Measuring bias in self-reported data

Robert Rosenman[*],

School of Economic Sciences, Washington State University, P.O. Box 646210, Pullman, WA 99164-6210, USA

Vidhura Tennekoon, and

School of Economic Sciences, Washington State University, P.O. Box 646210, Pullman, WA 99164-6210, USA vidhura@wsu.edu

Laura G. Hill

Department of Human Development, Washington State University, 523 Johnson Tower, Pullman WA 99164, USA laurahill@wsu.edu

Abstract

Response bias shows up in many fields of behavioural and healthcare research where self-reported data are used. We demonstrate how to use stochastic frontier estimation (SFE) to identify response bias and its covariates. In our application to a family intervention, we examine the effects of participant demographics on response bias before and after participation; gender and race/ethnicity are related to magnitude of bias and to changes in bias across time, and bias is lower at post-test than at pre-test. We discuss how SFE may be used to address the problem of 'response shift bias' – that is, a shift in metric from before to after an intervention which is caused by the intervention itself and may lead to underestimates of programme effects.

Keywords

response bias; response-shift bias; programme evaluation; stochastic frontier analysis; stochastic frontier estimation; SFE; prevention science

1 Introduction

In this paper, we demonstrate the potential of a common econometric tool, stochastic frontier estimation (SFE), to measure response bias and its covariates in self-reported data. We illustrate the approach using self-reported measures of parenting behaviours before and after a family intervention. We demonstrate that in addition to affecting targeted behaviours, an intervention may also affect any bias associated with self-assessment of those behaviours. We show that SFE can be used to identify and correct for bias in self-assessment both before and after treatment, resulting in more accurate estimates of treatment effects.

Response bias is a widely discussed phenomenon in behavioural and healthcare research where self-reported data are used; it occurs when individuals offer self-assessed measures of some phenomenon. There are many reasons individuals might offer biased estimates of self-assessed behaviour, ranging from a misunderstanding of what a proper measurement is to social-desirability bias, where the respondent wants to 'look good' in the survey, even if the survey is anonymous. Response bias itself can be problematic in programme evaluation and research, but is especially troublesome when it causes a recalibration of bias after an intervention. Recalibration of standards can cause a particular type of measurement bias known as 'response-shift bias' (Howard, 1980). Response-shift bias occurs when a respondent's frame of reference changes across measurement points, especially if the changed frame of reference is a function of treatment or intervention, thus, confounding the treatment effect with bias recalibration. More specifically, an intervention may change respondents' understanding or awareness of the target concept and the estimation of their level of functioning with respect to the concept (Sprangers and Hoogstraten, 1989), thus changing the bias at each measurement point. In fact, some treatments or interventions are intended to change how respondents look at the target concept. Further complicating matters is that an intervention may affect not only a respondent's metric for targeted behaviours across time points (resulting in response shift bias) but may also affect other types of response bias. For example, social desirability bias may decrease over the course of an intervention as respondents come to know and trust a service provider. Thus, it is necessary to understand the degree and type of response bias at both pretest and posttest in order to determine whether response shift has occurred.

When there is a potential for confusing bias recalibration with treatment outcomes, statistical approaches may be useful (Schwartz and Sprangers, 1999). In recent years, researchers have applied structural equation modelling (SEM) to the problem of decomposing error in order to identify response shift bias (Oort, 2005; Oort et al., 2005). In this paper, we suggest a different statistical approach which reveals response bias at a single time point as well as differences in bias across time points. Perhaps more importantly, it identifies covariates of these differences. When applied before and after an intervention, it reveals differences related to changes in respondents' frame of reference. Thus, it can be used to decompose errors so that recalibration of the bias occurring across time points can be distinguished from simple response bias within each time point. The suggested approach is based on SFE (Aigner et al., 1977; Battese and Coelli, 1995; Meeusen and van den Broeck, 1977), a technique widely used in economics and operational research.

Our approach has two significant advantages over that proposed by Oort et al. (2005). Their approach reveals only *aggregate changes* in the responses and requires a minimum of two temporal sets of observations on the self-rating of interest as well as multiple measures of the item to be rated. SFE, to its credit, can identify *response differences across individuals* (as opposed to simply aggregate response shifts) with a single temporal observation and a single measure, so is much less data intensive. Moreover, since it identifies differences at the individual level, it allows the analyst to identify not only that responses differ by individual, but what characteristics are at the root of the differences. Thus, as long as more than one temporal observation is available for respondents, SFE can be used to systematically identify

different types of response recalibration by looking at the changes at the individual level, and aggregating them. SFE again has an advantage because the causes of both bias and recalibration can be identified at the individual level.

