Contemporary Clinical Trials Communications 4 (2016) 222–228

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

Application of the marginal structural model to account for suboptimal adherence in a randomized controlled trial

James Rochon [a, *], Manjushri Bhapkar [b], Carl F. Pieper [c], William E. Kraus [b, d], for the CALERIE Study Group

[a] Rho Federal Systems, 6330 Quadrangle Drive, Chapel Hill, NC 27517, USA
[b] Duke Clinical Research Institute, 2400 Pratt Street, Durham, NC 27710, USA
[c] Department of Biostatistics and Bioinformatics, Duke University Medical Center, 2424 Erwin Road, Durham, NC 27710, USA
[d] Duke Molecular Physiology Institute, 300 North Duke Street, Durham, NC 27701, USA

ARTICLE INFO

Article history:
Received 12 July 2016
Received in revised form
14 October 2016
Accepted 25 October 2016
Available online 3 November 2016

Keywords:
Causal modeling
Marginal structural model
Weighted GEE model
Adherence
Aging
Calorie restriction

ABSTRACT

Background: There is considerable interest in adjusting for suboptimal adherence in randomized controlled trials. A per-protocol analysis, for example removes individuals who fail to achieve a minimal level of adherence. One can also reassign non-adherers to the control group, censor them at the point of non-adherence, or cross them over to the control. However, there are biases inherent in each of these methods. Here, we describe an application of causal modeling to address this issue.
Methods: The marginal structural model with inverse-probability weighting was implemented using a weighted generalized estimating equation model. Two ancillary models were developed to derive the weights. First, stepwise linear regression was used to model the observed percent weight loss, while stepwise logistic regression model was applied to model early discontinuation from the intervention. From these, participant- and time-specific weights were calculated.
Discussion: This model is complicated and requires careful attention to detail. Which variables to force into the ancillary models, how to construct interaction terms, and how to address time-dependent covariates must be considered. Nevertheless, it can be used to great effect to predict intervention effects at full adherence. Moreover, by contrasting these results against intention-to-treat results, insights can be gained into the intrinsic physiologic effect of the intervention.
Trial registration: ClinicalTrials.gov Identifier NCT00427193.

1. Introduction

Clinical trials are frequently described as the “gold standard” for evaluating the effectiveness of a biomedical intervention. Randomization is expected to create treatment groups equivalent with respect to known and unknown confounding factors prevailing at baseline. However, this balance can be quickly eroded by confounders arising post randomization. They include drop-out, adverse events, and the use of concomitant medications. Suboptimal adherence is particularly insidious. It undermines the credibility of the intervention, and casts doubt on the external validity of the results. The traditional intention-to-treat (ITT) analysis effectively ignores these factors. This attenuates the treatment effect towards the null robbing the study of power to detect significant differences. For this reason, there is generally strong interest in adjusting for the time-dependent confounder.

A number of approaches have been proposed. A per-protocol analysis, for example, removes individuals who fail to achieve a minimal level of adherence. However, this diminishes the external validity of the study, reduces power, and if applied differentially, can lead to bias in the between-group comparisons. One can also censor them at the point of non-adherence or cross them over to the control group at that point. Peduzzi et al. [1] however, demonstrated that there are biases in each of these methods. Another approach is to include adherence as a covariate in a linear model. However, adherence is an endogenous variable and satisfies the definition of a time-dependent confounder. That is, it is

Abbreviations: ITT, intention-to-treat; MSM, marginal structural model; GEE, generalized estimating equation; CALERIE, Comprehensive assessment of the long-term effects of reducing intake of energy; CR, calorie restriction; BMI, body mass index; RMR, resting metabolic rate; %WL, percent weight loss; s.e., standard error.

* Corresponding author.
E-mail address: James_Rochon@RhoWorld.com (J. Rochon).

http://dx.doi.org/10.1016/j.conctc.2016.10.005

influenced by past and current treatment effectiveness while simultaneously affecting future outcomes. Standard linear models are not appropriate under these circumstances.

In this paper, we provide a case study on applying a causal model to account for suboptimal adherence and predict intervention effects at full adherence. Specifically, the marginal structural model (MSM) using inverse-probability [2,3] is widely applied in the epidemiologic [4–6] and clinical trials literature [7,8] including accounting for adherence [9–11]. The MSM is a form of causal analysis based upon the theory of counterfactuals [12]. It allows inference to be drawn under the assumption that, possibly contrary to fact, all participants had adhered to the intervention at the prescribed level. It is implemented using a weighted generalized estimating equation (GEE) model [13]. Informally, the weights create a pseudo-population in which the confounding effects of the observed adherence are removed and the effect of intervention is observed. As such, it is a longitudinal version of the propensity score method [14]. This allows a participant's propensity to adhere to change as circumstances are perceived over the course of the study.

