recorded that day for evaluating the program. Students were instructed on how to use the balance beam scales to measure participants' weights and how to record these values on data collection forms. The data collection forms included a name, participant identification number, and group assignment (treatment vs. control) for each individual to be weighed along with a place for students to document each recorded weight. Each form was enclosed in randomized opaque envelopes with a random order of participants. This envelope was picked by the study personnel and handed one copy of the form to the student rater and kept the other so the study personnel was aware of the 'participants' assignment.

During data collection, the 59 student raters weighed 9-10 individuals who were actually standardized patients (SPs) trained to play the role of weight loss trial participants. These

individuals were recruited from the pool of all SPs affiliated with the academic medical center's clinical learning center with the assistance of the center's coordinator. The coordinator provided a brief description of the project needs to SPs and requested participation from those who self-identified as overweight or obese. Of the 10 SPs, six were women, and eight were Caucasian.

When playing the role of treatment or control participants, some SPs delivered brief scripts during the weigh-in encounter, which described the program as very helpful for promoting weight loss ('treatment' participants) or expressing disappointment at having been randomized to the 'control' group. Each SP played the role of both treatment and control at random for different raters. Thus, we attempted to elicit treatment expectations among student raters in the initial encounter with research staff as well as in the interactions with SPs during weigh-ins. To avoid raters from seeing SPs with another rater, each rater was placed in a separate room where the SP was brought to by research staff.

Upon completion of data collection with the 9-10 weight loss study 'participants', student raters returned their completed data collection forms to study personnel and were scheduled to return to the research clinic later that week to receive their compensation. In addition to receiving payment, student raters were also debriefed on the true purpose of the study at this subsequent visit. The debriefing sessions occurred after all student raters had completed the study protocol to ensure that no one learned about the true purpose of the study prior to participating. After the debriefing, retrospective informed consent was obtained, and student raters received \$30 for participating. No one declined to provide consent. This study protocol was approved by the UAB IRB. This study is registered at ClinicalTrials.gov as COB12083.

Statistical Approach

The SP role as either a 'control' or 'treatment' weight loss trial participant was randomized within each rater, such that five of the SPs were 'controls' and five were 'treatments' for any given rater. While 59 participants completed the study, one participant recognized several of the SPs; thus, we performed analyses with and without this participant, as a sensitivity analysis. Weights were not recorded for one SP (who was absent on one of the weigh-in days) for 31 raters; two of those 31 raters also did not record a weight for another SP (due to the SP having to leave the facility early on that day), leading to a total of 33 missing recorded weights. These missing values were handled via multiple imputation, using 20 imputations in total. The effect of raters' expectation was tested through ordinary least squares (OLS) linear regression of weights on the supposed treatment, controlling for SP (since there was no 'gold standard' measurement of any SP's true weight), the age of the rater, the sex of the rater, the day on which the weights were recorded, and the room in which the weights were recorded in the research clinic (one of four rooms, each with its own scale) as covariates. Because residuals from the analysis regressions exhibited 'fat tails', and hence a violation of the assumption of normality, a robust check of the regression's results was performed via permutation tests. As these were not qualitatively different, we report the results from the usual OLS.

Results

Of the 58 participants contributing data for analyses, the mean age was 20.1 ± 2.3 years, and approximately 78% were women. Approximately 41% were Caucasian and 36% were African American. The majority (83%) were undergraduate students.

In Figure 1, recorded weights are shown by SP under ‘control’ (panel A) and ‘treatment’ (panel B) conditions. Means and medians for each SP, calculated using recorded weights for both conditions, are superimposed using blue Xs and green pluses, respectively. Analyses of these data indicate that raters recorded a weight that was 0.293 kilograms heavier when weighing an SP in his/her ‘treatment’ role as opposed to the ‘control’ role; however, this difference was not statistically significantly greater than zero (95% CI $[-0.13, 0.72]$, $p=0.175$). Alternative models controlling for raters’ race, ethnicity, education, or academic status, as well as the sex of the SP, did not produce qualitatively different results. Similarly, an analysis that simply ignored missing weight values was not qualitatively different from the aforementioned results, nor did the inclusion or exclusion of the participant who recognized SPs qualitatively change the results.