What may superficially be seen as two disadvantages to SFE when compared to SEM approaches are actually common to both methods. First, both measure response (and therefore response shift) against a common subjective metric established by the norm of the data. In fact, any systematic difference by an individual from this norm is how we measure 'response bias'. With both SEM and SFE, if an objective metric exists, the difference between the self-rating and the objective measure is easily established. A second apparent disadvantage is that SFE requires a specific assumption of a truncated distribution of the bias (although it is possible to test this assumption statistically). While SEM can reveal response shift on individual bias without such a strong assumption, aggregate changes become manifest only if "many respondents experience the same shift in the same direction" [Oort, (2005), p.595]. Hence, operationally the assumptions are nearly equivalent.

In next section, we explain how we model response bias and response recalibration within the SFE framework. In Section 3, we present our empirical application including the results of our baseline model and a model with heteroscedastic errors as a robustness check. In Section 4, we discuss the relative merits of the method we propose, together with its limitations and offer some conclusions.

2 Response bias and SFE

We are concerned with situations where individuals do not have an objective measure of some variable of interest which we denote Y^*_{it}, and we have to use a subjective measure (denoted Y_{it}) as a proxy instead. An unbiased estimate of the variable of interest Y^*_{it} can be defined as,

$$Y_{it}|Y^*_{it}, Z^*_{it} = Y_{it}|Y^*_{it} \quad (1)$$

where Y_{it} denotes the observed measurement, Y^*_{it} is the true attribute being measured and Z_{it} represents variables other than Y^*_{it}. When Y_{it} is self-reported Z_{it} includes (often unobserved) variables affecting the frame of reference used by respondents for measuring Y^*_{it} and (1) is not assured. Within this context, *response bias* is simply the case that $Y_{it} | Y^*_{it}, Z_{it} \quad Y_{it} | Y^*_{it}$. The bias is upward if $Y_{it} | Y^*_{it}, Z_{it} > Y_{it} | Y^*_{it}$ and downward if the inequality goes the other way.

Our approach for measuring response bias and bias recalibration (change in response bias between two time periods) is based on the Battese and Coelli (1995) adaptation of the stochastic frontier model (SFE) independently proposed by Aigner et al. (1977), and Meeusen and van den Broeck (1977). Let

$$Y^*_{it} = T\beta_0 + X_{it}\beta_t + \varepsilon_{it} \quad (2)$$

where Y^*_{it} is the true (latent) outcome, T denotes some treatment or intervention,[1] X_{it} are variables other than the treatment that explain the outcome and ε_i is a random error term. For

identification, we assume that ε_{it} is distributed iid $N\left(0, \sigma_\varepsilon^2\right)$. The observed self-reported outcome is a combination of true outcome and the response bias Y_{it}^R.

$$Y_{it}=Y_{it}^*+Y_{it}^R \quad (3)$$

We consider the specific case that the bias term Y_{it}^R has a truncated-normal distribution

$$Y_{it}^R=u_{it}\cdot(u_{it}>0) \quad (4)$$

where u_{it} is a random variable which accounts for response shift away from a subjective norm response level (usually called the 'frontier' in SFE) and is distributed $N\left(\mu_{it}, \sigma_u^2\right)$ independent of ε_{it}. Moreover,

$$\mu_{it}=T\delta_0+z_{it}\delta_t \quad (5)$$

where the vector z_{it} includes variables (other than the treatment) that explain the specific deviation from the response frontier. Subscript i indexes the individual observation and, subscript t denotes time.[2] Substituting (2), (4) and (5) in (3) we can write,

$$E\left(Y_{it}\right)=T\beta_0+X_{it}\beta_t+T\delta_0+z_{it}\delta_t+\sigma_u\cdot\frac{\phi\left(\frac{T\delta_0+z_{it}\delta}{\sigma_u}\right)}{\Phi\left(\frac{T\delta_0+z_{it}\delta_t}{\sigma_u}\right)} \quad (6)$$

