Psychology of Addictive Behaviors
2015, Vol. 29, No. 1, 231–236

0893-164X/15/$12.00 http://dx.doi.org/10.1037/adb0000015

BRIEF REPORT

Two Alternative Approaches to Conventional Person-Mean Imputation Scoring of the Self-Rating of the Effects of Alcohol Scale (SRE)

Matthew R. Lee, Bruce D. Bartholow, and Denis M. McCarthy
University of Missouri and the Midwest Alcoholism Research Center, Columbia, Missouri

Sarah L. Pedersen
University of Pittsburgh

Kenneth J. Sher
University of Missouri and the Midwest Alcoholism Research Center, Columbia, Missouri

A low level of response to alcohol is considered a significant risk factor for alcohol use disorder. Survey measures of this construct assess the number of drinks required to experience various alcohol effects, so data will be missing for effects participants have not experienced. Furthermore, missingness will likely be more common for items with higher means, as more severe effects are likely experienced both less commonly and at higher consumption levels. We explored whether these atypical characteristics of response-to-alcohol survey data cause problems when using conventional person-mean imputation scoring. This scoring approach involves averaging across nonmissing items for each participant, implicitly assuming that missing items have similar distributional properties (e.g., means) as nonmissing items. Analyses used data from the most commonly utilized response-to-alcohol survey measure: The Self-Rating of the Effects of Alcohol Scale (SRE). Results (a) revealed a strong relationship between higher item means and greater item missingness, (b) established that this relation causes person-mean imputation to produce more downwardly biased response-to-alcohol summary scores for participants with more missing data, (c) established that this induced a spurious relationship between higher response-to-alcohol summary scores and higher alcohol-effect endorsement (i.e., the number of SRE alcohol effects experienced), and (d) found that these biases can be reduced with 2 alternative scoring approaches. We discuss these and other potential problems with person-mean imputation, and common and unique advantages of the 2 alternative approaches. We consider generalizability, including how the problems shown here may vary in practical significance across different populations and measures.

Keywords: SRE, low level of response to alcohol, alcohol sensitivity, psychometrics, assessment

Supplemental materials: http://dx.doi.org/10.1037/adb0000015.supp

Low level of response to alcohol can be defined as requiring larger quantities of alcohol than others to experience equivalent alcohol effects. A great deal of past research supports this construct's importance to the etiology of problem drinking and alcohol use disorders (AUDs; Morean & Corbin, 2010; Schuckit et al., 2011). Although level of response to alcohol can be assessed objectively in the laboratory, such assessments are impractical in many contexts, and survey measures are now widely used because (a) they can be easily administered to large samples; (b) they can be administered to subgroups where alcohol administration may be unethical (e.g., adolescents, those with AUDs); and (c) they can assess severe alcohol effects (e.g., passing out) that cannot be ethically induced via alcohol administration. The most widely used response-to-alcohol survey measure is the 12-item Self-Rating of the Effects of Alcohol Scale (SRE; Schuckit, Smith, & Tipp, 1997), and research has supported its use as a proxy for laboratory

This article was published Online First August 18, 2014.

Matthew R. Lee, Bruce D. Bartholow, and Denis M. McCarthy, Department of Psychological Sciences, University of Missouri and the Midwest Alcoholism Research Center, Columbia, Missouri; Sarah L. Pedersen, Department of Psychiatry, University of Pittsburgh; Kenneth J. Sher, Department of Psychological Sciences, University of Missouri, and the Midwest Alcoholism Research Center, Columbia, Missouri.

This research was supported by National Institute on Alcohol Abuse and Alcoholism Grants P60-AA011998 to Andrew Heath (sub# 5979 to Bruce D. Bartholow), R21 AA015218 to Denis M. McCarthy, F31 AA017571 to Sarah L. Pedersen, and T32-AA013526 and K05-AA017242 to Kenneth J. Sher.