The original MSM papers and subsequent applications were focused on a binary confounder using the Bernoulli distribution. A rough outline was provided, however, on how to implement this model for a continuous confounding measure using the Gaussian distribution. Here, we provide the details on how to do so. We describe the process to derive the weights and implement the weighted GEE model. We contrast the MSM results against the ITT results. We then use these differences to interpret the effects of the intervention.

2. The CALERIE study

2.1. Overview

To fix ideas, consider the Comprehensive Assessment of the Long-term Effects of Reducing Intake of Energy (CALERIE) study. CALERIE was a randomized controlled trial evaluating the effects of calorie restriction (CR) on the aging process. Here, CR is defined as a dietary regimen which reduces dietary energy intake while maintaining proper nutritional adequacy. Across a wide range of species including yeast, worm, spiders, flies, fish, mice and rats, CR has been shown to increase median and maximum lifespan [15]. In humans, supporting evidence has come from observational studies in longer-lived individuals [16] and from those who self-impose CR [17]. Moreover, pilot studies [18–20] conducted in preparation for the present study pointed to significant short-term effects of CR on resting metabolic rate, cardiovascular and diabetes risk factors, and cognitive function. CALERIE was the first clinical trial to evaluate CR effects in a non-obese population over an extended interval of 24 months. The overall goal was to determine if two years of sustained 25% CR would lead to same improvements in the biomarkers of aging and chronic diseases as seen in the animal studies.

2.2. Study design

CALERIE was a single-blind, multi-center, clinical trial [21]. Healthy individuals from both genders and all races were eligible. Men were restricted to 21–50 years of age, and women from 21 to 47 years of age; body mass index (BMI) was restricted to the range, $22 \leq \text{BMI} < 28\ \text{kg/m}^2$. Participants were assigned at random with 2:1 allocation to the CR intervention or an *ad libitum* control. Randomization was stratified by site, sex and BMI, with BMI dichotomized into normal weight ($22 \leq \text{BMI} < 25\ \text{kg/m}^2$) versus overweight strata ($25 \leq \text{BMI} < 28\ \text{kg/m}^2$). All participants provided written informed consent, the study protocol was approved by Institutional Review Boards at all participating institutions, and oversight was provided by an independent Data and Safety Monitoring Board.

Although there were a number of primary and secondary outcomes, here we focus on the outcomes related to the metabolic adaptation hypothesis for the effects of CR. It postulates that CR reduces metabolic rate more than that predicted by the changes in fat mass and fat-free mass that would ordinarily accompanying weight loss. This results in increased metabolic efficiency that may be accompanied by a drop in core body temperature and a decrease in oxidative stress. Resting metabolic rate (RMR) was measured in all participants at baseline and at months 12 and 24 (M12 and M24); additional measurements were performed in CR intervention arm at months 6 and 18 (M6 and M18). RMR was adjusted for changes in body composition over the intervening interval, and is henceforward referred to as the adjusted RMR. Core body temperature was measured in all participants at baseline, M6, M12 and M24. Data from all available time points were included in the statistical analyses; however, we focus specifically on the results at M12 and M24.

2.3. CR intervention

A complete description of the CR intervention has been provided elsewhere [22]. Briefly, an intensive behavioral approach with appropriate dietary modifications was applied. Participants made dietary selections (under the supervision of intervention staff) that most effectively allowed them to achieve the CR goal. Participants were allowed to vary their dietary choices as needed or desired over the course of the intervention. Psychologists and nutritionists supervised the delivery of the CR intervention in a structured and consistent manner. Behavioral strategies known to be effective in long-term weight-loss studies [23] were applied together with dietary composition changes known to enhance satiety and reduce hunger [24]. In contrast, the controls were not given any specific calorie goal. They received no dietary or behavioral counseling and continued unrestricted in their habitual diets. As such, the control group was conceptualized as the natural history of this study population as it aged over the two years.

2.4. Adherence to the CR intervention

Two hundred, eighteen participants started their assigned intervention, including 143 and 75 in the CR and control groups respectively. Percent weight loss (%WL) from baseline was the primary measure for monitoring adherence to the CR intervention. Based on pilot studies, an algorithm [25] was developed providing the specific %WL profile required to achieve the prescribed level of calorie restriction: 11% at M6, and 15.5% at M12, M18 and M24.

Fig. 1 presents box-and-whisker plots of the %WL observed by CR participants over the four follow-up intervals. The mean ± standard error (s.e.) was 9.9 ± 0.3%, 11.5 ± 0.4%, 11.4 ± 0.4% and 10.4 ± 0.4% at M6, M12, M18 and M24 respectively. The %WL among the controls was 0.8 ± 0.4% over the first 6 months; this subsequently returned to baseline levels for the remaining 18 months of the study (not shown). Between-group differences were all highly significant. Nevertheless, despite the considerable degree of weight loss realized among the CR participants, it fell well short of the %WL profile targeted for this study.