where $\phi(.)$ and $\Phi(.)$ are the standard normal probability density function and cumulative probability functions, respectively. Any *treatment effect* is given by β_0 in equation (6). The normal relationship between the *Xs* and *Y* are given by β_t. The last three terms on the right hand side represent the observation-specific response bias from this normal relationship. Treatment can affect both the maximum possible value of the measured outcome of a given individual (as defined by $X_{it}\beta_t$), and the response bias. If treatment changes the response bias it will be indicated by the term δ_0 and the bias recalibration is given by

$$E\left(Y_{i2}-Y_{i1}\right)-\beta_0=\delta_0+\sigma_u\cdot\frac{\phi\left(\frac{T\delta_0+z_{it}\delta_t}{\sigma_u}\right)}{\Phi\left(\frac{T\delta_0+z_{it}\delta_t}{\sigma_u}\right)}-\sigma_u\cdot\frac{\phi\left(\frac{z_{it}\delta_t}{\sigma_u}\right)}{\Phi\left(\frac{z_{it}\delta_t}{\sigma_u}\right)}. \quad (7)$$

The estimated δ_0 coefficient on treatment indicates how treatment has changed response bias. If $\delta_0 = 0$ there is no recalibration and the response bias, if it exists, is not affected by the treatment. Cross terms of treatment and other variables (that is, slope dummy variables) may be used if the treatment is thought to change the general way these other variables interact with functioning.

Recalibration can occur independently of the treatment effect. In fact, recalibration is sometimes a goal of the treatment or intervention in addition to the targeted outcome, which

---

[1]We present a single model that allows for pre- and post-intervention measurement of the outcome of interest and bias. If the self-reported data is not related to an intervention, β_0 and δ_0 (below) are identically 0 and there is only one time period, t.
[2]Due to symmetry of the normal distribution, without loss of generality we can also assume that the bias distribution is right truncated.

means a desired outcome is that $\delta \quad 0$ and $Y_{i1} \mid Y^*_{it} \quad Y_{i2} \mid Y^*_{it}$ *for* $t \in \{1,2\}$. In other words, there is a change in individual measurement scale caused (and intended) by the intervention.

3 An application to evaluation of a family intervention

We applied SFE to examine response bias and recalibration in programme evaluations of a popular, evidence-based family intervention (the Strengthening Families Program for Parents and Youth 10–14, or SFP) (Kumpfer et al., 1996). Families attend SFP once a week for seven weeks and engage in activities designed to improve family communication, decrease harsh parenting practices, and increase parents' family management skills. At the beginning and end of a programme, parents report their level of agreement with various statements related to skills and behaviours targeted by the intervention (e.g., 'I have clear and specific rules about my child's association with peers who use alcohol'). Consistent with the literature on response shift, we hypothesised that non-random bias would be greater at pretest than at posttest as parents changed their standards about intervention-targeted behaviours and became more conservative in their self-ratings. In other words, we expected that after the intervention parents would recalibrate their self-ratings downward, resulting in an underestimate of the programme's effects.

3.1 Sample

Our data consisted of 1437 parents who attended 94 SFP cycles in Washington State and Oregon from 2005 through 2009. 25% of the participants identified themselves as male, 72% as female, and 3% did not report gender. 27% of the participants identified themselves as Hispanic/Latino, 60% as White, 2% as Black, 4% as American Indian/Alaska Native, 3% as other or multiple race/ethnicity, and 3% did not report race/ethnicity. Almost 74% of the households included a partner or spouse of the attending parent, and 19% reported not having a spouse or partner. For almost 8% of the sample, the presence of a partner or spouse is unknown. Over 62% of our observations are from Washington State, with the remainder from Oregon.

3.2 Measures

The outcome measure consisted of 13 items assessing parenting behaviours targeted by the intervention, including communication about substance use, general communication, involvement of children in family activities and decisions, and family conflict. Items were designed by researchers of the programme's efficacy trial and information about the scale has been reported on elsewhere (Spoth et al., 1995; Spoth et al., 1998). Cronbach's alpha (a measure of internal consistency) in the current data was .85 at both pretest and posttest. Items were scored on a 5-point Likert-type scale ranging from 1 ('strongly disagree') to 5 ('strongly agree').

Variables used in the analysis, including definitions and summary statistics, are presented in Table 1. The average family functioning, as measured by the change in self-assessed parenting behaviours from the pretest to the posttest, increased from 3.98 to 4.27 after participation in SFP.