Correspondence concerning this article should be addressed to Matthew R. Lee, 200 South 7th Street, Department of Psychological Sciences, University of Missouri, Columbia, MO 65211. E-mail: LeeMat@Missouri.edu

assessments (Schuckit, Smith, Trim, Tolentino, & Hall, 2010). Given the potential importance of response-to-alcohol survey measures, surprisingly little attention has been paid to their psychometric properties. As we discuss below, these measures produce data with some atypical characteristics that may cause problems when computing summary scores with person-mean imputation (a commonly used scoring approach with these and various other measures).

Person-mean imputation is an approach for computing summary scores while including participants with incomplete data. It involves simply averaging responses across nonmissing items for each participant. However, this means that summary scores often will be computed from different subsets of items for different participants, given that different participants often will be missing different items. This can lead to problems if all items are not similarly distributed. For instance, response-to-alcohol survey measures ask participants the number of drinks they require to experience different alcohol effects, so means will likely be higher for items assessing more versus less severe effects (e.g., "pass out" vs. "feel any different"). This may be particularly problematic because these higher-mean items also will likely have more missing data, as fewer participants will have experienced these more severe effects, and participants cannot report the drinks required for effects they've never experienced. Thus, there will likely be a relationship between higher item means and greater item missingness due to their mutual relations with item severity. Methodologists have warned that such a relationship may cause person-mean imputation to be particularly problematic (Enders, 2010; Tsikriktsis, 2005).

This relationship between item means and item missingness may be problematic because, if higher-mean items have more missing data, then when averaging across nonmissing items (i.e., conducting person-mean imputation), summary scores for participants with more missing data will tend to be based on fewer higher-mean items. Thus, there will be a downward bias in summary scores that is more pronounced for those with more missing data. This is particularly critical with response-to-alcohol survey measures because the number of nonmissing items reflects another drinking outcome, *alcohol-effect endorsement* (i.e., the number of alcohol effects experienced). Thus, a downward bias that is greater for participants with more missing data could introduce a spurious association between response-to-alcohol summary scores and alcohol-effect endorsement. Further, this may spuriously alter associations of response-to-alcohol summary scores with other drinking outcomes that correlate with alcohol-effect endorsement.

The current study used the SRE to explore these potential problems with conventional person-mean imputation. We also investigated two alternative scoring approaches that may minimize these problems: (a) *standardized person-mean imputation* where the SRE items were transformed to have more comparable distributions (i.e., converted to z-scores) before averaging nonmissing items; and (b) *factor-score estimation* where factor modeling of the SRE items was used to obtain response-to-alcohol factor scores. Analyses were conducted to (a) test the expected relationship between higher item means and greater item missingness, (b) assess whether this relationship causes person-mean imputation to produce a spurious association between SRE response-to-alcohol summary scores and SRE alcohol-effect endorsement, (c) map an item-level depiction of the summary score bias produced by person-mean imputation, and (d) assess whether person-mean imputation also spuriously alters relations of response-to-alcohol summary scores with other drinking outcomes. Additional analyses used data with simulated missingness among those with complete SRE data to bolster confidence in our conclusions.

Method

The current sample of young adult current drinkers ($N = 881$; $M_{age} = 22.9$, $SD_{age} = 2.4$; 52% female, 77% White, 16% African American; 85% attending/attended college) was formed by combining baseline data from two independent alcohol administration studies. Both samples were recruited via advertisements at a large Midwestern University and in the surrounding area. Both samples excluded nondrinkers and those reporting past substance problems or other serious physical/mental illness. Comparing demographic and study variables between Samples 1 and 2 ($n = 705$, Ellingson, Fleming, Verges, Bartholow, & Sher, in press; $n = 176$, Pedersen & McCarthy, 2013; respectively; see Online Supplements Table 1), significant differences were found only for age ($M = 23.1$ vs. $M = 22.0$; $t(877) = 6.82$, $p < .001$) and race (86% vs. 43% European American; 5% vs. 56% African American; 8% vs. 0% other; $\chi^2(2) = 286.33$, $p < .001$).[1]

