

RESEARCH

Open Access



# Effects of differential measurement error in self-reported diet in longitudinal lifestyle intervention studies

David Aaby\*  and Juned Siddique

## Abstract

**Background:** Lifestyle intervention studies often use self-reported measures of diet as an outcome variable to measure changes in dietary intake. The presence of measurement error in self-reported diet due to participant failure to accurately report their diet is well known. Less familiar to researchers is *differential* measurement error, where the nature of measurement error differs by treatment group and/or time. Differential measurement error is often present in intervention studies and can result in biased estimates of the treatment effect and reduced power to detect treatment effects. Investigators need to be aware of the impact of differential measurement error when designing intervention studies that use self-reported measures.

**Methods:** We use simulation to assess the consequences of differential measurement error on the ability to estimate treatment effects in a two-arm randomized trial with two time points. We simulate data under a variety of scenarios, focusing on how different factors affect power to detect a treatment effect, bias of the treatment effect, and coverage of the 95% confidence interval of the treatment effect. Simulations use realistic scenarios based on data from the Trials of Hypertension Prevention Study. Simulated sample sizes ranged from 110–380 per group.

**Results:** Realistic differential measurement error seen in lifestyle intervention studies can require an increased sample size to achieve 80% power to detect a treatment effect and may result in a biased estimate of the treatment effect.

**Conclusions:** Investigators designing intervention studies that use self-reported measures should take differential measurement error into account by increasing their sample size, incorporating an internal validation study, and/or identifying statistical methods to correct for differential measurement error.

**Keywords:** Bias, Sample size, Coverage, Behavior, Intervention, Clinical trial

## Introduction

Lifestyle intervention studies—which aim to change a participant’s weight or eating behavior—often use self-reported measures of diet, such as interviewer-assisted 24-hour dietary recalls or food frequency questionnaires. These measures are prone to error for various reasons including poor quantification of portion sizes and social desirability [1]. More reliable and accurate measures, such

as recovery biomarkers, require a 24-hour urine collection, and are expensive and cumbersome for participants to collect. Thus self-reported measures are commonly used as a surrogate for the true quantity of interest.

Measurement error in intervention studies can result in biased estimates of the treatment effect and reduced power to detect treatment effects [2]. Measurement error interferes with analyses to determine if an intervention is effective and limits the ability of researchers to design and implement effective interventions to reduce behavioral risk factors. Thus, the presence of measurement error may

\*Correspondence: [david.aaby@northwestern.edu](mailto:david.aaby@northwestern.edu)

Department of Preventive Medicine, Northwestern University, Chicago, IL



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

lead researchers to adopt or discard interventions that are actually (in)effective.

Most measurement error research has focused primarily on problems associated with measurement error in predictor variables [3], particularly those situations where an exposure is measured with error, thus attenuating or distorting the relationship between exposure and outcome. Less work has been done investigating the implications of measurement error in outcome variables in a longitudinal intervention setting. Longitudinal dietary intervention studies involve repeated dietary assessments over time and produce unique measurement error issues that are not encountered in cross-sectional studies. Participants may modify their reporting behavior to appear compliant with dietary recommendations of the study [4], or they may attempt to reduce interview duration and reporting difficulty during follow-up assessments by omitting items or by erroneously reporting foods that are easier to measure or describe [5]. Alternatively, their accuracy may improve over time due to training in portion size assessment and a more general awareness of their dietary intake [6, 7].

Differential measurement error is where the nature of measurement error (bias and precision) differs over time and/or by treatment condition [8, 9]. In terms of bias, participants may (1) become more accurate in their reports of diet due to improved self-monitoring; (2) misreport their diet in order to appear compliant with the intervention; or (3) report with the same accuracy as seen at baseline. Similarly, the precision of reporting may increase, decrease, or stay the same as at baseline. These changes in bias and precision may differ by treatment condition. While a biased treatment effect is clearly undesirable, reduced power due to additional variability is not a trivial matter in lifestyle interventions, where effect sizes tend to be small and the additional variation due to measurement error can result in a failure to detect a treatment effect.

In this paper, using the setting of a longitudinal clinical trial, we use simulation to derive how various forms of differential measurement error influence sample size, bias, and coverage of the 95% confidence interval when estimating treatment effects. We provide recommendations for investigators when designing intervention trials that use dietary intake as an outcome variable.

## Methods

In this section, we describe a simulation study to assess the consequences of outcome measurement error and differential measurement error on the ability to estimate treatment effects. Our simulation-based scenarios reflect the settings of intervention studies where diet-based outcomes are measured repeatedly over time in both a treatment and control group. The analysis model is a covariance pattern regression model [10] where the outcome is modeled as a function of time, a treatment by time

interaction, and where an unstructured covariance matrix is used to estimate the variance-covariance parameters.

We examine 1) Differential measurement error with respect to time which is reflected in differences in measurement error variability between baseline and follow-up as well as differences in over/under reporting at baseline as compared to follow-up; 2) Differential measurement error with respect to treatment condition, with participants in the treatment group having different measurement error (variability and bias) as compared to those in the control group. To explore these settings, we simulate data under a variety of scenarios, focusing on how different factors influence: the power to detect the treatment effect, the bias of the treatment effect, and the coverage of the 95% confidence interval of the treatment effect.

Let  $z_{ij}$  be the true value (i.e. true dietary intake) of the quantity we wish to measure on participant  $i$ ,  $i = 1, \dots, N$  at time  $j$ ,  $j = 0, 1$ . This quantity is assumed to be measured without error. Let  $y_{ij}$  be the observed outcome of interest measured with error. That is,  $y_{ij}$  is  $z_{ij}$  measured with error, such as a self-reported dietary measure. Let  $d_i$  be an indicator as to whether a participant has been randomized to the treatment group ( $d_i = 1$ ) or the control group ( $d_i = 0$ ). The variable  $t_{ij}$  indicates the time points at which the quantities are measured, a baseline measurement ( $t_{ij} = 0$ ) and a follow-up measurement ( $t_{ij} = 1$ ). Finally, let  $n_d$  denote the sample size in each treatment and control group.

The distribution of  $z_{ij}$  has the following form:

$$z_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 (t_{ij} \times d_i) + \varepsilon_{ij} \quad (1)$$

where  $\beta_0$  is an intercept term and  $\beta_1$  is the effect of time. We assume no differences in intervention conditions at baseline, and thus do not include a main effect for treatment. The regression coefficient  $\beta_2$  is the estimand of interest, the expected true difference in change over time between the two treatment conditions. In all of our simulations, we fix the values of the regression coefficients in Eq. (1) at the values listed in Table 1. That is,  $\beta_0 = 8.21$ ,  $\beta_1 = -0.037$ , and  $\beta_2 = -0.25$ . These values are based on data from the Trials of Hypertension Prevention (TOHP) study (described below).

In Eq. (1),  $\varepsilon_{ij}$  is a random error term with distribution  $\varepsilon_{ij} \sim N(0, \Sigma_z)$ . The variance-covariance matrix  $\Sigma_z$  is

$$\Sigma_z = \begin{bmatrix} \sigma_z^2 & \rho\sigma_z^2 \\ \rho\sigma_z^2 & \sigma_z^2 \end{bmatrix} \quad (2)$$

where  $\sigma_z^2$  is the variance at baseline and follow-up and  $\rho$  is the correlation between baseline and follow-up. Again, these values are listed in Table 1 and are fixed across all simulation scenarios ( $\sigma_z^2 = 0.17$ ,  $\rho = 0.5$ ).