3.3 Procedure

Pencil-and-paper pretests were administered as part of a standard, ongoing programme evaluation on the first night of the programme, before programme content was delivered; posttests were administered on the last night of the programme. All data are anonymous; names of programme participants are not linked to programme evaluations and are unknown to researchers. The Institutional Review Board of Washington State University issued a Certificate of Exemption for the procedures of the current study.

We used SFE to estimate (pre- and post-treatment) family functioning scores as a function primarily of demographic characteristics. Based on previous literature (Howard and Dailey, 1979), we hypothesised that the one-sided errors (response bias) would be downward, and preliminary analysis supported that hypothesis.[3] Additional preliminary analysis of which variables to include among z_i (including a model using all the explanatory variables) led us to conclude that three variables determined the level of bias in the family functioning assessment – age, Latino/Hispanic ethnicity, and whether or not the functioning measure was a pretest or posttest assessment. We used the 'xtfrontier' routine in Stata to estimate the parameters of our models. Unlike the applications of SFE to technical efficiency estimation our model does not require log transforming the dependent variable.

3.4 The baseline model

The results of the baseline SFE model are shown in Table 2. The Wald χ^2 statistic indicated that the regression was highly significant. Several demographic variables were found to influence the assessment of family functioning with conventional statistical significance. Males gave lower estimates of family functioning than did females and those with unreported gender. All non-White ethnic groups (and those with unreported race/ethnicity) assessed their family's functioning more highly than did White respondents. Participation in the Strengthening Families Program increased individuals' assessments of their family's functioning.

We assessed bias, and its change, from the coefficient estimates for the δ parameters where $\mu_i = z_i\delta$. Our first overall question was if, in fact, there was a one-sided error. Three measures of unexplained variation are shown in Table 2: $\sigma^2 = E(\varepsilon_i - u_i)^2$ is the variance of the total error, which can be broken down into component parts, $\sigma_u^2 = E\left(u_i^2\right)$ and $\sigma_\varepsilon^2 = E\left(\varepsilon_i^2\right)$. The statistic $\gamma = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}$ gives the percent of total unexplained variation attributable to the one-sided error. To ensure $0 \leq \gamma \leq 1$ the model was parameterised as the inverse logit of γ and reported as inlgtgamma. Similarly, the model estimated the natural log of σ^2, reported as lnsigma2, and used these estimates to derive σ^2 , σ_ε^2, σ_u^2 and γ. As seen in the table the estimates for inlgtgamma was highly significant but the estimate for lnsigma2 had a p-value of 0.317, which means we cannot reject a hypothesis that all of the variation in the responses

---

[3]When we tried to estimate the parameters of a model with one-sided errors upward the maximisation procedure failed to converge. A specification with one-sided errors upward but without a constant term converged, but a null hypothesis that there is a one-side error term was rejected with near certainty, indicating that there is no sizable upward response bias. A similar analysis but with the one-sided upward errors completely random (rather than dependent on treatment and other variables) was also rejected, again with near certainty. Thus, upward bias was robustly rejected.

is due to respondent-specific bias. Hence, we found strong support for the one-sided variation that we call bias, and we saw that by far the most substantial portion of the unexplained variation in our data came from that source.

Three variables explained the level of bias. Latino/Hispanic respondents on average had more biased estimates of their family functioning. Looking again at equation (3), we see that this means they, relative to other ethnic groups, underestimated their family functioning. However, we found that older participants had smaller biases, thus giving closer estimates of their family's relative functioning. Of primary interest is the estimate of the treatment effect. Participation in SFP strongly lowered the bias, on average.

3.5 Decomposing the measured change in functioning

The total change in the functioning score averaged 0.295. This total change consisted of two parts as indicated by the following:

Total change = Measured prescore − Measured postscore = (Real prevalue − Prevalue bias) − (Real postvalue − Postvalue bias) = Real change − (Postvalue bias − Prevalue bias) The term in parentheses is negative (the estimation indicates that treatment lowered the bias). Thus, the total change in the family functioning score underestimated the improvement due to SFP, although the measured post-treatment family functioning was not as large as it would seem from the reported family functioning scores, on average. Table 3 shows the average estimated bias by pre- and post-treatment, and the average change in bias, which was –0.133. Thus, the average improvement in family functioning was underestimated by this amount.

Table 4 shows the results of a regression on bias change and demographic and other characteristics. Males and Black respondents had marginally larger bias changes, while those with race/ethnicity unreported had smaller bias changes. Since the bias change was measured as postscore bias minus prescore bias, this means that the bias changed less, on average, for male and Black respondents, but more, on average, for those whose race was unreported.