The key measure of the current study was the SRE (Schuckit et al., 1997), a 12-item scale that asks participants the number of drinks they require to experience four different alcohol effects (feel any different, feel dizzy or slur speech, stumble or walk uncoordinatedly, and unintentionally pass out or fall asleep), each within three different time frames (in the first five lifetime drinking occasions, in the most recent 3-month drinking period, and in the period of heaviest lifetime drinking). SRE r*esponse to alcohol* was indexed with the three aforementioned scoring approaches. SRE *alcohol-effect endorsement* was indexed by summing the number of nonmissing SRE items. Other drinking outcomes included *past-month alcohol consumption* (frequency*quantity), *past-month heavy drinking frequency* (≥ 5 drinks per occasion), *past month maximum alcohol consumption*, and *past 3-month drinking consequences* (sum of 12 social consequences). For more on measures, see Online Supplements Table 1.

Results

We first tested the hypothesized relationship between higher item means and greater item missingness. A very strong association was revealed when the 12 SRE items were treated as a sample of 12 cases and a correlation was estimated between the items' means and their percentages of missing data ($r = .75$; $p < .001$; see Online Supplements Figure 1).

Next, we computed SRE summary scores with conventional person-mean imputation and the two alternative scoring approaches. In person-mean imputation, for each participant, nonmissing SRE item scores were averaged. In *standardized* person-mean imputation, each of the 12 SRE item variables was converted to z-scores (i.e., transformed to have a mean of zero and a standard deviation of one), and then for each participant, nonmissing *stan-*

[1] By design, Samples 1 and 2 were limited to ages 21–30 and 21–26, respectively. Sample 2 was limited to Whites and African Americans to compare these groups on response to alcohol.

dardized SRE item scores were averaged. For *factor-score estimation*, a series of exploratory and confirmatory factor analyses led us to retain a bifactor model with a global factor indicated by all 12 SRE items and six specific factors for the four SRE alcohol effects and two of the three SRE time frames (CFI = 0.96; RMSEA = 0.12; $\chi^2(34) = 452.9$, $p < .001$; see Online Supplements Figure 2). Factor scores from the global factor were used as the summary scores from factor-score estimation. See Online Supplements for more details of the three approaches.

Using summary scores obtained from the above three scoring approaches, we next investigated whether person-mean imputation appeared to produce a spurious association of SRE summary scores with SRE alcohol-effect endorsement. Indeed, there was a medium-sized (Cohen, 1992) correlation for person-mean imputation ($r = .29$; $p < .001$), whereas this correlation was nonsignificant for standardized person-mean imputation ($r = .02$; $p = .48$) and significant but small for factor-score estimation ($r = .08$; $p = .03$; see Online Supplements Figure 3).

Next, we conducted an item-level investigation of the imputation bias caused by conventional person-mean imputation. Enders (2010) notes that person-mean imputation is equivalent to replacing participants' missing scores with the average of their nonmissing scores. For illustrative purposes, we carried out both person-mean imputation and standardized person-mean imputation in this manner.[2] Figure 1 compares the distribution of observed (i.e., nonmissing) scores with the distribution of imputed scores for each of the 12 SRE items. Figure 1 confirms that, because items with higher means (means of blue points) also have more missing data (red points), person-mean imputation tends to replace missing values on higher-mean items with downwardly biased averages of mostly lower-mean items. For example, the three items assessing the number of drinks required to pass out have both particularly high means (means of blue points) and particularly prevalent missing data (red points), and imputed values for these items appear particularly downwardly biased relative to observed values for these items (red points vs. blue points). Figure 1 also suggests that this problem was avoided with the alternative *standardized* person-mean imputation approach. Supporting these conclusions, when treating the 12 SRE items as a sample of 12 cases, intraclass correlations (ICCs) assessing agreement of the 12 items' imputed score means with their observed score means showed a negative, nonsignificant ICC for person-mean imputation (ICC = −.22; $p = .77$) and a positive, very strong, and significant ICC for standardized person-mean imputation (ICC = .99; $p < .001$).