Discussion

Contrary to our hypothesis, we found no evidence that manipulating the expectancy related to the effectiveness of a weight loss program biased the measurement and recording of body weight. This suggests that, within this simulated clinical research context, there was no apparent measurement bias due to non-blinding.

There are several potential limitations that may mitigate our findings. First, despite multiple promptings to the students about the effectiveness of the weight loss “treatment”, the students had no vested interest in the results of the “trial” (i.e., they were participating briefly and merely for modest monetary compensation). As such, in the absence of any personal or professional interest in the outcome of the trial, the expectancy manipulations may not have been sufficiently powerful to elicit bias. Second, the information provided by researchers about the perceived effectiveness of the intervention may have negated or diminished any conscious or unconscious inclination for raters to bias their weight measurements; such might not have been the case if a state of equipoise had been perceived by the raters. Third, some of the student raters may have questioned the true purpose of their participation and may have altered their behaviors accordingly.

Fourth, the accuracy of the weight measurements may have been compromised by the use of a balance beam scale. Despite training on how to use the scale, a majority (39/59) of the students recorded a weight that deviated from the median weighing of that SP by at least one kilogram; roughly half had at least one weighing that deviated by 2.5 kilograms (see Figure 1). These deviations, which are many times greater in magnitude than the effect sizes observed in the data, add a good deal of noise around any signal that might exist. The use of a digital scale, which requires less skill or practice, would almost surely have produced more precise results and thus attenuating this noise in the experimental system.

Research on the accuracy of blood pressure measurement has shown that assessors demonstrate a bias to round up or down (most often to a number ending in 0 or 5) when recording blood pressure values.¹⁴ When applied to the current measurement context, use of digital scales could reduce or eliminate one source of bias (i.e., selection of the scale's balance point when weighing), but potential bias in the transfer of the value to the data collection form remains for both methods of weight measurement. Use of digital scales was not chosen because it was believed that it allowed too little opportunity for judgment and hence for bias to enter the procedure.

It is also worth noting that the lack of observed effects for non-blinding was specific to the measurement of body weight. No conclusions can be made about the effects of non-blinding on other, less objective outcomes such as assessor-led interviews or other research measures requiring some degree of assessor interpretation or subjective scoring. In fact, it is plausible that assessors' expectancy effects and non-blinding may be confined to, or at least more pronounced with, more subjective outcome measures.

Strengths of this study included a novel research design that modeled scenarios experienced in a typical weight loss trial to examine its influence on the raters' evaluation of weight recordings, the manipulation of expectancies related to the success of the intervention, and the randomizing of SP's to play the role of both a treatment and control participant across the student raters.

Research shows that under some circumstances RCTs with non-blinded evaluators produce different effects than do studies with blinded evaluators,¹⁵ suggesting that ideally evaluators should be blinded to treatment assignment. In principle, such blinding remains a sound precaution. Nevertheless, though further research is warranted, the current results suggest that non-blinded evaluators in weight loss RCTs do not produce obvious biases in results, at least under the conditions simulated in the current investigation (i.e., measurement of body weight by student raters). Future research should explore this topic with actual research staff involved in the implementation of weight loss trials rather than one-time, untrained evaluators. Future investigation examining potential effects on less "objective" outcome measures, such as interviewer-administered surveys, may also be informative.

Acknowledgements

GD, KF, AA, and DA conceived the experiment and developed the study protocol. GD, KF, AA, and PC conducted the experiment. JD and DA designed and conducted the statistical analyses. All authors were involved in writing the manuscript and had final approval of the submitted work. This project was supported by award numbers P30DK056336 and K23DK081607 from the National Institute of Diabetes and Digestive and Kidney Diseases; T32HL072757 and R25HL124208 from the National Heart, Lung, and Blood Institute; and K12GM088010 from the National Institute of General Medical Sciences. The opinions expressed are those of the authors and not necessarily those of the NIH or any other organization.