2.5. Intention-to-treat results

The primary analytic vehicle was a Gaussian repeated measures analysis [26,27] with treatment, time, and the treatment × time

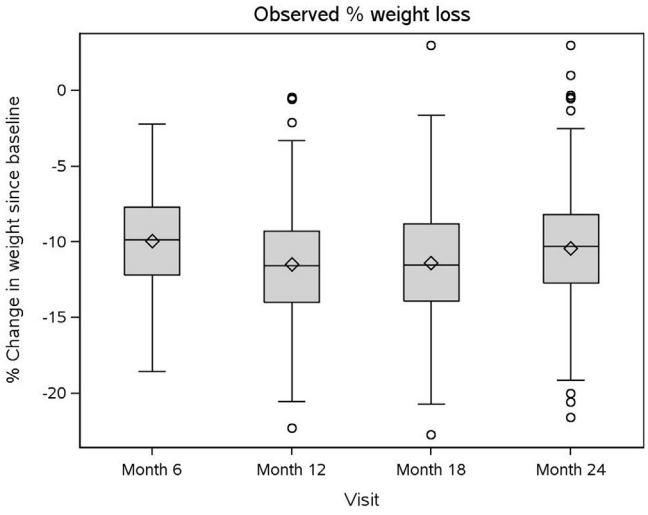

Fig. 1. Box-and-whisker plot of the percent weight change in the CR intervention over the four follow-up intervals.

interaction as independent variables. Design variables, site, sex, and BMI stratum as well as the baseline value of the outcome, were included as covariates to increase precision. Time was treated as a categorical variable, and an unstructured covariance matrix was applied for the repeated observations.

ITT results were reported in Ravussin et al. [28] and Table 1 presents the adjusted mean ± s.e. change for the two metabolic adaptation outcomes by treatment group over time. Small and non-significant changes from baseline were observed in the adjusted RMR in the control group. In the CR group, significant decreases were observed: 48.2 ± 9.2 kcal/d at M12, and 38.2 ± 11.3 kcal/d at M24. The between-group difference was statistically significant at M12 but failed to reach significance at M24. For core temperature, significant decreases of 0.05 ± 0.02 °C were observed in the CR group at both time points; however, because of small and non-significant decreases among the controls, the between-group differences failed to reach statistical significance. Thus, the ITT analysis provided only muted support for the metabolic adaptation hypotheses for the effects of CR.

3. Implementation of the marginal structural model

3.1. Derivation of the weights for the percent weight loss model

The original MSM papers [2,3] provided a detailed review on how to implement the model when adherence is binary. Weights are derived from an ancillary regression model with adherence (Yes/No) as the dependent variable. It is treated as a survival outcome, i.e., the time to becoming non-adherent, and analyzed using a logistic regression model [29]. That is, a longitudinal dataset is created consisting of multiple observations per subject. One proceeds sequentially through the different time intervals. If the subject was adherent in the interval, an observation is added to the dataset with the adherence outcome coded as “Yes”. When the interval during which the subject became non-adherent is reached (if any), an observation is added with the outcome coded as “No.” No further observations are added beyond this interval. A participant who is adherent throughout the study has observations for all the time intervals with adherence always coded as “Yes.”

The dataset is then analyzed using a logistic regression model including a term for the time interval as well as the fixed and time-dependent covariates of interest. The former is treated as a categorical variable effectively adding an intercept for each time interval. Because observations are only added to the dataset if the subject was adherent in the previous interval, one is modeling the logit of the probability of remaining adherent (or, becoming non-adherent) at any visit conditional on being adherent at the previous visits. The model is fit in the usual manner, and the logits are estimated from the estimated regression parameters. The conditional probabilities of remaining adherent are derived by taking the inverse logit transformation; the conditional probability of becoming non-adherent is one minus this quantity. The cumulative product of these conditional probabilities over a set of follow-up visits therefore represents the joint probability of the participant's adherence profile up to that visit. This quantity is inverted and used as the weight for that participant at that visit.

In the case of a continuous confounder, Section 6.2 of Robins et al. [2] suggested that the logistic regression model can replaced by ordinary least-squares regression. Thus, an ancillary regression model among the CR participants was applied using %WL as the dependent variable. A key assumption is sequential ignorability, i.e., no unmeasured confounders conditional on past %WL and covariate

Table 1
Baseline values and changes from baseline in the control and calorie restriction treatment groups from the intention-to-treat analysis and predicted at the targeted %WL profile from the MSM model.