Next, we evaluated the expectation that conventional person-mean imputation may also spuriously alter associations between SRE summary scores and other drinking outcomes due to their associations with SRE alcohol-effect endorsement. Although subtle, Table 1 shows that correlations of SRE summary scores with four different drinking outcomes were consistently stronger for person-mean imputation than for the two alternative approaches. Furthermore, Table 1 supports the argument that this is due to a "third variable" influence of SRE alcohol-effect endorsement by showing these correlations both overall and within subgroups at low, moderate, and high levels of alcohol-effect endorsement. For instance, for person-mean imputation only, the correlation of SRE summary scores with past-month maximum alcohol consumption was larger overall than in any of the three alcohol-effect endorsement subgroups.

Finally, to bolster confidence in our conclusions, additional analyses were conducted with the full sample's missing data pattern simulated among a subsample with complete SRE data ($n = 516$).[3] As expected, person-mean imputation performed worse than the two alternative approaches regarding the correlations of the three approaches' SRE summary scores with those computed from complete observed data (although all correlations were very large; see Table 2, row 4). Perhaps more important, only conventional person-mean imputation produced SRE summary scores that were significantly correlated with "simulated alcohol-effect endorsement" (i.e., the number of nonmissing items produced by the missing data simulation). Specifically, this correlation was roughly medium-sized, and positive for person-mean imputation ($r = .28$; $p < .001$), but nonsignificant and close to zero for standardized person-mean imputation ($r = .00$; $p = .93$), for factor-score estimation ($r = .01$; $p = .76$), and (most importantly) for complete observed data ($r = .02$; $p = .65$; see Table 2, row 5). This strongly supports the conclusion that this relationship is spuriously driven by bias from person-mean imputation, particularly because in this case alcohol-effect endorsement was randomly assigned through missing data simulation, so a relationship with SRE summary scores would be otherwise unexpected.[4]

Discussion

Our findings warrant caution regarding the common practice of computing scale summary scores with person-mean imputation. Given our expectation that atypical psychometric properties of response-to-alcohol survey data may make person-mean imputation particularly problematic, we used the SRE to show that this scoring approach can introduce systematic bias into summary scores and thereby produce spurious associations with other constructs.

As expected, results showed a very strong relationship where missing data was far more likely for higher-mean (i.e., more severe) SRE items. Our results show that this causes person-mean imputation to produce summary scores that are more downwardly biased for participants with more missing data. That is, results support the conclusion that person-mean imputation assumes all items are similarly distributed, and violating this assumption can cause nontrivial systematic bias in summary scores when distributional variability relates to missingness.

Perhaps most important, results confirmed that systematic bias from person-mean imputation can give the false impression of a substantial association between response to alcohol and alcohol-effect

[2] For item-level *standardized* person-mean imputation, after replacing missing standardized item scores with the average of nonmissing standardized item scores, the resulting 12 variables were returned to an unstandardized metric (by multiplying by the standard deviations and adding the mean of the original variable) to facilitate comparability of results between the two approaches.

[3] For example, 15.8% ($n = 140$) of the full sample was missing only the three "pass out" items, so 15.8% ($n = 82$) of the complete data subsample was randomly assigned to have these three items' scores deleted.

[4] We also used this simulated missingness dataset to conduct item-level person-mean imputation and standardized person-mean imputation (as depicted in Figure 1 for the full sample). Person-mean imputation was less accurate than standardized person-mean imputation in reproducing observed item scores and item means from complete data (see Online Supplements Table 3).