**Table 1** Simulation Study Parameter Values

Parameter	TOHP Value	Range of Simulation Values	Equation
$\beta_0$	8.21	Fixed	2.1
$\beta_1$	-0.037	Fixed	2.1
$\beta_2$	-0.25	Fixed	2.1
$\sigma_z^2$	0.17	Fixed	2.2
$\rho$	0.5	Fixed	2.2
$\gamma_0$	5.29	Fixed	2.3
$\gamma_1$	0.09	Fixed	2.3
$\gamma_2$	0.33	1, 0.33	2.3
$\gamma_3$	-0.006	-1.5 to 1.5 by 0.01	2.3
$\gamma_4$	-0.034	-0.10 to 0.05 by 0.001	2.3
$\lambda_1$	1.86	Fixed	2.4
$\lambda_2$	1.17	0.5 to 2.5 by 0.1	2.4
$\lambda_3$	1.32	0.5 to 2.5 by 0.1	2.4

To generate  $y_{ij}$  we add measurement error to the  $z_{ij}$  values as follows:

$$y_{ij} = \gamma_0 + \gamma_1 (d_i \times t_{ij}) + \gamma_2 z_{ij} + \gamma_3 (z_{ij} \times t_{ij}) + \gamma_4 (z_{ij} \times t_{ij} \times d_i) + \delta_{ij} \tag{3}$$

where  $\gamma_0$  is an intercept term reflecting the overall difference between self-report and true intake at baseline for a given value of  $z_{ij}$ ,  $\gamma_1$  is the additional change in intercept between treatment conditions at follow-up,  $\gamma_2$  is the slope of the regression of  $y_{ij}$  on  $z_{ij}$  at baseline that reflects how  $y_{ij}$  varies as a function of true intake at baseline,  $\gamma_2 + \gamma_3$  is the slope in the control group at time 1, and  $\gamma_4$  is the difference in slopes between the treatment and control groups at time 1.

The error term  $\delta_{ij}$  in Eq. (3) is normally distributed with  $\delta_{ij} \sim N(0, \Sigma_y)$ . The variance-covariance matrix  $\Sigma_y$  is

$$\Sigma_y = \begin{bmatrix} \lambda_1 \sigma_z^2 & \rho \sigma_z^2 \\ \rho \sigma_z^2 & \lambda_1 \lambda_2 \lambda_3^d \sigma_z^2 \end{bmatrix} \tag{4}$$

where  $\lambda_1$  is a factor for inflating variance at every time point,  $\lambda_2$  is a factor for inflating variance at follow-up, and  $\lambda_3$  is a factor for inflating variance in the treatment group at follow-up. Eq. (4) allows for repeated measures on participant  $i$  to be correlated and can also allow for heterogeneous variances.

Equations (1) through (4) provide a very flexible framework for simulating data and incorporate a number of scenarios for simulating differential measurement error. Table 2 summarizes—in terms of our model—various types of measurement error and how the parameters are set or varied in our simulation scenarios. For example, when the parameters  $\gamma_0, \gamma_1, \gamma_3, \gamma_4 = 0$  and  $\gamma_2 = 1$  in Eq. (3),  $y_{ij}$  follows a classic measurement error model, where  $y_{ij}$  is an unbiased measure of  $z_{ij}$ , but measured with additional variability (Table 2, row 1). We focus on

the last three rows of Table 2: differential measurement with respect to time, differential measurement error with respect to treatment, and differential measurement error with respect to time and treatment.

In these scenarios, the parameters  $\gamma_3$  and  $\gamma_4$  allow for changes in bias at follow-up and by treatment condition, respectively. The parameters  $\lambda_2$  and  $\lambda_3$  allow for additional variability at follow-up and within the treatment group, respectively.

Using this simulation framework, and varying the sample size as well as the parameters  $\gamma_3, \gamma_4, \lambda_2,$  and  $\lambda_3$  in Eqs. (3) and (4), we simulate data under a variety of scenarios. Each set of simulations is centered at non-differential measurement error (Table 2, row 2). We then expand our simulations around this central assumption to investigate how differential measurement error impacts estimates of the treatment effect.

To ensure that we are simulating realistic scenarios, we calibrate the simulation parameter values in Eqs. (1) through (4) using data on sodium intake from the Trials of Hypertension Prevention Study (TOHP), a randomized controlled trial of 2811 participants who received lifestyle interventions and nutritional supplement interventions for hypertension prevention [11]. TOHP collected 24-hour recalls as well as urinary sodium on 744 participants, both at baseline and at follow-up. This allows us to posit realistic values for the parameters involving true intake in Eqs. (1) and (2) as well for the parameters involving measurement error in Eqs. (3) and (4).

Table 1 summarizes the parameters estimated from the TOHP data as well as the varying values used in the simulations. The true treatment effect,  $\beta_2$  in Eq. (1), is -0.25 on the log scale so that at follow-up, participants in the treatment condition have sodium intake  $(1 - \exp(-0.25)) \times 100 = 22\%$  less than those in the control group.

We define the *naive* treatment effect as the difference in change from baseline between treatment and control groups using the error prone self-reported values  $y$ . Using Eq. (3), the naive treatment effect is given by:

$$\begin{aligned} \Psi^{naive} &= E(y_1 - y_0 \mid d_i = 1) - E(y_1 - y_0 \mid d_i = 0) \\ &= \gamma_1 + \beta_2(\gamma_2 + \gamma_3) + \gamma_4(\beta_0 + \beta_1 + \beta_2). \end{aligned} \tag{5}$$

See **Bias** in the appendix for details. An estimate of the treatment effect is unbiased when  $\Psi^{naive} - \beta_2 = 0$ .

**Variability simulation**

We estimated the naive treatment effect in (5) under a variety of simulation scenarios by varying the parameters in Table 1. We examined how simultaneously increasing measurement error variability at follow-up ( $\lambda_2$ ) and increasing measurement error variability in the treatment condition at follow-up ( $\lambda_3$ ) increases the required sample size to achieve 80% power. We assume that the trial was

**Table 2** Different measurement error scenarios described in terms of Eqs. (3) and (4) where  $z_{ij}$  is true intake for participant  $i$  at time  $j$ .  $y_{ij}$  is  $z_{ij}$  measured with error

Type of Measurement Error	Model in terms of Equations (3) and (4)	Varying sim parameters (bias)	Varying sim parameters (variance)
Classical ME	$y_{ij} = z_{ij} + \delta_{ij}$	NA	NA
Systematic ME (non-DME)	$y_{ij} = \gamma_0 + \gamma_2 z_{ij} + \delta_{ij}$	NA	NA
DME w.r.t. time	$y_{ij} = \gamma_0 + \gamma_2 z_{ij} + \gamma_3 (z_{ij} \times t_{ij}) + \delta_{ij}$	$\gamma_3$	$\lambda_2$
DME w.r.t. tx	$y_{ij} = \gamma_0 + \gamma_1 (d_i \times t_{ij}) + \gamma_2 z_{ij} + \gamma_4 (z_{ij} \times t_{ij} \times d_i) + \delta_{ij}$	$\gamma_4$	$\lambda_3$
DME w.r.t. time and tx	$y_{ij} = \gamma_0 + \gamma_1 (d_i \times t_{ij}) + \gamma_2 z_{ij} + \gamma_3 (z_{ij} \times t_{ij}) + \gamma_4 (z_{ij} \times t_{ij} \times d_i) + \delta_{ij}$	$\gamma_3, \gamma_4$	$\lambda_2, \lambda_3$

Abbreviations: ME measurement error; DME differential measurement error; w.r.t. with respect to; tx treatment; NA not applicable

powered assuming non-differential measurement error based on existing self-reported data. Thus, the parameter for increased variability across all participants regardless of time point or treatment condition ( $\lambda_1$ ) was fixed across all scenarios and equal to 1.86. Calculation of power was based on a two-sample z-test, as defined in [Power and sample size](#) of the Appendix.