Outcome	Intention-to-treat analysis					Predicted at %WL profile[a]		
	Controls		CR intervention		Between-group p-value[c]	CR intervention		Between-group p-value[c]
	Mean (s.e.)[b]	Within-group p-value[c]	Mean (s.e.)[b]	Within-group p-value[c]		Mean (s.e.)[d]	Within-group p-value[c]	
Adjusted[e] RMR (kcal/d)								
Baseline	1393 (24)		1418 (17)		0.33			
Month 12	−13.9 (12.3)	0.52	−48.2 (9.2)	<0.001	0.04	−57.7 (13.8)	<0.001	0.007
Month 24	−22.6 (14.6)	0.25	−38.2 (11.3)	0.002	0.78	−58.6 (15.0)	<0.001	0.06
Core temperature (°C)								
Baseline	37.0 (0.03)		37.0 (0.02)		0.41			
Δ Month 12	−0.03 (0.02)	0.54	−0.05 (0.02)	0.006	0.70	−0.07 (0.03)	0.02	0.26
Δ Month 24	−0.02 (0.02)	0.64	−0.05 (0.02)	0.02	0.84	−0.06 (0.03)	0.06	0.32

%WL = percent weight loss; MSM = marginal structural model; CR = calorie restricted; RMR = resting metabolic rate.
[a] Percent weight loss profile is: 11% at month 6, and 15.5% at months 12, 18 and 24 (see text).
[b] Baseline values are the observed mean (s.e.); change scores are the least-squares adjusted means (s.e.) from the ITT repeated measures analysis.
[c] Within-group *p*-value tests for a significant change from baseline to the follow-up visit in that group; between-group *p*-value tests for a significant between-group difference in the change score at that visit.
[d] Change scores are the least-squares adjusted means (s.e.) from the marginal structural model.
[e] RMR is adjusted for changes in body composition over time (see text).

history. To address this, the broad list of potential independent variables shown in Table 2 was proposed by the investigators. The general rule was to err on the side of being inclusive rather than being parsimonious. A multi-step process was applied.

1. The analysis dataset pooled all the available observations across the 4 follow-up visits.
2. The design variables, study visit and clinical site, were forced into the model.
3. Expert opinion suggested that age, sex and BMI stratum would be important predictors of %WL. They were forced into the model.
4. As justified in more detail below, the lagged %WL value was forced into the model.
5. In the regression model for a specific outcome (e.g., the adjusted RMR), the lagged value of that outcome was forced into the model
6. For the time-dependent covariates in group 3, both the contemporaneous and lagged values were investigated. The baseline value was used for the lagged value for the M6 observation.
7. All interactions with age, sex, BMI stratum and site were investigated.
8. A stepwise regression model was applied to determine which variables were significantly related to %WL. All predictors significant at $\alpha = 0.05$ plus those forced into the model advanced to the final model.
9. Consistent with good statistical practice, if an interaction term was significant, then the corresponding main effects also advanced irrespective of their own levels of statistical significance.

From the final model, the probability of the observed %WL value for each CR participant at each visit was derived from the Gaussian probability density function. The mean was estimated using the estimated regression parameters while the residual variance was used for the estimate of σ^2. Because the lagged %WL was included as a covariate for the observations at M12, M18 and M24, the probability represented the conditional probability given the previous %WL value (and the other covariates). This was not done for the M6 observation, and the corresponding value represented the marginal %WL probability at M6. The cumulative product of these probabilities over a set of follow-up visits therefore represented the joint probability of the %WL profile up to that visit. For inverse-probability weighting, this quantity was placed in the denominator of the participant- and time-specific weights.

Hernán et al. [3] observed that the derived weights can vary dramatically across participants and over time and may adversely affect the numerical stability of the GEE algorithm. They advocated using stabilized weights instead. Thus, a second ancillary regression model restricted to the time-independent covariates derived above (and the lagged %WL) was applied. The joint probability of the participant's %WL profile up to any visit was again derived as the cumulative product of the probabilities derived at the different time points. This quantity was placed in the numerator of the participant- and time-specific weights. The stabilized weight for each participant at each visit, therefore, was the ratio of the cumulative probabilities from excluding and including the time-dependent covariates.

3.2. Deriving the weights for early discontinuation of the CR intervention

A similar approach was applied to derive the inverse-probability of censoring weights [11], i.e., discontinuing prior to the scheduled end of follow-up. As described in detail above, this was treated as a survival outcome and analyzed using the logistic regression model [29]. A stepwise logistic regression analysis was performed using

Table 2
Independent variables considered for the ancillary regression models.

Category	Variables
1. Demographic	• Age • Sex • BMI stratum • Race • Height • Ethnicity • Marital status • Housing situation • Education • Family income
2. Baseline covariates	
Self-reported nutrition variables	• kcal/day • %fat • %protein • %carbohydrate
Physical activity	• Total minutes of physical activity
Safety markers	• BDI (marker of depression) • MAEDS subscales (markers of eating disorders)
3. Time-dependent covariates	
Self-reported nutrition variables	• kcal/day • %fat • %protein • %carbohydrate
Physical activity	• Total minutes of physical activity
Safety markers	• BDI (marker of depression) • Hemoglobin (marker of anemia) • MAEDS subscales (markers of eating disorders)
Intervention variables	• Percent attendance at individual intervention sessions • Percent attendance at group intervention sessions

BMI = body mass index; kcal = kilocalories; BDI=Beck Depression Inventory; MAEDS = Multi-axial Assessment of Eating Disorder Symptoms.

the same broad list of potential predictor variables outlined in Table 2. All significant predictors plus those forced into the model advanced to the final model. The probability of the participant's disposition at each time point, i.e., either continuing the intervention or discontinuing at that visit, conditional on undertaking the intervention at the previous time point, was derived using the estimated regression parameters and applying the inverse logit transformation. The cumulative product of these probabilities over a set of follow-up visits represented the joint probability of the participant's participation profile up to that follow-up visit. This model was performed with and without the time-dependent covariates, and the stabilized weights were derived as the ratio of the cumulative probabilities from excluding and including the time-dependent covariates.