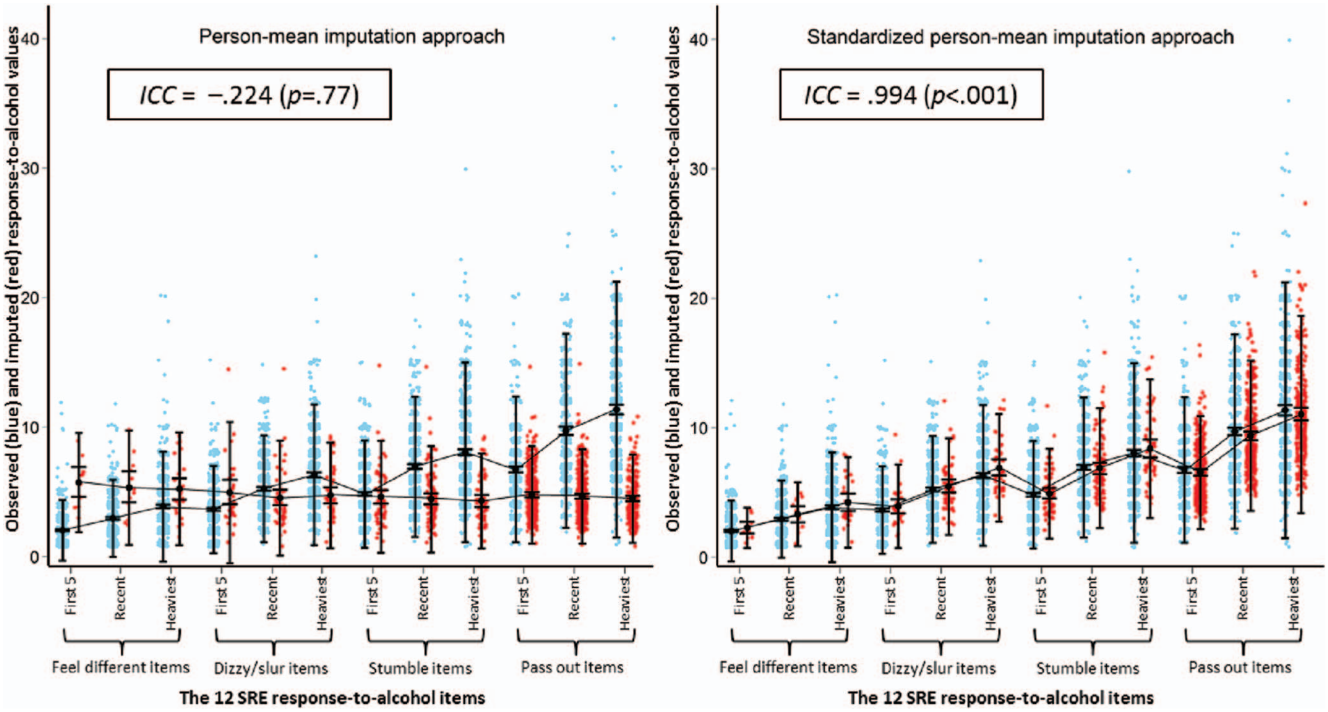

Figure 1. Characterizing the item-level imputation bias of person-mean imputation: Contrasting distributions of response to alcohol (*y*-axis) observed scores (blue points) versus imputed scores (red points) across the 12 SRE items (*x*-axis), separately for person-mean imputation (left) and standardized person-mean imputation (right). Black dots (with connecting lines) show means, wide bars show 95% mean confidence limits, and narrow bars encompass scores within two standard deviations of the mean. Reported intraclass correlations (ICCs; equation (3,1) from Shrout & Fleiss, 1979) assess agreement between observed value means and imputed value means when treating the 12 SRE items as a sample of 12 cases. For *x*-axis labels that distinguish among the 12 SRE items, "First 5" = during the first five lifetime drinking occasions, "Recent" = during the most recent 3-month drinking period, and "Heaviest" = during the period of heaviest lifetime drinking.

endorsement. This could lead to inaccurate substantive conclusions regarding these constructs' relations to one another and to problem drinking etiology, and it calls for reanalysis of past evidence for this relationship from studies where person-mean imputation was used (e.g., Schuckit, Smith, Trim, Fukukura, & Allen, 2009). Results similarly suggested that person-mean imputation can alter associations of response to alcohol with other drinking outcomes, and although this was relatively subtle in the current analyses, it could be more pronounced in other contexts. For instance, our participants were young adult current drinkers volunteering for an alcohol admin-