**Treatment effect simulation**

Next, we examined how simultaneously varying the change in slope for the control group at follow-up ( $\gamma_3$ ) and varying the change in slope for the treatment group at follow-up ( $\gamma_4$ ) increases the bias of the treatment effect in terms of the percent increase in the bias of the treatment effect. We fix the parameters  $\gamma_1$  and  $\gamma_2$  to the TOHP values displayed in Table 1. Thus only the  $\gamma$  parameters that affect measurement error differentially (i.e.  $\gamma_3$  and  $\gamma_4$ ) influence the percent increase in bias of the treatment effect. In our setting, an increase in slope results in greater self-reported values at follow-up as compared to baseline (or the control group) for a fixed value of true intake.

**Coverage simulation**

Finally, we varied both the differential measurement error parameters in Table 2 affecting bias ( $\gamma_3, \gamma_4$ ) and the differential measurement error parameters that affect variance ( $\lambda_2, \lambda_3$ ) to generate different combinations of high/low bias and high/low variance. The parameters that do not affect differential measurement error ( $\gamma_1, \gamma_2, \lambda_1$ ) were fixed at their TOHP values. We calculated the naive treatment effect and its 95% confidence interval ([Appendix Coverage](#)). To compare these scenarios to each other and that based on true intake, we display our results using a forest plot.

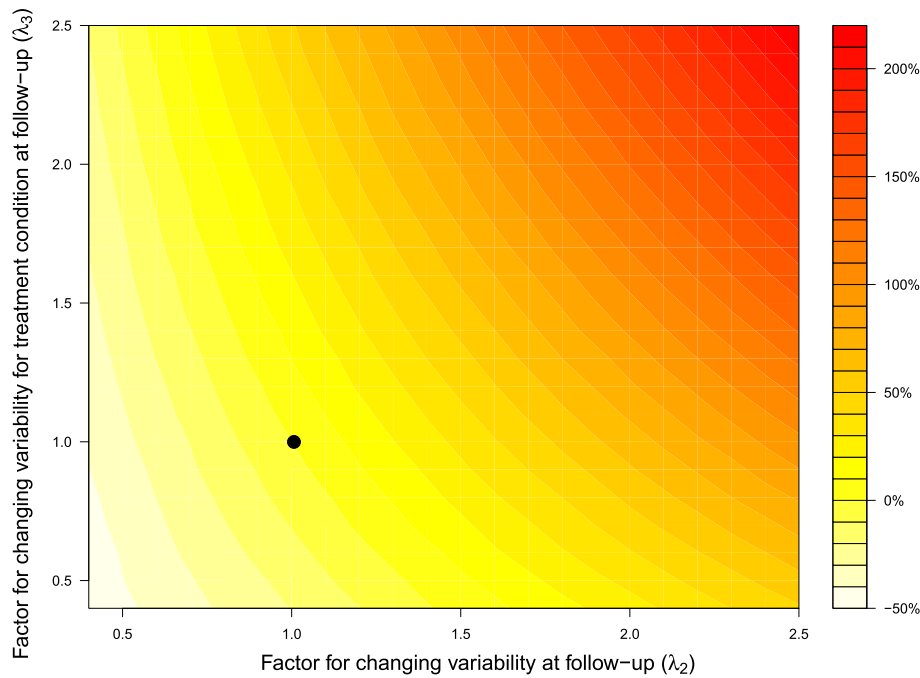
The coverage probability of a confidence interval is the proportion of the time that the interval contains the true

quantity of interest  $\beta_2$ . Coverage can be affected by both bias and variability and as a result, provides a good summary of how different parameters affecting measurement error can impact estimates of the treatment effect. Let  $\Psi_{lower}$  and  $\Psi_{upper}$  be the lower and upper endpoints of a 95% confidence interval of an estimate of the naive treatment effect. Ideally, an estimator exhibits *nominal* coverage, such that the coverage of its 95% confidence interval is also 95%. We calculate the coverage of the naive treatment effect as the probability that the true treatment effect lies within the 95% confidence interval of the naive treatment effect. Details are in [Coverage](#) of the appendix.

**Results**

Figure 1 is a contour plot of the percent increase in sample size needed to achieve 80% power to detect a treatment effect. The x-axis displays values for the measurement error parameter for the additional variability at follow-up ( $\lambda_2$ ). The y-axis displays values for the measurement error parameter for increasing variability for the treatment condition at follow-up ( $\lambda_3$ ). As these parameters increase, so does the sample size needed to achieve 80% power.

Using estimates from the TOHP data, under a scenario of no increased variability at follow-up in both the treatment and control conditions ( $\lambda_2 = 1, \lambda_3 = 1$ ), the sample size needed to achieve 80% power is  $n = 117$  per group (indicated by the black dot in Fig. 1). Differential measurement error with respect to time ( $\lambda_2 > 1, \lambda_3 = 1$ ) has a greater impact on required sample size to achieve 80% power than does differential measurement error with respect to treatment ( $\lambda_2 = 1, \lambda_3 > 1$ ). For example, under a scenario where there is additional variability at follow-up ( $\lambda_2 = 2$ ) but no additional variability for treatment condition ( $\lambda_3 = 1$ ), the sample size must increase by 65.0% in order to achieve 80% power, which corresponds to a sample size of  $n = 193$  per group. For scenarios where there is



**Fig. 1** Contour plot for the percent change in sample size required to achieve 80% power to detect a fixed treatment effect across a range of parameters that change measurement error variability differentially with respect to follow-up and treatment condition at follow-up. The x-axis displays values for the parameter for changing variability at follow-up ( $\lambda_2$ ), the y-axis displays values for the parameter for changing variability for the treatment condition at follow-up ( $\lambda_3$ ). The point plotted at (1,1) corresponds to 0% change in sample size due to no change in variability at follow-up or treatment condition at follow-up

no additional variability at follow-up in the control group ( $\lambda_2 = 1$ ) but additional variability only in the treatment condition at follow-up, ( $\lambda_3 = 2$ ), the sample size must increase by only 32.5%, ( $n = 155$  per group). For situations where there is both increased variability at follow-up and treatment condition at follow-up, ( $\lambda_2 = 2, \lambda_3 = 2$ ), the sample size must increase by 130.8%, which corresponds to a sample size of  $n = 270$  per group. Under scenarios of decreased variability, the required sample size decreases. For example, when  $\lambda_2 = .5$  and  $\lambda_3 = .5$ , the sample decreases by 40% ( $n = 70$  per group).

Figure 2 is a contour plot of the percent increase in bias of the treatment effect for varying values of  $\gamma_3$ —the measurement error parameter for change in slope for the control group (x-axis), and  $\gamma_4$ —the additional change in slope for the treatment group (y-axis). As measurement error increases in the intervention groups, so does the bias of the treatment effect.