3.6 The SFE model with heteroscedastic error

One alternative to our baseline model (known as the total effects model in SFE terminology) which generated the results in Table 2 is a SFE model which allows for heteroscedasticity in ε_i, u_i, or both. More precisely, for this model, we maintained equation (3) but had $E(\varepsilon^2) = \omega_\varepsilon w_i$ and $E(u) = \omega_u w_i$ where ω_ε and ω_u are parameters to be estimated and w_i are variables that explain the heteroscedasticity. We note that w_i need not be the same in the two expressions, but since elements of ω_ε and ω_u can be zero we lose no generality by showing it as we do, and in fact in our application we used the same variables in both expressions, those that we used to explain μ in the first model. Table 5 reports the results of such a model. In this case, the one-sided error we ascribe to bias is evident from statistically significant parameters in the explanatory expressions for σ_u^2.

We note first that the estimates in the main body of the equation were quantitatively and qualitatively very similar to those for the non-heteroscedastic SFE model. The only

substantive change is that age was no longer significant at an acceptable p-value, and race unreported had a p-value of 0.1. All signs and magnitudes were similar. Once again, results indicated that participation in SFP (treatment) strongly improved functioning. Additionally, treatment lowered the variability of both sources of unexplained variation across participants. Th e decreased unexplained variation due to ε is likely explained by individuals having a better idea of the constructs assessed by scale items. For our purposes, the key statistic here is the coefficient of treatment explaining σ_u^2. The estimated parameter was negative and significant with a p-value = 0.03. Since the bias was one-sided we clearly can conclude that going through SFP lowered the variability of the bias significantly. Moreover, these estimates can be used to predict the bias of each observation, and with this model the average bias fell from 0.545 to 0.492, so while the biases were larger with this model, the decrease in the average (–0.63) was about one-half the decrease we saw in the first model.

4 Discussion and conclusions

As we noted earlier, bias in self-rating is of concern in a variety of research areas. In particular, the potential for recalibration of self-rating bias as a function of material or skills learned in an intervention has long been a concern to programme evaluators as it may result in underestimates of programme effectiveness (Howard and Dailey, 1979; Norman, 2003; Pratt et al., 2000; Sprangers, 1989). However, in the absence of an objective performance measurement, it has not been possible to determine whether lower posttest scores truly represent response-shift bias or instead an actual decrement in targeted behaviours or knowledge (i.e., an iatrogenic effect of treatment). By allowing evaluators to test for a decrease in response bias from pretest to posttest, SFE provides a means of resolving this conundrum.

The SFE method, however, is not without problems. The main limitation is that the estimates rely on assumptions about the distributions of the two error components. Model identification requires that one of the error terms, the bias term in our application, to be one-sided. This, however, is not as strong an assumption as it looks, for two reasons. First, often there is prior information or theory that indicates the most likely direction for the bias. Second, the validity of the assumption can be tested statistically.

We presented SFE as a method to identify response bias and changes in response bias, within the context of self-reported measurements at individual and aggregate levels. Even though we proposed a novel application, the techniques not new, and has been widely used in economics and operational research for over three decades. The procedure is easily adoptable by researchers, since it is already supported by several statistical packages including Stata (StataCorp., 2009) and Limdep (Econometrica Software, Inc., 2009).

Response bias has long been a key issue in psychometrics, with response shift bias a particular concern in programme evaluation. However, almost all statistical attempts to address the issue have been confined to using SEM to test for response shift bias at the aggregate level. As noted in the introduction, our approach has three significant advantages over SEM techniques that try to measure response bias. SEM requires more data – multiple time periods and multiple measures, and measures bias only in the aggregate. SFE can

identify bias with a single time period (although multiple observations are needed to identify bias recalibration) and identifies response biases across individuals. Perhaps the biggest advantage over SEM approaches is that SFE not only identifies bias but also provides information about the root causes of the bias. SFE allows simultaneously analysis about treatment effectiveness, causal factors of outcomes, and covariates to the bias, improving the statistical efficiency of the analysis over traditional SEM which often cannot identify causal factors and covariates to bias, and when it can, it requires two-step procedures. And since SFE allows the researcher to identify bias and causal factors at the individual level, it expands our ability to identify, understand, explain, and potentially correct for, response shift bias. Of course, bias at the individual level can be aggregated to measures comparable to what is learned through SEM approaches.