Table 1
Comparing the Three Scoring Approaches on Associations of Response-to-Alcohol Summary Scores With Other Drinking Outcomes: Correlations Overall and Among Three Subgroups at Three Different Levels of Alcohol-Effect Endorsement

	Response-to-alcohol summary scores from the three approaches											
	Person-mean imputation				Standardized person-mean imputation				Factor-score estimation			
		Endorsement levels				Endorsement levels				Endorsement levels		
Drinking outcomes	Overall	1–9	10–11	12	Overall	1–9	10–11	12	Overall	1–9	10–11	12
Alcohol-effect endorsement	.29**	—	—	—	.018	—	—	—	.077*	—	—	—
Other drinking outcomes												
Past-month consumption	.48**	.47**	.48**	.41**	.40**	.38**	.42**	.40**	.46**	.46**	.50**	.43**
Past-month heavy drinking	.47**	.49**	.52**	.37**	.38**	.40**	.46**	.36**	.44**	.50**	.51**	.39**
Past-month max. consumption	.57**	.48**	.54**	.54**	.47**	.38**	.46**	.52**	.54**	.48**	.47**	.56**
Past-3-month consequences	.18**	.05	.18	.14**	.12**	.01	.14	.14**	.16**	.05	.14	.16**

Note. For the three endorsement level subgroups, $n = 255$ for subgroup 1–9; $n = 110$ for subgroup 10–11; and $n = 516$ for subgroup 12.
** $p < .01$.

Table 2
Contrasting the Three Scoring Approaches Using Data With Simulated Missingness Among Those With Complete SRE Data (n = 516): Correlations of SRE Summary Scores From the Three Approaches With SRE Summary Scores From Observed Complete Data and With "Simulated Alcohol-Effect Endorsement"

	1	2	3	4	5
1. Person-mean imputation SRE summary scores	1	—	—	—	—
2. Standardized person-mean imputation SRE summary scores	.93***	1	—	—	—
3. Factor-score estimation SRE summary scores	.89***	.93***	1	—	—
4. Observed complete data SRE summary scores	**.91****	**.94****	**.94****	1	—
5. Simulated alcohol-effect endorsement[a]	**.28****	**.00**	**.01**	**.02**	1

Note. Rows 4 and 5 are in bold text to highlight (row 4) correlations of SRE summary scores from the 3 scoring approaches with SRE summary scores from observed complete data and (row 5) correlations of SRE summary scores from the 3 scoring approaches and from observed complete data with "simulated alcohol effect endorsement." Only participants who were randomly assigned to have at least one missing score were included ($n = 228$) because participants who were randomly assigned to have no missing scores ($n = 288$) had exactly the same values across the four response-to-alcohol summary scores.
[a] The number of nonmissing items for a given participant, which in this case of simulated missingness was purely a function of the random assignment of different missing data patterns to different participants.
*** $p < .001$.

istration study, and the problems demonstrated here may be exacerbated in younger samples with more missing data (i.e., nonendorsement) for relatively severe alcohol effects. This is important, given that survey measures can be especially useful in younger samples (e.g., adolescents) where laboratory assessment may be considered unethical. Conversely, the problems shown here may be less pronounced among older individuals or more severe drinkers where less missing data would be expected. To extend our findings, it may be particularly useful to (a) use prospective data to assess the impact of decreases in missing data as participants age, and (b) use simulated data to assess the impact of systematically varying different potentially relevant factors (e.g., missing patterns and mean differences across items).