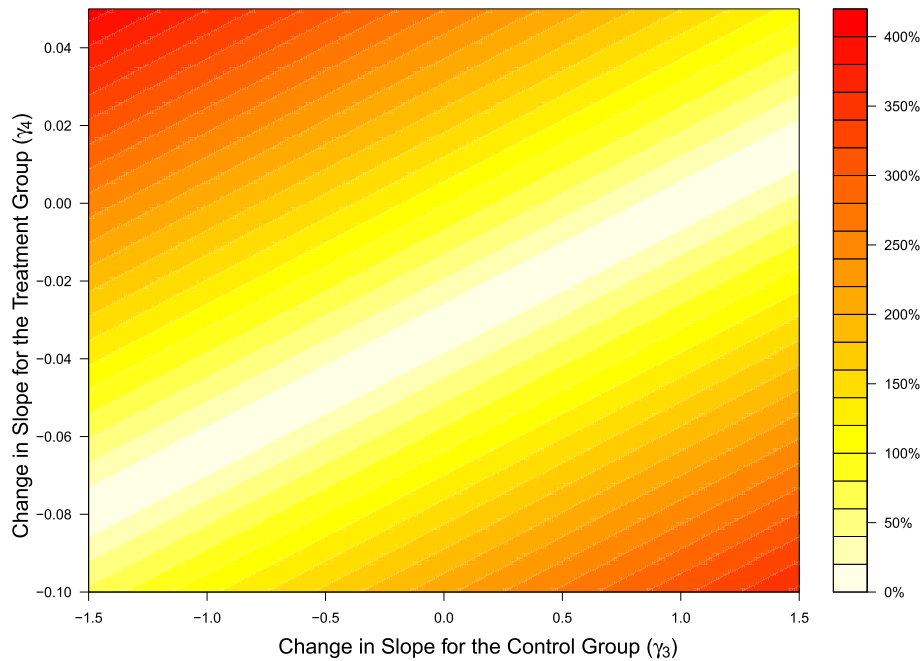
Unlike the parameters governing variance, here differential measurement error with respect to treatment ( $\gamma_4$ ) does have a substantial effect. For example, when there is no additional change in slope for the treatment group at follow-up ( $\gamma_4 = 0$ ), and a small increase in slope for the control group ( $\gamma_3$  increases from 1.02 to 1.10), then the bias increases by only 8%, so that the naive treatment effect reflects a 23.7% reduction in sodium intake in the

treatment group versus the control group (as compared to the true treatment effect of a 22% reduction). When there is no additional change in slope for the control group at follow-up ( $\gamma_3 = 0$ ) and a small increase in slope for the treatment group at follow-up ( $\gamma_4$  increases from -0.032 to -0.05), the bias increases by 56.5% (a naive treatment effect of 32.4%). For an increase in slope for both the control group ( $\gamma_3 = 1.05$ ) and the treatment group ( $\gamma_4 = -0.05$ ), the bias increases by 161.5% (treatment participants have 48% less sodium at follow-up as compared to control participants).

Figure 3 displays a forest plot of estimates of the naive treatment effect, their associated 95% confidence intervals, and the coverage of the confidence interval of the treatment effect under a range of different measurement error parameters. The vertical solid line shows the true treatment effect of -0.25. The vertical dotted line at 0 indicates no treatment effect.

Under classical measurement error, the estimate of the treatment effect is unbiased, but has increased variability. The (+) and (-) refer to whether the  $\gamma_3$  and  $\gamma_4$  parameters governing measurement error in Table 2 are greater than or less than 0, respectively. Bias in the treatment effect as well as increased variability occurs in systematic measurement error and differential measurement error with respect to time, treatment, or both time and treatment.



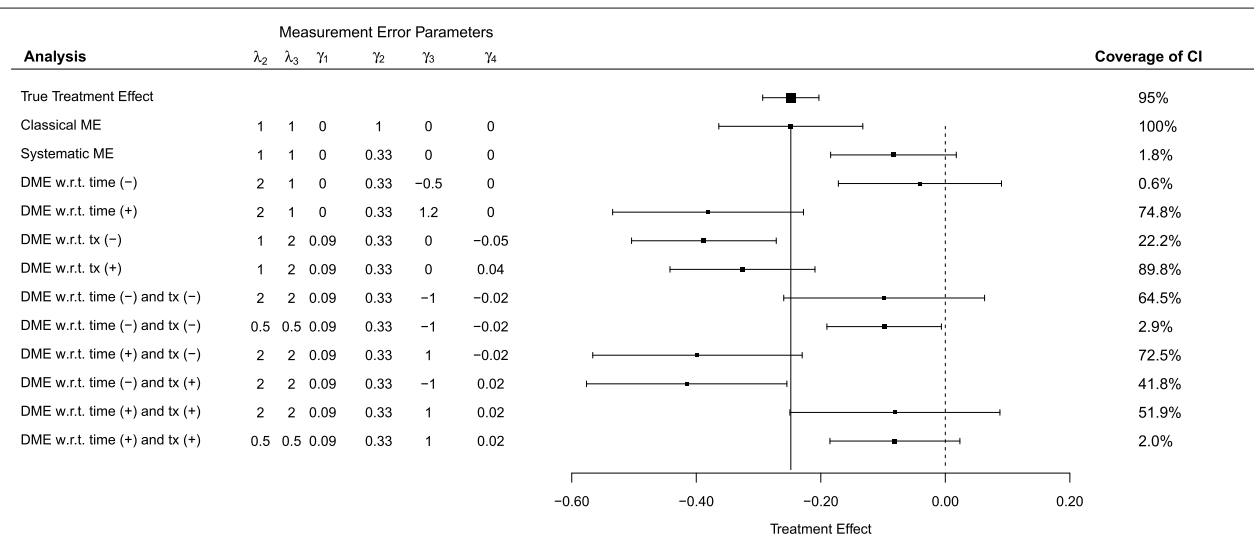


**Fig. 2** Contour plot for the percent increase in bias across a range of parameters for the slope of the treatment effect. The x-axis displays values for the measurement error parameter for the change in slope for the control group ( $\gamma_3$ ), the y-axis displays values for the measurement error parameter for the change in slope for the treatment group ( $\gamma_4$ ). Increases in differential measurement error with respect to treatment have a stronger impact on bias than do increases in differential measurement error with respect to time

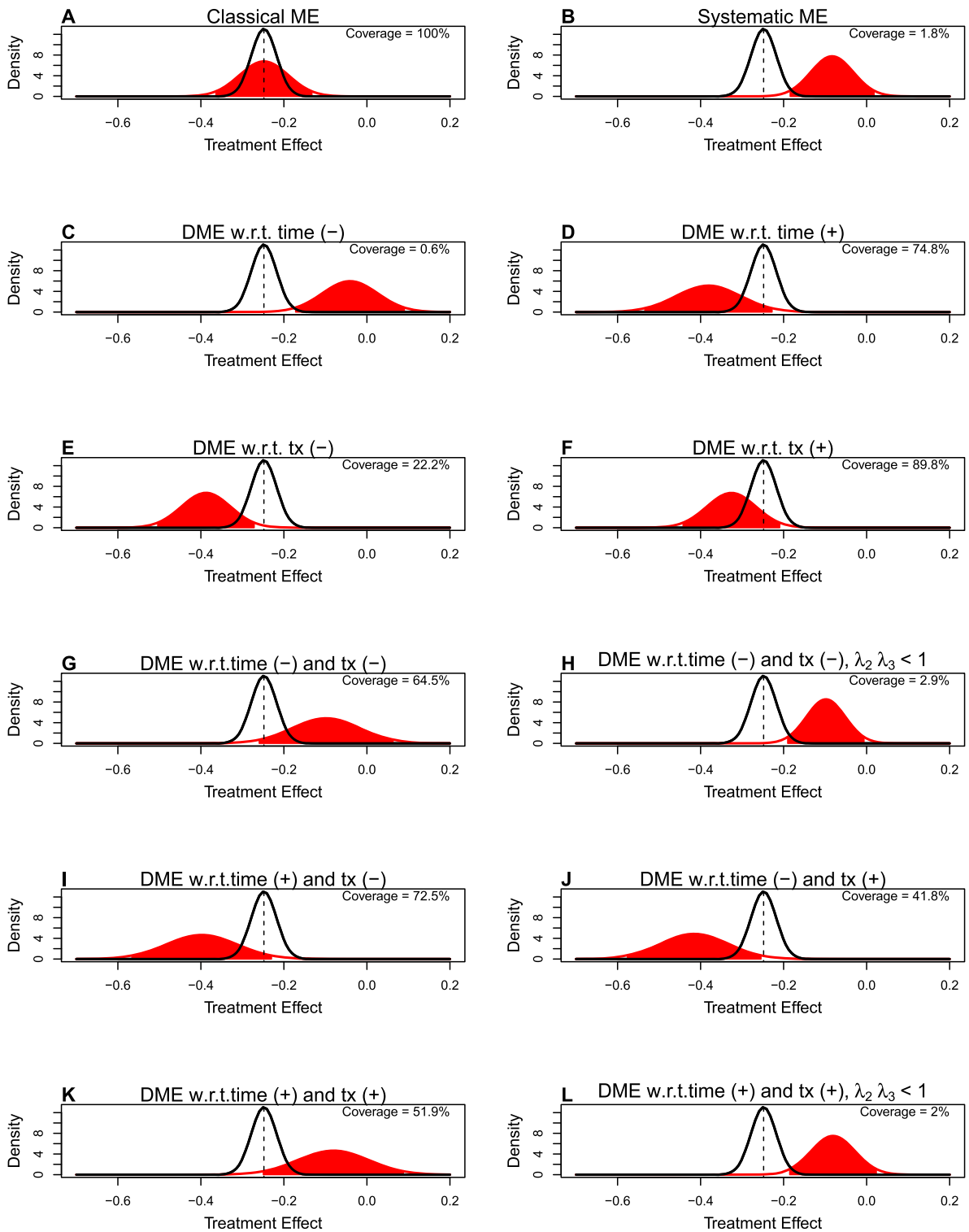
Under some scenarios (systematic ME, DME w.r.t. time (-), DME w.r.t. time (-) and tx (-), DME w.r.t. time (+) and tx (+)), the bias and increased variability can be so great that the 95% confidence interval contains 0, such that the naive treatment effect is no longer significant. Under other scenarios, the bias is in the opposite direction, so