The final weight for each CR participant at each visit was the product of the stabilized weights from the two sets of ancillary analyses.

4. Results

4.1. Stabilized weights

The final list of predictors for the %WL ancillary model in the analysis of the adjusted RMR is shown in Table 3. The adjusted R^2 value was 0.71 indicating that the predictors collectively accounted for a considerable amount of the variation in %WL. Among the variables forced into the model, neither the lagged value of the adjusted RMR nor study site were statistically significant after accounting for other variables in the model. Age and BMI stratum were significant as main effects and through their interactions with other variables. Two markers from the dietary recall were significant predictors: the lagged number of kilocalories (marginally significant at $p = 0.06$) and the contemporaneous percent calories from carbohydrates. Both the contemporaneous and lagged values of the binge eating subscale from the Multi-axial Assessment of Eating Disorder Symptoms [30] reached statistical significance. We return to the clinical significance of these predictors in the Discussion.

Table 3
Final set of independent variables derived for the ancillary regression model for %WL in the analysis of the adjusted resting metabolic rate.

Variable	*p*-value
Lagged %WL[a]	<0.001
Lagged adjusted RMR[a]	0.52
Study visit[a]	<0.001
Study site[a]	0.97
Age[a]	<0.001
Age × Lagged %WL	0.02
Sex[a]	0.03
BMI Stratum[a]	0.04
BMI stratum × study site	<0.001
BMI stratum × lagged adjusted RMR	0.003
BMI stratum × self-reported kcal − lagged	0.04
Others	
Marital status	<0.001
Housing situation[c]	0.006
Self-reported kcal − lagged	0.06[b]
Pct calories from carbohydrates − contemporaneous	0.04
MAEDS Binge eating subscale − contemporaneous	<0.001
MAEDS Binge eating subscale − lagged	0.02
Pct attendance at group intervention sessions	<0.001

%WL = percent weight loss; RMR = resting metabolic rate; BMI = body mass index; kcal = kilocalories; Pct = percent; MAEDS = Multi-axial Assessment of Eating Disorder Symptoms.
[a] Effect forced into the model.
[b] Main effect included due to a significant interaction.
[c] Housing situation refers to house, apartment, dormitory, etc.

With respect to the discontinuation model, 26 participants discontinued the CR intervention prior to the end of the study including 6 (23.1%) before M6, 8 (30.8%) before M12, 11 (42.3%) before M18 and 1 (3.9%) before M24. The final list of predictors is shown in Table 4. Among the variables forced into the model, study site, age (marginally at $p = 0.07$), sex and the lagged adjusted RMR were significant predictors; BMI stratum was not. One of the macronutrients from the dietary recall was significant, i.e., the percent calories from carbohydrates at baseline, overall and through its interaction with sex. The score on the Beck Depression Inventory [31] at baseline also reached significance through its interaction with sex.

Fig. 2 presents box-and-whisker plots of the final stabilized weights at the four follow-up visits in the analysis of the adjusted RMR. The tops of the whiskers are all below 2.0, with small numbers of larger values. The maximum value of 6.9 was for a participant who withdrew after the M6 visit.

4.2. Metabolic effects predicted by the MSM

The MSM was implemented using a weighted GEE model using the identity link and the Gaussian variance. To ensure a consistent basis for comparing the ITT and MSM results, the same terms described above for the ITT model were included as independent variables. Additionally, linear and quadratic terms in %WL, centered at the targeted %WL profile, were added for the CR group to provide for the association between the change in outcome and %WL achieved. Because controls were not given any specific %WL goal, "adherence" was a meaningless concept in this group (even if it could be calculated). Thus, no MSM adjustments were performed in this group. This also ensures a consistent basis for comparing treatment differences between the ITT and MSM analyses. The linear and quadratic terms in the control group were therefore set to 0, and the participant- and time-specific weights were all set to 1.0.

The unstructured covariance matrix led to convergence problems, and was replaced by compound symmetry supplemented by the robust estimator [32] of the covariance matrix among the regression parameters. A preliminary test to determine if the quadratic trend differed across the four time points was not significant ($p = 0.58$), and the corresponding interaction terms were

Table 4
Final set of independent variables derived for the ancillary logistic regression model for early discontinuation of the CR intervention in the analysis of the adjusted resting metabolic rate.