Of course, the current evidence for problems with person-mean imputation should not be generalized to all applications of this approach. Indeed, this approach has been validated in other contexts (e.g., Hawthorne & Elliott, 2005; Roth, Switzer, & Switzer, 1999), and as stated earlier, the extent of the problems shown here can likely be attributed to atypical psychometric properties of the SRE and other response-to-alcohol surveys (e.g., the Alcohol Sensitivity Questionnaire; Bartholow et al., 2003). However, other measures may share these properties and thus warrant similar concerns. For instance, certain fear/anxiety measures ask participants to rate the fear they experience from various stimuli (e.g., the Fear Survey Schedule for Children-II; Gullone & King, 1992), and person-mean imputation may be problematic with these measures if more fear-provoking stimuli are also less frequently experienced and reported upon (i.e., if item means relate to item missingness). Further, person-mean imputation may be at least somewhat problematic even with measures where there is substantial item-mean variability that is unrelated to missingness. Although this may not cause systematic bias, it would likely introduce random error into summary scores (i.e., randomly distributed deviations from "true scores"), which can decrease statistical power to detect associations with other variables (Williams, Zimmerman, & Zumbo, 1995). Future research (e.g., simulations) should assess whether there are circumstances where this could result in practically important reductions in statistical power. In addition to avoiding this problem, as discussed below, there are other likely advantages to our alternative scoring approaches that were not directly highlighted in our analyses (see Online Supplements for a detailed discussion of considerations when choosing among the three scoring approaches).

Standardized person-mean imputation was shown to be an alternative approach that requires minimal statistical expertise, but nonetheless addresses a key problem with conventional person-mean imputation by avoiding the assumption that all items have similar distributions (for support for a similar approach called "corrected person-mean imputation," see Huisman, 2000). By transforming items to have more comparable distributions (i.e., converting to z-scores) prior to averaging, standardized person-mean imputation assumes only that participants' elevation relative to the mean should be similar across items. This assumption of participant rank-order stability across items is implicit for any scale where multiple items are considered to indicate a single construct (i.e., items are assumed to be at least moderately intercorrelated; e.g., Clark & Watson, 1995). Beyond the advantages directly highlighted in our analyses, it is also noteworthy that conventional person-mean imputation implicitly weights higher-mean items more heavily, whereas standardized person-mean imputation achieves more equal weighting, thus avoiding a disproportionately large influence of higher-mean items on summary scores.[5] This advantage likely applies generally to any scale with distributional (e.g., mean) variability across items.

Factor-score estimation, although a more complex alternative approach, offered the additional advantage of providing a relatively pure index of response to alcohol by partialing out other sources of item-level variance. This appears to be particularly important for the SRE, given that factor modeling revealed complex multidimensionality among the 12 SRE items. Due to this

[5] However, standardized person-mean imputation with unequal weighting can also be achieved by imputing standardized missing values at the item level and then returning the resulting standardized item variables to their original metric prior to averaging (as we did for Figure 1).

multidimensionality, a unitary index of response to alcohol could only be obtained through modeling additional specific latent factors for different SRE alcohol effects and time frames. However, advantages of factor-score estimation likely extend beyond measures with this type of multidimensionality, given that factor modeling could still then be useful for partialing out item-specific sources of measurement error (e.g., see Ledgerwood & Shrout, 2011).

Limitations of our study include use of a somewhat demographically and age-restricted sample of mostly college/postcollege young adults. As noted above, the concerns raised by our results likely vary in practical significance across populations, so future research should explore this variability. Our approach is also limited relative to data simulation methods where relevant factors can be systematically varied, although our approach has the advantage of showing that these problems can occur under conditions found in real data. Our findings also have limited direct relevance to research using laboratory assessments of response to alcohol, as they primarily inform research using survey assessments, although this is a worthwhile objective given the common use and unique functions of response-to-alcohol surveys. Despite its limitations, this study contributes important insights that relate broadly to the common practice of person-mean imputation, with particularly clear and direct implications for survey-based studies of response to alcohol. Although clearly raising concerns regarding person-mean imputation in at least some circumstance, findings also support two straightforward alternative approaches for addressing these concerns.