that the naive treatment effect is greater than the true effect.

Figure 4 displays density plots for the distribution of the treatment effect comparing the true treatment effect (in black) and the naive treatment effect under different scenarios of measurement error (in red). This provides a



**Fig. 3** Forest plot of treatment effects, their associated 95% confidence intervals, and coverage of those intervals for the true treatment effect (top row), and naive treatment effects for a range of analyses based on altering the measurement error parameters in Table 2. The (+) and (-) refer to whether the  $\gamma_3$  and  $\gamma_4$  parameters are greater than or less than 0, respectively. The vertical solid line shows the true effect. The vertical dotted line at 0 indicates no treatment effect. In all analyses,  $\lambda_1 = 1.86$  and the sample size is  $n = 372$  per group



**Fig. 4** Density plots for the distribution of the naive treatment effect under different scenarios of measurement error (in red) compared to the true treatment effect (in black). Each panel A-J represents a different measurement error scenario. The dashed lines indicates the mean true treatment effect. The 95% confidence interval of the naive treatment effect is the shaded area under the curve (in red). Measurement error parameters for each scenario are equal to values found in Fig. 3. Coverage of the true treatment effect is 95%. In all analyses, the sample size is  $n = 372$  per group

graphical illustration of coverage of the confidence interval of the treatment effect under the same measurement error-corrected scenarios in Fig. 3. Coverage of the true treatment effect is 95%. Under classical measurement error, the coverage is 100%. Coverage ranges from 0.6% to 89.8% depending on the differential measurement error scenario.

## Discussion

We found that when using self-reported dietary measures as outcomes in a lifestyle intervention study, differential measurement error with respect to treatment condition and time can result in a biased treatment effect and can impact the sample size needed to achieve 80% power in detecting a treatment effect. Increased variability in the outcomes measured with error ( $y_{ij}$ ), increases the sample size needed to achieve 80% power. The impact on sample size differs depending on the type of differential error: increased variability at follow-up ( $\lambda_2$ ) increases the required sample size at a faster rate than increased variability for treatment condition alone at follow-up ( $\lambda_3$ ). This is because increasing  $\lambda_2$  affects all observations, while increasing  $\lambda_3$  affects only those in the treatment group. Similarly, decreasing  $\lambda_2$  and/or  $\lambda_3$ , decreases the sample size required to reach 80% power, with  $\lambda_2$  decreasing the required sample size at a faster rate than  $\lambda_3$ . Naturally, when both factors increased/decreased variability, that had the largest percent increase/decrease on the sample size needed to achieve 80% power. By ignoring the possibility of increased variability at follow-up, trials may be under-powered.

Bias of the treatment effect is also affected by differential measurement error but here, differential measurement error with respect to treatment has a greater impact than does differential measurement error with respect to time. There is little additional bias when there is additional change in slope for the control group ( $\gamma_3$ ). When we set the other parameters in the measurement error model to 0 or 1 ( $\gamma_1, \gamma_4 = 0$  and  $\gamma_2 = 1$ ), then bias is equal to  $(1 + \gamma_3)\beta_2$ . Thus small values of  $\gamma_3$  have little effect on bias. However, even a small increase in slope for the treatment group ( $\gamma_4$ ) can have a substantial impact on bias (see Eq. 5).

In our simulations for power and bias, we fixed  $\gamma_1$ , the additive difference in measurement error between treatment conditions at follow-up. Since  $\gamma_1$  does not appear in the variance calculations (see Eqs. 26 and 27) and only appears in the estimation of the naive treatment effect ( $\Psi^{naive}$ ), it will affect the sample size itself, but not the percent increase. For smaller values of  $\gamma_1$ , the required sample size needed is smaller and for larger values of  $\gamma_1$ , the required sample size is larger. When varying  $\gamma_1$  and keeping the other  $\gamma$  parameters constant at the TOHP values, the percent increase in sample size is still

approximately 32.5%, 65%, and 130% when ( $\lambda_2 = 1, \lambda_3 = 2$ ), ( $\lambda_2 = 2, \lambda_3 = 1$ ), and ( $\lambda_2 = 2, \lambda_3 = 2$ ) respectively. In terms of bias,  $\gamma_1$  is a constant term, shifting the naive treatment effect by the same amount across all values of  $\gamma_2$  and  $\gamma_3$  as it increases or decreases. Thus the shape of the contour plot in Fig. 2 stays the same, only its values change.

Similarly,  $\rho$ , the correlation between baseline and follow-up, was fixed throughout the simulations. The correlation  $\rho$  does not affect bias. As  $\rho$  increases, the sample size required to achieve 80% power decreases and as  $\rho$  decreases, the sample size must increase.

There is a relationship between the values of  $\lambda_2, \lambda_3$  and the percent increase in sample size required to achieve 80% power. Figure 1 reports the percent increase in sample size relative to a referent scenario ( $\lambda_2 = 1, \lambda_3 = 1$ ). Given the calculation of the sample size in Eq. 36, the ratio of sample sizes is equal to the ratio of the variances (Eq. 37). Solving Eq. 37 for a set of  $\lambda$  parameters, one can directly calculate the percent increase in sample size needed under scenarios of increased variability. As Eq. (39) makes clear, values of  $\lambda_2$  have both additive and multiplicative effects on the increase in sample size. This is why there is a doubling in the percent increase in sample size in Fig. 1 as the  $\lambda$  parameters change from ( $\lambda_2 = 1, \lambda_3 = 2$ , 32.5% increase), ( $\lambda_2 = 2, \lambda_3 = 1$ , 65% increase), ( $\lambda_2 = 2, \lambda_3 = 2$ , 130% increase).

The required sample size to achieve 80% power also depends on the  $\gamma$  parameters (Eq. 36) However, the ratio of the percent increase in sample size relative to the referent scenario for two different differential measurement error scenarios is invariant to the values of the  $\gamma$  parameters (Eq. 39).