Variable	*p*-value
Lagged adjusted RMR[a]	0.047
Study visit[a]	0.25
Study site[a]	0.02
Site × pct attendance at group intervention sessions	0.02
Age[a]	0.07
Sex[a]	0.008
Sex × pct calories from carbohydrates - baseline	0.009
Sex × BDI − baseline	0.02
BMI stratum[a]	0.54
Others	
Pct calories from carbohydrates − baseline	0.05
BDI − baseline	0.11[b]
Pct attendance at group intervention sessions	0.005
Ethnicity	0.02

RMR = resting metabolic rate; pct = percent; BDI=Beck Depression Inventory; BMI = body mass index; MAEDS = Multi-axial Assessment of Eating Disorder Symptoms.
[a] Effects forced into the model.
[b] Main effect included due to a significant interaction.

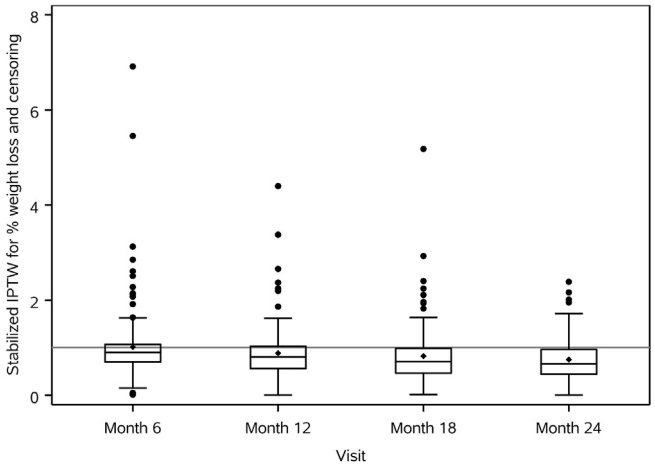

Fig. 2. Box-and-Whisker plots of the final stabilized weights, by time point, in the analysis of the adjusted resting metabolic rate.

removed from the model. In the final model, the quadratic term was marginally significant ($p = 0.07$); however, it was retained in the model to ensure robustness. Fig. 3 provides a scatterplot of the adjusted RMR values at M12 and M24 with the overlaid quadratic curve. It shows that the greatest effect occurred around the targeted %WL level of 15.5%.

Table 1 reports the predicted means ± s.e. change in the adjusted RMR at the targeted %WL profile from the MSM model. A deeper decrease was observed in the adjusted RMR: 57.7 ± 13.8 and 58.6 ± 15.0 kcal/d at M12 and M24, respectively. Both were highly significant; the between-group comparisons reached statistical significance (marginally significant at $p = 0.06$ at M24). This suggests that at the targeted %WL profile, CR would result in an immediate drop in the adjusted RMR which would be sustained with continued CR exposure. The MSM analysis of core temperature, however, was little changed from the ITT analysis. The linear and quadratic terms in %WL failed to reach statistical significance ($p = 0.30$ and 0.70, respectively) indicating that core temperature was largely unresponsive to the actual degree of %WL. The predicted decrease from baseline was 0.07 ± 0.03 and 0.06 ± 0.03 °C at months 12 and 24, respectively. The between-group comparisons continued to be non-significant.

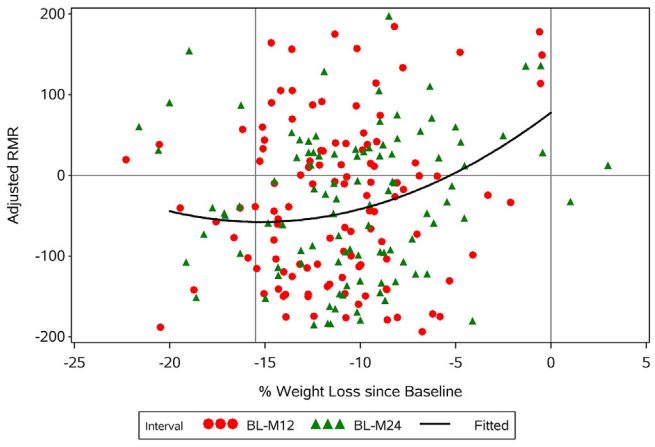

Fig. 3. Scatterplot of the adjusted resting metabolic rate against percent weight loss in the CR group at months 12 and 24 overlaid with the fitted quadratic curve. The vertical reference line corresponds to the targeted weight loss level.

5. Discussion

The MSM was developed in the context of the epidemiologic cohort design which does not enjoy the advantages conferred by randomization in randomized controlled trials. Under specific and detailed assumptions, it adjusts for the confounding effects of covariates observed post randomization. Whereas the value of the intention-to-treat analysis is that it anticipates the effect of the intervention if it were introduced into general practice, the value of the MSM is that it anticipates the effect of the intervention under the specific counterfactual condition, i.e., following the targeted adherence profile.