Acknowledgments

Biography

Robert Rosenman is a Professor of Economics in the School of Economic Sciences at Washington State University. His current research aims to develop new approaches to measure of economic benefits of substance abuse prevention programmes. His research has appeared in journals such as the *American Economic Review*, *Health Economics*, *Clinical Infectious Diseases* and *Health Care Management Science*.

Vidhura Tennekoon is a Graduate student in the School of Economic Sciences at Washington State University. His research interests are in health economics, applied econometrics and prevention science with a current research focus in dealing with misclassification in survey data.

Laura G. Hill is a Psychologist and Associate Professor of Human Development at Washington State University. Her research focuses on the translation of evidence-based prevention programmes from research to practice and measurement of programme effectiveness in uncontrolled settings.

References

Aigner D, Lovell CAK, Schmidt P. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics. 1977; 6(1):21–37.

Battese GE, Coelli TJ. A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics. 1995; 202:325–332.

Econometrica Software, Inc.. LIMDEP Version 9.0 [Computer Software]. Econometrica Software, Inc.; Plainview, NY: 2009.

Howard GS. Response-shift bias: a problem in evaluating interventions with pre/post self-reports. Evaluation Review. 1980; 41:93–106. DOI: 10.1177/0193841x8000400105.

Howard GS, Dailey PR. Response-shift bias: a source of contamination of self-report measures. Journal of Applied Psychology. 1979; 64(2):144–150.

Kumpfer, KL.; Molgaard, V.; Spoth, R. The strengthening families program for the prevention of delinquency and drug use. In: Peters, RD.; McMahon, RJ., editors. Preventing Childhood Disorders, Substance Abuse, and Delinquency, Banff International Behavioral Science Series. Vol. 3. Sage Publications, Inc.; Thousand Oaks, CA, USA: 1996. p. 241-267.

Meeusen W, van den Broeck J. Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review. 1977; 182:435–444.

Norman G. Hi! How are you? Response shift, implicit theories and differing epistemologies. Quality of Life Research. 2003; 123:239–249. [PubMed: 12769136]

Oort FJ. Using structural equation modeling to detect response shifts and true change. Quality of Life Research. 2005; 143:587–598. [PubMed: 16022054]

Oort FJ, Visser MRM, Sprangers MAG. An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. Quality of Life Research. 2005; 143:599–609. [PubMed: 16022055]

Pratt CC, McGuigan WM, Katzev AR. Measuring program outcomes: using retrospective pretest methodology. American Journal of Evaluation. 2000; 21(3):341–349.

Schwartz CE, Sprangers MAG. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Social Science & Medicine. 1999; 48(11):1531–1548. [PubMed: 10400255]

Spoth R, Redmond C, Shin C. Direct and indirect latent-variable parenting outcomes of two universal family-focused preventive interventions: Extending a public health-oriented research base. Journal of Consulting and Clinical Psychology. 1998; 66(2):385–399. DOI: 10.1037/0022-006x.66.2.385. [PubMed: 9583342]

Spoth R, Redmond C, Haggerty K, Ward T. A controlled parenting skills outcome study examining individual difference and attendance effects. Journal of Marriage and Family. 1995; 57(2):449–464. DOI: 10.2307/353698.

Sprangers M. Subject bias and the retrospective pretest in retrospect. Bulletin of the Psychonomic Society. 1989; 27(1):11–14.

Sprangers M, Hoogstraten J. Pretesting effects in retrospective pretest-posttest designs. Journal of Applied Psychology. 1989; 74(2):265–272. DOI: 10.1037/0021-9010.74.2.265.

StataCorp.. Stata Statistical Software: Release 11 [Computer Software]. StataCorp LP; College Station, TX: 2009.

Table 1

Variable names, descriptions and summary statistics

Name	Description	M	SD
Pretest functioning	Semi-continuous (0-5)	3.979	0.546
Posttest functioning	Semi-continuous (0-5)	4.273	0.461
Male	If Male = 1	0.250	0.433
Gender missing	If gender not reported = 1	0.030	0.170
White	If White = 1	0.601	0.490
Black	If Black = 1	0.023	0.150
Latino/Hispanic	If Latino/Hispanic = 1	0.269	0.443
Native American	If Native American = 1	0.040	0.195
Other	If Other race/ethnicity = 1	0.034	0.182
Race missing	If race not reported = 1	0.034	0.182
Age	Integer (17-73)	38.822	7.846
Partner or spouse	If Partner or spouse in family = 1	0.736	0.441
Partner or spouse missing	If Partner or spouse in family not reported = 1	0.077	0.266
Partner or spouse attends	If Partner or spouse attended SFP = 1	0.499	0.500
Washington State	If family lives in Washington State = 1	0.622	0.485