Depending on the size and the sign of the measurement error parameters, the bias can be in the positive direction, towards zero, or in the negative direction. The combination of positive bias and increased variability can make the estimate of the treatment effect overlap with zero, resulting in a non-significant observed treatment effect. This can be seen in Fig. 3 under Systematic ME, as well as several differential measurement error scenarios where the 95% confidence interval crosses the dotted vertical line at zero. Bias in the negative direction (seen in estimates of the treatment effect to the left of the solid vertical line in Fig. 3) can make the treatment effect appear much larger than the true effect, which could lead investigators to think the intervention was much more successful than it actually was in reality. An extreme case of bias (not shown in Fig. 3) would not only bias in the positive direction, but show a significant positive effect, greater than zero. The treatment effect would be significant but in the opposite direction, and thus yielding a wrong conclusion.

Coverage of the treatment is also affected by differential measurement error. Coverage tends to be higher when



there is less bias in the treatment effect measured with error, but increased variability, such as DME w.r.t. tx (+) in Fig. 3 (Panel F in Fig. 4). Although the true treatment effect is contained within the confidence interval of the naive effect when coverage is high, the naive estimate is highly variable. Coverage is low when there is a large bias, even if variability is increased. When coverage is low, the naive treatment effect differs greatly from the true treatment effect, due to bias. Coverage decreases when  $\lambda_2$  and  $\lambda_3 < 1$ , due to decreased variability. This can be seen comparing panels G and H, and comparing panels K and L in Fig. 4.

In lifestyle intervention studies, it has been shown that measurement error in self-reported dietary measures can differ both with respect to treatment condition and over time. Natarajan et al. [7] investigated measurement error of dietary self-report in an intervention trial in which self-reported and plasma carotenoid biomarker data were available on all participants at each time point. Using a model which took into account measurement error in both self-report and biomarker-based measures, they found that self-reported accuracy improved in participants randomized to the intervention condition. They also found increases in variability among follow-up measurements in the intervention condition.

Espeland et al. [4] fit a longitudinal model to self-reported and urinary sodium to longitudinal data from a lifestyle intervention trial of 900 individuals with hypertension who were randomized to one of four conditions. They found that self-reported sodium intake was less than urinary sodium at all visits and within each study group. Interestingly, the ratio of self-reported to urinary sodium intake was smallest at follow-up compared to baseline in the most intensive intervention condition. The authors hypothesized that this was due to compliance bias and noted that, “subjective pressures to please staff and meet intervention goals led to under-reporting intakes.” Also, unlike the analyses by Natarajan et al. [7], measurement errors were *less* variable during follow-up than at baseline for all cohorts. This was attributed to better recall of foods containing sodium based on knowledge gained from the interventions.

Together, these results suggest that the presence of differential measurement error is likely to be intervention specific and may depend on the population being studied. For example, in a intervention study of youth with type 1 diabetes, Sanjeevi et al. [12] found no evidence of differential measurement error.

While our findings demonstrate the impact of differential measurement error, there are some limitations to this work. We only used two time points in developing our models. As the number of time points increases, so does the number of differential measurement error scenarios.

We assumed continuous, normally-distributed outcomes. We used a linear measurement model as this is a common approach for modeling measurement error and empirically has been shown to provide a good fit to the data when both true intake and its version measured with error are available [6, 7, 13], especially after values have been log-transformed. In practice, investigators are often interested in outcomes such as number of fruits and vegetables [14], which are not normally distributed and the impact of non-differential measurement error could be different. Future work will look at the impact of differential measurement error in non-normal outcomes. We based our simulations on self-reported sodium intake using parameters from the TOHP study. Examining measurement error for other components of diet, such as total intake, would require different parameter values, although one would expect to see results similar to those presented here. Finally, we focused on the setting of measurement error in dietary interventions. However, differential measurement error with respect to treatment and/or time can also exist in observational studies and an area of future work is to better understand the role of measurement error when estimating treatment effects using observational data.

## Conclusions

When designing a longitudinal lifestyle intervention study, researchers using self-reported dietary measures need to consider the impact of measurement error and differential measurement error. Recruiting a larger sample size can help overcome the loss of power associated with the additional variability due to measurement error. However, this approach does nothing to correct for bias. A more expansive approach that would allow the researcher to diagnose and correct for both bias and variance due to measurement error is to include an internal validation study with recovery biomarkers and implement methods that allow for measurement correction using internal validation studies [9, 15]. When an internal validation study is not possible, methods that use external validation studies [16, 17] are possible although they require the user to make additional assumptions regarding transportability of the measurement error model [18]. Still, we feel that these additional efforts to correct for measurement error are worthwhile, as they require only marginally more effort than conducting the intervention itself, and allow researchers to make inferences with greater accuracy and precision.

## Appendix

We present formulas for the bias of the naive treatment effect, its variance, and the coverage of its 95% confidence interval.

**Bias**

Let  $Z_0$  refer to true intake at baseline, and  $Z_1$  refer to true intake at follow-up. Based on Eq. (1) in the main text, the mean value of true intake for each treatment condition and time point is given by

$$E(Z_0 | d_i) = \beta_0 \tag{6}$$

$$E(Z_1 | d_i = 0) = \beta_0 + \beta_1 \tag{7}$$

$$E(Z_1 | d_i = 1) = \beta_0 + \beta_1 + \beta_2. \tag{8}$$

Using Eqs. (6) through (8) the expected change in true intake in the control condition is

$$E[Z_1 - Z_0 | d_i = 0] = \beta_1 \tag{9}$$

and the expected change in true intake in the treatment condition is

$$E[Z_1 - Z_0 | d_i = 1] = \beta_1 + \beta_2 \tag{10}$$

so that the expected true treatment effect is

$$E[Z_1 - Z_0 | d_i = 1] - E[Z_1 - Z_0 | d_i = 0] = \beta_2. \tag{11}$$

Let  $Y_t$  refer to self-reported intake at time  $t$ ,  $t = 0, 1$ . The mean value of self-reported intake for each treatment condition and time point is given by

$$E(Y_t | d) = E_z[E_{Y|Z}(y_t | z_t, d)]$$

so that, using Eq. (3) in the main text, the mean self-reported values by time and treatment group are:

$$E(Y_0 | d_i = 0) = \gamma_0 + \gamma_2\beta_0 \tag{12}$$

$$E(Y_0 | d_i = 1) = \gamma_0 + \gamma_2\beta_0 \tag{13}$$

$$E(Y_1 | d_i = 0) = \gamma_0 + (\gamma_2 + \gamma_3)(\beta_0 + \beta_1) \tag{14}$$

$$E(Y_1 | d_i = 1) = \gamma_0 + \gamma_1 + (\gamma_2 + \gamma_3 + \gamma_4)(\beta_0 + \beta_1 + \beta_2). \tag{15}$$

Using Eqs. (12) through (15), the expected change in self-reported intake in the control condition is

$$E[Y_1 - Y_0 | d_i = 0] = (\gamma_2 + \gamma_3)\beta_1 + \gamma_3\beta_0 \tag{16}$$

and the expected change in self-reported intake in the treatment condition is

$$E[Y_1 - Y_0 | d_i = 1] = \gamma_1 + (\gamma_2 + \gamma_3 + \gamma_4)(\beta_0 + \beta_1 + \beta_2) - \gamma_2\beta_0. \tag{17}$$

The expected treatment effect from the self-reported intake (i.e. the naive treatment effect) is given by

$$\Psi^{naive} = E[Y_1 - Y_0 | d_i = 1] - E[Y_1 - Y_0 | d_i = 0] = \gamma_1 + \beta_2(\gamma_2 + \gamma_3) + \gamma_4(\beta_0 + \beta_1 + \beta_2). \tag{18}$$

The mean bias of the naive treatment effect is obtained by subtracting Eq. (11) from Eq. (18) to obtain.