References

1. Jensen MD, Ryan DH, Apovian CM, et al. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Obesity Society. *J Am Coll Cardiol*. 2013 doi:10.1016/j.jacc.2013.11.004.
2. Michael RB, Garry M, Kirsch I. Suggestion, cognition, and behavior. *Curr Dir Psychol Sci*. Jun; 2012 21(3):151–6.

3. Rosenthal R. Interpersonal expectancy effects: a 30-year perspective. *Curr Dir Psychol Sci.* Dec; 1994 3(6):176–9.
4. Rosenthal R. Covert communication in laboratories, classrooms, and the truly real world. *Curr Dir Psychol Sci.* Oct; 2003 12(5):151–4.
5. Weber SJ, Cook TD. Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inferences. *Psychol Bull.* Apr; 1972 77(4):273–95.
6. Klien O, Doyen S, Leys C, et al. Low hopes, high expectations: Expectancy effects and the replicability of behavioral experiments. *Perspect Psychol Sci.* Nov; 2012 7(6):572–84.
7. Rosenthal R, Lawson R. A longitudinal study of the effects of experimenter bias on the operant learning of laboratory rats. *J Psychiat Res.* Jun.1964 2:61–72. [PubMed: 14177091]
8. Rosenthal R. Experimenter and clinical effects in scientific inquiry and clinical practice. *Prev & Treat.* Oct.2002 5(1) Article 38. Available at: <http://journals.apa.org/prevention/volume5/pre0050038c>.
9. Colagiuri B. Participant expectancies in double-blind randomized placebo-controlled trials: potential limitations to trial validity. *Clin Trials.* Jun; 2010 7(3):246–55. [PubMed: 20421243]
10. Bausell R, Lao L, Bergman S, Lee WL, Berman BM. Is acupuncture analgesia an expectancy effect? Preliminary evidence based on participants' perceived assignments in two placebo-controlled trials. *Eval Health Prof.* Mar; 2005 28(1):9–26. [PubMed: 15677384]
11. Thomas JL, Guo H, Lynam IM, et al. The impact of perceived treatment assignment on smoking cessation outcomes among African-American smokers. *J Gen Intern Med.* Sep; 2008 23(9):1361–6. [PubMed: 18587620]
12. Colagiuri B, Morley KC, Boakes RA, Haber PS. Expectancy in double-blind placebo-controlled trials: an example from alcohol dependence. *Psychother Psychosom.* 2009; 78(3):167–71. [PubMed: 19270472]
13. Hróbjartsson A, Forfang E, Haahr MT, Als-Nielsen B, Brorson S. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *Int J Epidemiol.* Jun; 2007 36(3):654–63. [PubMed: 17440024]
14. Burnier M, Gasser UE. End-digit preference in general practice: A comparison of the conventional auscultatory and electronic oscillometric methods. *Blood Pressure.* 2008; 17:104–109. [PubMed: 18568699]
15. Hróbjartsson A, Thomsen A, Emanuelsson F, et al. Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *CMAJ.* Mar; 2013 185(4):E201–11. [PubMed: 23359047]

What is already known about this subject

- A methodological strength of double-blind weight loss trials is that researchers are unaware of participants' treatment assignment.
- Information about a research trial conveyed to participants by researchers can influence participants' response to a protocol.

What this study adds

- Less is known about how researchers' behaviors or attitudes may be influenced by knowledge of the trial.
- This is the first study to examine the effect of research staff expectations on the accuracy of data collection in a weight loss trial.

Figure 1.

Recorded weights for each standardized patient (SP) are presented according to whether the participants were told that the SPs were part of the 'control' (A) or 'treatment' (B) groups. Average SP weights calculated across both conditions are indicated by blue Xs. Median SP weights calculated across both conditions are indicated by green pluses.