Table 2

SFE - total effects model

Variable	P	SE	Z	p < Z
Functioning				
Treatment	0.156	0.027	5.87	0.000
Male	−0.119	0.020	−6.03	0.000
Gender missing	−0.018	0.058	−0.30	0.760
Black	0.167	0.054	3.11	0.002
Latino/Hispanic	0.256	0.029	8.86	0.000
Native American	0.090	0.043	2.08	0.038
Other	0.174	0.045	3.83	0.000
Race missing	0.113	0.054	2.08	0.038
Age	−0.005	0.001	−3.92	0.000
Partner or spouse	−0.026	0.022	−1.18	0.237
Partner or spouse missing	−0.062	0.037	−1.70	0.090
Washington State	0.023	0.018	1.31	0.189
Constant	4.605	0.054	85.63	0.000
μ				
Treatment	−1.195	0.407	−2.94	0.003
Hispanic	1.100	0.383	2.87	0.004
Age	−0.052	0.028	−1.88	0.061
lnsigma2	0.291	0.201	1.00	0.317
inlgtgamma	2.559	0.263	9.72	0.000
σ^2	1.338	0.389		
γ^2	0.928	0.018		
σ^2_u	1.242	0.383		
σ^2_ε	0.096	0.010		
Wald $\chi^2(15) = 331.46$				
Prob > $\chi^2 = 0.000$				

Table 3

Averages of bias and change

Variable	M	SD
Estimated u, pre-treatment	0.469	0.368
Estimated u, post-treatment	0.335	0.273
Change in u, post minus pre	−0.133	0.346

Table 4

Regression of bias change

Dependent variable: change in bias	β	SE	t	p < t
Male	0.050	0.023	2.19	0.029
Gender missing	0.100	0.064	1.55	0.122
Black	0.114	0.062	1.84	0.066
Latino/Hispanic	0.015	0.022	0.68	0.496
Native American	0.048	0.047	1.02	0.308
Other	0.078	0.051	1.54	0.125
Race/ethnicity missing	−0.147	0.061	−2.42	0.016
Age	0.003	0.001	2.74	0.006
Partner or spouse	0.032	0.028	1.13	0.258
Partner or spouse information missing	0.051	0.040	1.27	0.203
Washington State	−0.002	0.020	−0.11	0.912
Partner or spo use attended	−0.009	0.024	−0.36	0.721
Constant	−0.303	0.054	−5.65	0.000
Source	Sum of square errors		df	F(12, 1424) = 2.4
Model	3.408042		12	Prob. >F = 0.0044
Residual	168.2181		1,424	R-squared = 0.019
Total	171.6262		1,436	

Table 5

SFE with heteroscedasticity

Variable	P	SE	Z	p<Z
Functioning				
Treatment	0.222	0.032	6.94	0.000
Male	-0.098	0.019	-5.11	0.000
Gender missing	0.002	0.057	0.04	0.970
African Americans	0.159	0.054	2.95	0.003
Hispanic	0.344	0.035	9.95	0.000
Native American	0.096	0.042	2.27	0.023
Other	0.158	0.044	3.63	0.000
Race missing	0.090	0.053	1.69	0.091
Age	−0.001	0.002	−0.65	0.516
Partner or spouse	−0.027	0.021	−1.29	0.199
Partner or spouse missing	−0.044	0.035	−1.25	0.213
Washington State	0.017	0.017	0.98	0.325
Constant	4.532	0.088	51.55	0.000
Ln (σ^2_s)				
Treatment	−0.715	0.187	−3.81	0.000
Hispanic	−1.132	0.288	−3.94	0.000
Age	−0.007	0.010	−0.66	0.512
Constant	−1.906	0.434	−4.39	0.000
ln (σ^2_u)				
Treatment	−0.247	0.116	−2.13	0.033
Hispanic	0.913	0.123	7.42	0.000
Age	−0.005	0.007	−0.67	0.504
Constant	−0.761	0.319	−2.39	0.017
Wald $\chi^2(12) = 253.60$				
Prob. $> \chi^2 = 0.000$				