References

Bartholow, B. D., Pearson, M., Sher, K. J., Wieman, L. C., Fabiani, M., & Gratton, G. (2003). Effects of alcohol consumption and alcohol susceptibility on cognition: A psychophysiological examination. *Biological Psychology, 64,* 167–190. doi:10.1016/S0301-0511(03)00108-X

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological assessment, 7,* 309–319. doi:10.1037/1040-3590.7.3.309

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. doi:10.1037/0033-2909.112.1.155

Ellingson, J. M., Fleming, K. A., Verges, A., Bartholow, B. D., & Sher, K. J. (in press). Working memory as a moderator of impulsivity and alcohol involvement: Testing the cognitive motivational theory of alcohol use with prospective and working memory updating data. *Addictive Behaviors*.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Gullone, E., & King, N. J. (1992). Psychometric evaluation of a revised fear survey schedule for children and adolescents. *The Journal of Child Psychology and Psychiatry, 33,* 987–998. doi:10.1111/j.1469-7610.1992.tb00920.x

Hawthorne, G., & Elliott, P. (2005). Imputing cross-sectional missing data: Comparison of common techniques. *Australian and New Zealand Journal of Psychiatry, 39,* 583–590. doi:10.1080/j.1440-1614.2005.01630.x

Huisman, M. (2000). Imputation of missing item responses: Some simple techniques. *Quality and Quantity, 34,* 331–351. doi:10.1023/A:1004782230065

Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology, 101,* 1174–1188. doi:10.1037/a0024776

Morean, M. E., & Corbin, W. R. (2010). Subjective response to alcohol: A critical review of the literature. *Alcoholism: Clinical and Experimental Research, 34,* 385–395. doi:10.1111/j.1530-0277.2009.01103.x

Pedersen, S. L., & McCarthy, D. M. (2013). Differences in acute response to alcohol between African Americans and European Americans. *Alcoholism: Clinical and Experimental Research*. Advance online publication. doi:10.1111/acer.12068

Roth, P. L., Switzer, F. S., & Switzer, D. M. (1999). Missing data in multiple item scales: A Monte Carlo analysis of missing data techniques. *Organizational Research Methods, 2,* 211–232. doi:10.1177/109442819923001

Schuckit, M. A., Smith, T. L., & Tipp, J. E. (1997). The self-rating of the effects of alcohol (SRE) form as a retrospective measure of the risk for alcoholism. *Addiction, 92,* 979–988. doi:10.1111/j.1360-0443.1997.tb02977.x

Schuckit, M. A., Smith, T. L., Trim, R. S., Allen, R. C., Fukukura, T., Knight, E. E., . . . Kreikebaum, S. A. (2011). A prospective evaluation of how a low level of response to alcohol predicts later heavy drinking and alcohol problems. *The American Journal of Drug and Alcohol Abuse, 37,* 479–486. doi:10.3109/00952990.2011.598590

Schuckit, M. A., Smith, T. L., Trim, R., Fukukura, T., & Allen, R. (2009). The overlap in predicting alcohol outcome for two measures of the level of response to alcohol. *Alcoholism: Clinical and Experimental Research, 33,* 563–569. doi:10.1111/j.1530-0277.2008.00870.x

Schuckit, M. A., Smith, T. L., Trim, R. S., Tolentino, N. J., & Hall, S. A. (2010). Comparing structural equation models that use different measures of the level of response to alcohol. *Alcoholism: Clinical and Experimental Research, 34,* 861–868. doi:10.1111/j.1530-0277.2010.01158.x

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428. doi:10.1037/0033-2909.86.2.420

Tsikriktsis, N. (2005). A review of techniques for treating missing data in OM survey research. *Journal of Operations Management, 24,* 53–62. doi:10.1016/j.jom.2005.03.001

Williams, R. H., Zimmerman, D. W., & Zumbo, B. D. (1995). Impact of measurement error on statistical power: Review of an old paradox. *Journal of Experimental Education, 63,* 363–370. doi:10.1080/00220973.1995.9943470

Received January 17, 2014
Revision received June 10, 2014
Accepted June 10, 2014 ■