In the context of clinical trials, the credibility of the MSM analysis is critically dependent on the degree to which the counterfactual condition can actually be realized. Given the limited long-term success of weight-loss studies, there was never any ambition of introducing 25% CR into mainstream dietary practice. Rather, the purpose of the study was to determine if the physiologic effects observed in carefully controlled animal studies would be replicated in humans. CALERIE anticipated that the strict adherence imposed in animal studies would not be realized in human trials. Causal modeling, therefore, provided the only alternative to address this scientific question, and was written directly into the protocol.

The original MSM papers were focused on a binary confounder using the Bernoulli distribution. A recent monograph from the SAS Institute [33] provides a good overview of the model and some programming code from which to start. Here, we have provided a detailed case study on the application of this model for a continuous confounding measure using the Gaussian distribution. All MSM applications have emphasized that there are a number of extra steps and careful attention to detail is required, and this was no less true here. The construction of the ancillary models predicting adherence and withdrawal from the intervention are critical to deriving the inverse-probability weighting. Care must also be taken to identify which variables to force into the model, how to construct interaction terms, and how to address time-dependent covariates. In the weighted GEE model, the unstructured covariance matrix led to convergence problems, and some dexterity is required to overcome these types of problems. To satisfy the assumption of no unmeasured confounders, the study must anticipate and collect all the data predictive of these measures. A careful review of the literature and expert opinion among the clinical investigators are important first steps. Nevertheless, one can never be certain that all such predictors have been identified. Moreover, the weights are estimated from the ancillary regression models and there is sampling variability associated with the derived values. Consistent with other applications of the MSM model [4–11], however, this variability was not taken into consideration in our analysis. It would be worthwhile developing techniques to reflect this uncertainty using, for example, bootstrapping procedures.

Despite this, the exercise is well worth the additional effort and it provided important insight into the physiologic mechanisms at play in calorie restriction. The MSM lent additional support to the metabolic adaptation theory of calorie restriction. Adjusted RMR was sensitive to the level of %WL achieved. At full adherence, an immediate drop in the adjusted RMR is predicted which would be sustained with continued CR exposure. On the other hand, core temperature was largely unresponsive to %WL. At full adherence, significant decreases are predicted, but would fail to exceed those in the control group. The implications are that CR at the prescribed and ideal amount would have replicated findings already observed in animal models. Future work will apply this model to investigate other pathways suggested for the effects of CR. They include the pituitary-thyroid axis as measured by triiodothyronine, thyroid

stimulating hormone and catecholamines, and an inflammatory model measured by tumor necrosis factor alpha, C-reactive protein, other proinflammatory cytokines.

Although beyond the scope of this paper, the ancillary models provided insight into the factors which affected the %WL observed. Not unexpectedly, greater %WL was associated with a lower calorie intake, a larger percent of calories from carbohydrates and greater attendance at the group intervention sessions. Married, male sex and older age were also positively associated. These variables may prove useful as adherence markers in future studies.

Conflicts of interest

None.

Authors' contributions

All authors were responsible for identifying the research question, designing the study, obtaining ethics approval, the acquisition of funding, and the analysis and interpretation of data. All authors were responsible for the drafting of this manuscript, and have read and approved the final version.

Acknowledgements

This work was supported by the National Institutes of Health [grant numbers U01AG022132, U01AG020478, U01AG020487, U01AG020480 and P30AG028716]. Consistent with the NIH cooperative agreement mechanism, NIA staff participated with the principal investigators as equal members on the CALERIE Steering Committee and its various subcommittees. Although they reviewed the manuscript prior to submission, they exercised no editorial control over its form or content.