$$\text{Bias}(\Psi^{naive}) = \gamma_1 + \beta_2(\gamma_2 + \gamma_3) + \gamma_4(\beta_0 + \beta_1 + \beta_2) - \beta_2. \tag{19}$$

**Variance**

The variance of self-reported intake for each treatment condition and time point is given by

$$\text{Var}[Y_t | d] = E[\text{Var}(Y_t | Z_t, d)] + \text{Var}[E(Y_t | Z_t, d)]$$

so that,

$$\text{Var}[Y_0 | d_i = 0] = \sigma_z^2(\lambda_1 + \gamma_2^2) \tag{20}$$

$$\text{Var}[Y_0 | d_i = 1] = \sigma_z^2(\lambda_1 + \gamma_2^2) \tag{21}$$

$$\text{Var}[Y_1 | d_i = 0] = \sigma_z^2(\lambda_1\lambda_2 + (\gamma_2 + \gamma_3)^2) \tag{22}$$

$$\text{Var}[Y_1 | d_i = 1] = \sigma_z^2(\lambda_1\lambda_2\lambda_3 + (\gamma_2 + \gamma_3 + \gamma_4)^2). \tag{23}$$

The covariance between baseline and follow-up self-reported intake is

$$\text{Cov}(Y_1, Y_0) = E[\text{Cov}(Y_1, Y_0 | Z_1, Z_0)] + \text{Cov}[E(Y_1 | Z_1, Z_0), E(Y_0 | Z_1, Z_0)]$$

so that, for the control condition ( $d_i = 0$ ), the covariance is

$$\begin{aligned} \text{Cov}(Y_1, Y_0 | d_i = 0) &= \rho\sigma_z^2 + \text{Cov}[(\gamma_0 + \gamma_2z_{ij}), \\ &\quad (\gamma_0 + (\gamma_2 + \gamma_3)z_{ij})] \\ &= \rho\sigma_z^2 + \gamma_2(\gamma_2 + \gamma_3)\text{Cov}(z_{ij}, z_{ij}) \\ &= \rho\sigma_z^2(1 + \gamma_2(\gamma_2 + \gamma_3)), \end{aligned} \tag{24}$$

and for the intervention condition ( $d_i = 1$ ), the covariance is

$$\begin{aligned} \text{Cov}(Y_1, Y_0 | d_i = 1) &= \rho\sigma_z^2 + \text{Cov}[(\gamma_0 + \gamma_2z_{ij}), \\ &\quad (\gamma_0 + \gamma_1 + (\gamma_2 + \gamma_3\gamma_4)z_{ij})] \\ &= \rho\sigma_z^2 + \gamma_2(\gamma_2 + \gamma_3 + \gamma_4)\text{Cov}(z_{ij}, z_{ij}) \\ &= \rho\sigma_z^2(1 + \gamma_2(\gamma_2 + \gamma_3 + \gamma_4)). \end{aligned} \tag{25}$$

The variance in self-reported intake in the control condition is given by

$$\begin{aligned} \text{Var}(Y_1 - Y_0 | d_i = 0) &= \text{Var}(Y_1 | d_i = 0) \\ &\quad + \text{Var}(Y_0 | d_i = 0) - 2\text{Cov}(Y_0, Y_1) \\ &= \sigma_z^2(\lambda_1\lambda_2 + (\gamma_2 + \gamma_3)^2 + \lambda_1 + \gamma_2^2 \\ &\quad - 2\rho(1 + \gamma_2(\gamma_2 + \gamma_3))). \end{aligned} \tag{26}$$

The variance in self-reported intake in the treatment condition is given by

$$\begin{aligned} \text{Var}(Y_1 - Y_0 | d_i = 1) &= \sigma_z^2(\lambda_1\lambda_2\lambda_3 + (\gamma_2 + \gamma_3 + \gamma_4)^2 \\ &\quad + \lambda_1 + \gamma_2^2 \\ &\quad - 2\rho(1 + \gamma_2(\gamma_2 + \gamma_3 + \gamma_4))). \end{aligned} \tag{27}$$

**Coverage**

The coverage probability of a confidence interval is the proportion of the time that the interval contains the true quantity of interest. From Eq. (11), the true treatment effect is equal to  $\beta_2$ .

For either treatment condition ( $d_i = 0, 1$ ), let  $\bar{z}_1$  be the sample mean of  $z$  at time 1 and let  $\bar{z}_0$  be the sample mean of  $z$  at time 0. Let  $\rho$  be the correlation between  $z_0$  and  $z_1$  and let  $n_d$  be the sample size in either treatment condition. Regardless of treatment condition, the variance of  $\bar{z}_1 - \bar{z}_0$  is defined as

$$\begin{aligned} \text{Var}(\bar{z}_1 - \bar{z}_0) &= \frac{\sigma_z^2}{n_d} + \frac{\sigma_z^2}{n_d} - 2 \frac{\rho\sigma_z^2}{n_d} \\ &= \frac{2\sigma_z^2(1 - \rho)}{n_d}. \end{aligned} \tag{28}$$

Let  $\delta_1 = \bar{z}_1 - \bar{z}_0$  for  $d_i = 1$  and let  $\delta_0 = \bar{z}_1 - \bar{z}_0$  for  $d_i = 0$ . The estimate of the treatment effect  $\beta_2$  is  $\hat{\beta}_2 = \delta_1 - \delta_0$ . The variance of the estimated treatment effect is

$$\text{Var}(\hat{\beta}_2) = \frac{4\sigma_z^2(1 - \rho)}{n_d} \tag{29}$$

So that assuming Normality as in (1) in the main text, the sampling distribution of  $\hat{\beta}_2$  is

$$\hat{\beta}_2 \sim N\left(\beta_2, \frac{4\sigma_z^2(1 - \rho)}{n_d}\right). \tag{30}$$

Let  $\bar{y}_1$  be the sample mean of  $y$  at time 1 and let  $\bar{y}_0$  be the sample mean of  $y$  at time 0. Let  $\delta_1^* = \bar{y}_1 - \bar{y}_0$  for  $d_i = 1$  and let  $\delta_0^* = \bar{y}_1 - \bar{y}_0$  for  $d_i = 0$ . An estimate of the naive treatment effect is  $\hat{\Psi}^{naive} = \delta_1^* - \delta_0^*$ . Its variance is

$$\text{Var}(\hat{\Psi}^{naive}) = \frac{\text{Var}(Y_1 - Y_0 | d_i = 1)}{n_d} + \frac{\text{Var}(Y_1 - Y_0 | d_i = 0)}{n_d} \tag{31}$$

where the variance terms in the numerator were defined in Eqs. (26) and (27).

A 95% confidence interval for the naive treatment effect is:

$$\hat{\Psi}^{naive} \pm 1.96 * \text{SE}(\hat{\Psi}^{naive}) \tag{32}$$

The coverage of this confidence interval is the probability that it contains the true quantity of interest

$$\text{Coverage} = \text{Pr}(\Psi_{lower} < \hat{\beta}_2 < \Psi_{upper}) \tag{33}$$

where  $\Psi_{lower}$  and  $\Psi_{upper}$  are the endpoints of the confidence interval in (32).

**Power and sample size**

We power based on a two-sample z-test for the difference in mean change scores between treatment and control groups. The difference in means under the alternative hypothesis is the naive treatment effect  $\Psi^{naive}$ , given by

Eq. (18). The non-centrality parameter is

$$\Delta = \frac{\Psi^{naive}}{\sqrt{\text{Var}(\hat{\Psi}^{naive})}} \tag{34}$$

where the denominator is the square root of the variance in Eq. (31). Under a two-sample z-test, the critical value under the two-sided null hypothesis at a Type 1 error rate of .05 is -1.96, so that power is calculated by

$$\text{Power} = \Phi(-1.96 - \Delta), \tag{35}$$

where  $\Phi$  represents the standard normal distribution function.