References

[1] P. Peduzzi, J. Wittes, K. Detre, et al., Analysis as-randomized and the problem of non-adherence: an example from the veterans affairs randomized trial of coronary bypass surgery, Stat. Med. 12 (1993) 1185–1195.
[2] J.M. Robins, M.A. Hernán, B. Brumback, Marginal structural models and causal inference in epidemiology, Epidemiology 11 (2000) 550–560.
[3] M.A. Hernán, B. Brumback, J.M. Robins, Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men, Epidemiology 11 (2000) 561–570.
[4] K.L. Kahn, J.L. Adams, J.C. Weeks, et al., Adjuvant chemotherapy use and adverse events among older patients with stage III colon cancer, JAMA 303 (2010) 1037–1045.
[5] F.J. Palella Jr., C. Armon, K. Buchacz, et al., The association of HIV susceptibility testing with survival among HIV-infected patients receiving antiretroviral therapy: a cohort study, Ann. Intern. Med. 151 (2009) 73–84.
[6] J.A. Sterne, M.A. Hernán, B. Ledergerber, et al., Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study, Lancet 366 (2005) 378–384.
[7] S. Toh, S. Hernández-Díaz, R. Logan, et al., Coronary heart disease in postmenopausal recipients of estrogen plus progestin therapy: does the increased risk ever disappear? A randomized trial, Ann. Intern. Med. 152 (2010) 211–217.
[8] T. Yamaguchi, Y. Ohashi, Adjusting for differential proportions of second line treatment in cancer clinical trials. Part I: structural nested models and marginal structural models to test and estimate treatment arm effects, Stat. Med. 23 (2004) 1991–2003.
[9] S. Toh, S. Hernández-Díaz, R. Logan, et al., Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomization, Epidemiology 21 (2010) 528–539.
[10] L.E. Cain, S.R. Cole, Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident AIDS or death, Stat. Med. 28 (2009) 1725–1738.
[11] J.M. Robins, D.M. Finkelstein, Correcting for non-compliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests, Biometrics 56 (2000) 779–788.
[12] P.W. Holland, Statistics and causal inference (with discussion), J. Am. Stat. Assoc. 81 (1986) 945–970.
[13] S.L. Zeger, K.-Y. Liang, Longitudinal data analysis for discrete and continuous outcomes, Biometrics 42 (1986) 121–130.
[14] P.R. Rosenbaum, D.B. Rubin, The central role of the propensity score in observational studies for causal effects, Biometrika 70 (1983) 41–55.
[15] L. Fontana, L. Partridge, V.D. Longo, Extending healthy life span – from yeast to humans, Science 328 (2010) 321–326.
[16] D.C. Willcox, B.J. Willcox, Q. He, et al., They really are that old: a validation study of centenarian prevalence in Okinawa, J. Gerontol. A Biol. Sci. Med. Sci. 63 (2008) 338–349.
[17] T.E. Meyer, S.J. Kovacs, A.A. Ehsani, et al., Long-term caloric restriction ameliorates the decline in diastolic function in humans, J. Am. Coll. Cardiol. 47 (2006) 398–402.
[18] L.K. Heilbronn, L. de Jonge, M.I. Frisard, et al., Effect of 6-month calorie restriction on biomarkers of longevity, metabolic adaptation, and oxidative stress in overweight individuals: a randomized controlled trial, JAMA 295 (2006) 1539–1548.
[19] S.K. Das, C.H. Gilhooly, J.K. Golden, et al., Long-term effects of 2 energy-restricted diets differing in glycemic load on dietary adherence, body composition, and metabolism in CALERIE: a 1-y randomized controlled trial, Am. J. Clin. Nutr. 85 (2007) 1023–1030.
[20] S.B. Racette, E.P. Weiss, D.T. Villareal, et al., One year of caloric restriction in humans: feasibility and effects on body composition and abdominal adipose tissue, J. Gerontol. A Biol. Sci. Med. Sci. 61 (2006) 943–950.
[21] J. Rochon, C.W. Bales, E. Ravussin, et al., Design and conduct of the CALERIE study: comprehensive assessment of the long-term effects of reducing intake of energy, J. Gerontol. A Biol. Sci. Med. Sci. 66 (2011) 97–108.
[22] A.D. Rickman, D.A. Williamson, C.K. Martin, et al., The CALERIE Study: design and methods of an innovative 25% caloric restriction intervention, Contemp. Clin. Trials 32 (2011) 874–881.
[23] T.A. Wadden, M.L. Butryn, K.J. Byrne, Efficacy of lifestyle modification for long-term weight control, Obes. Res. 12 (Suppl.) (2004) 151S–162S.
[24] N.C. Howarth, E. Saltzman, S.B. Roberts, Dietary fiber and weight regulation, Nutr. Rev. 59 (2001) 129–139.
[25] C. Pieper, L. Redman, S. Racette, et al., Development of adherence metrics for caloric restriction interventions, Clin. Trials 8 (2011) 155–164.
[26] R.I. Jennrich, M.D. Schluchter, Unbalanced repeated-measures models with structured covariance matrices, Biometrics 42 (1986) 805–820.
[27] P.J. Diggle, P.J. Heagerty, K.-Y. Liang, S.L. Zeger, Analysis of Longitudinal Data, second ed., Oxford University Press, New York, 2002.
[28] E. Ravussin, L.M. Redman, J. Rochon, et al., A two-year randomized controlled trial of human caloric restriction: feasibility and effects on predictors of health span and longevity, J. Gerontol. A Biol. Sci. Med. Sci. 70 (2015) 1097–1104.
[29] E. Efron, Logistic regression, survival analysis, and the Kaplan-Meier curve, J. Am. Stat. Assoc. 83 (1988) 414–425.
[30] D.A. Anderson, D.A. Williamson, E.G. Duchmann, et al., Development and validation of a multifactorial treatment outcome measure for eating disorders, Assessment 6 (1999) 7–20.
[31] A.T. Beck, A. Beamesderfer, Assessment of depression: the depression inventory, Mod. Probl. Pharmacopsychiatry 7 (1974) 151–169.
[32] R.M. Royall, Model robust confidence intervals using maximum likelihood estimators, Int. Stat. Rev. 54 (1986) 221–226.
[33] D.E. Faries, Z.A. Kadziola, Analysis of longitudinal data using marginal structural models, in: D.E. Faries, A.C. Leon, J.M. Haro, et al. (Eds.), Analysis of Observational Heath Care Data Using SAS, SAS Institute, Cary, NC, 2010, pp. 211–230.