Solving Eq. (35) for  $n_d$ , we can obtain an equation for the sample size of each treatment group. Assuming power of 80% and a Type 1 error rate of .05, the sample size of each group is given by

$$\begin{aligned} n_d &= (\Phi^{-1}(.80) + 1.96)^2 \\ &\times \frac{\text{Var}(Y_1 - Y_0 | d_i = 1) + \text{Var}(Y_1 - Y_0 | d_i = 0)}{(\Psi^{naive})^2} \end{aligned} \tag{36}$$

The total sample size required is thus  $2n_d$ .

Figure 1 reports the percent increase in sample size relative to a referent scenario. Let  $n$  be the sample size under the referent scenario and let  $n^*$  be the sample size under an alternative scenario with values  $\lambda_2^*, \lambda_3^*$ . The proportion increase in sample size reported in Fig. 1 is  $\frac{n^*}{n} - 1$  or  $\frac{n^* - n}{n}$ . Using Eq. (36), the ratio of sample sizes is equal to the ratio of variances.

$$\begin{aligned} \frac{n^*}{n} &= \frac{\text{Var}^*(Y_1 - Y_0 | d_i = 1) + \text{Var}^*(Y_1 - Y_0 | d_i = 0)}{\text{Var}(Y_1 - Y_0 | d_i = 1) + \text{Var}(Y_1 - Y_0 | d_i = 0)} \\ &= \frac{(\lambda_2^* + 2 + \lambda_2^*\lambda_3^*) + c}{(\lambda_2 + 2 + \lambda_2\lambda_3) + c} \end{aligned} \tag{37}$$

where  $c$  is a constant term that does not depend on any values of  $\lambda$ . The ratio of percent increases under two different scenarios, for example, non-differential measurement error with respect to treatment and non-differential measurement error with respect to time is:

$$\left(\frac{n^* - n}{n^{**} - n}\right) = \frac{\lambda_2^*(1 + \lambda_3^*) - \lambda_2(1 + \lambda_3)}{\lambda_2^{**}(1 + \lambda_3^{**}) - \lambda_2(1 + \lambda_3)}. \tag{38}$$

Figure 1 used non-differential measurement error with respect to treatment and time as the reference scenario ( $\lambda_2 = 1, \lambda_3 = 1$ ). Under this reference scenario, Eq. (38) reduces to

$$\left(\frac{n^* - n}{n^{**} - n}\right) = \frac{\lambda_2^*(1 + \lambda_3^*) - 2}{\lambda_2^{**}(1 + \lambda_3^{**}) - 2}. \tag{39}$$

**Abbreviations**

ME: measurement error; DME: differential measurement error; w.r.t.: with respect to; tx: treatment; CI: confidence interval; TOHP: Trials of Hypertension Prevention Study

## Acknowledgements

None

## Authors' contributions

J.S. formulated the research question. J.S. and D.A. designed the study and wrote the article. D.A. carried out the simulations. Both authors read and approved the final manuscript.

## Funding

This research was supported by the National Institutes of Health (R01 HL127491). The funding sources played no role in study design, collection, analysis, or interpretation of data, the writing of the manuscript or the decision to submit for publication.

## Availability of data and materials

The simulation parameter values were calibrated using data on sodium intake from the Trials of Hypertension Prevention Study (TOHP). The data is available upon request at <https://biolincc.nhlbi.nih.gov/studies/tohp/>.

## Declarations

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable

### Competing interests

The authors declare that they have no competing interests.

Received: 9 October 2020 Accepted: 2 August 2021

Published online: 16 September 2021

## References

1. Willet W. *Nutritional Epidemiology*, Third Edition. New York: Oxford University Press; 2013.
2. Forster JL, Jeffery RW, VanNatta M, Pirie P. Hypertension Prevention Trial: Do 24-h food records capture usual eating behavior in a dietary change study? *Am J Clin Nutr*. 1990;51:253–7.
3. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement Error in Nonlinear Models: A Modern Perspective*, Second Edition. New York: Chapman & Hall/CRC; 2006.
4. Espeland MA, Kumanyika S, Wilson AC, Wilcox S, Chao D, Bahnson J, Reboussin DM, Easter L, Zheng B. Lifestyle interventions influence relative errors in self-reported diet intake of sodium and potassium. *Ann Epidemiol*. 2001;11:85–93.
5. Buzzard IM, Faucett CL, Jeffery RW, McBane L, McGovern P, Baxter JS, Shapiro AC, Blackburn GL, Chlebowski RT, Elashoff RM, Wynder EL. Monitoring dietary change in a low-fat diet intervention study: Advantages of using 24-hour dietary recalls vs food records. *J Am Diet Assoc*. 1996;96:574–9.
6. Espeland MA, Kumanyika S, Wilson AC, Wilcox S, Chao D, Bahnson J, Reboussin DM, Easter L, Zheng B, Group TCR, et al. Lifestyle interventions influence relative errors in self-reported diet intake of sodium and potassium. *Ann Epidemiol*. 2001;11:85–93.
7. Natarajan L, Pu M, Fan J, Levine RA, Patterson RE, Thomson CA, Rock CL, Pierce JP. Measurement error of dietary self-report in intervention trials. *Am J Epidemiol*. 2010;172:819–27.
8. Kristal AR, Andrilla CHA, Koepsell TD, Diehr PH, Cheadle A. Dietary assessment instruments are susceptible to intervention-associated response set bias. *J Am Diet Assoc*. 1998;98(1):40–3.
9. Keogh RH, Carroll RJ, Toozé JA, Kirkpatrick SI, Freedman LS. Statistical issues related to dietary intake as the response variable in intervention trials. *Stat Med*. 2016;35:4493–508.
10. Hedeker D, Gibbons RD. *Longitudinal Data Analysis*. Hoboken, NJ: Wiley; 2006.
11. Satterfield S, Cutler JA, Langford HG, Applegate WB, Borhani NO, Brittain E, Cohen JD, Kuller LH, Lasser NL, Oberman A, et al. Trials of Hypertension Prevention phase I design. *Ann Epidemiol*. 1991;1(5):455–71.
12. Sanjeevi N, Lipsky L, Liu A, Nansel T. Differential reporting of fruit and vegetable intake among youth in a randomized controlled trial of a behavioral nutrition intervention. *Int J Behav Nutr Phys Act*. 2019;16(1):15.
13. Freedman LS, Commins JM, Moler JE, Willett W, Tinker LF, Subar AF, Spiegelman D, Rhodes D, Potischman N, Neuhauser ML, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for potassium and sodium intake. *Am J Epidemiol*. 2015;181(7):473–87.
14. Spring B, Schneider K, McFadden H, Vaughn J, Kozak A, Smith M, Moller A, Epstein L, Russell S, DeMott A, Hedeker D. Make Better Choices (MBC): Study design of a randomized controlled trial testing optimal technology-supported change in multiple diet and physical activity risk behaviors. *BMC Public Health*. 2010;10:586.
15. Talitman M, Gorfine M, Steinberg DM. Estimating the intervention effect in calibration substudies. *Stat Med*. 2020;39(3):239–51.
16. Siddique J, Daniels MJ, Carroll RJ, Raghunathan TE, Stuart EA, Freedman LS. Measurement error correction and sensitivity analysis in longitudinal dietary intervention studies using an external validation study. *Biometrics*. 2019;75(3):927–37.
17. Nab L, Groenwold RHH, Welsing PMJ, van Smeden M. Measurement error in continuous endpoints in randomised trials: Problems and solutions. *Stat Med*. 2019;38(27):5182–96.
18. Ackerman B, Siddique J, Stuart EA. Calibrating validation samples when accounting for measurement error in intervention studies. *Stat Methods Med Res*. 2021;30(5):1235–48.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

