Data supplement for Ray et al., Efficacy of Combining Varenicline and Naltrexone for Smoking Cessation and Drinking Reduction: A Randomized Clinical Trial. Am J Psychiatry (doi: 10.1176/appi.ajp.2020.20070993)

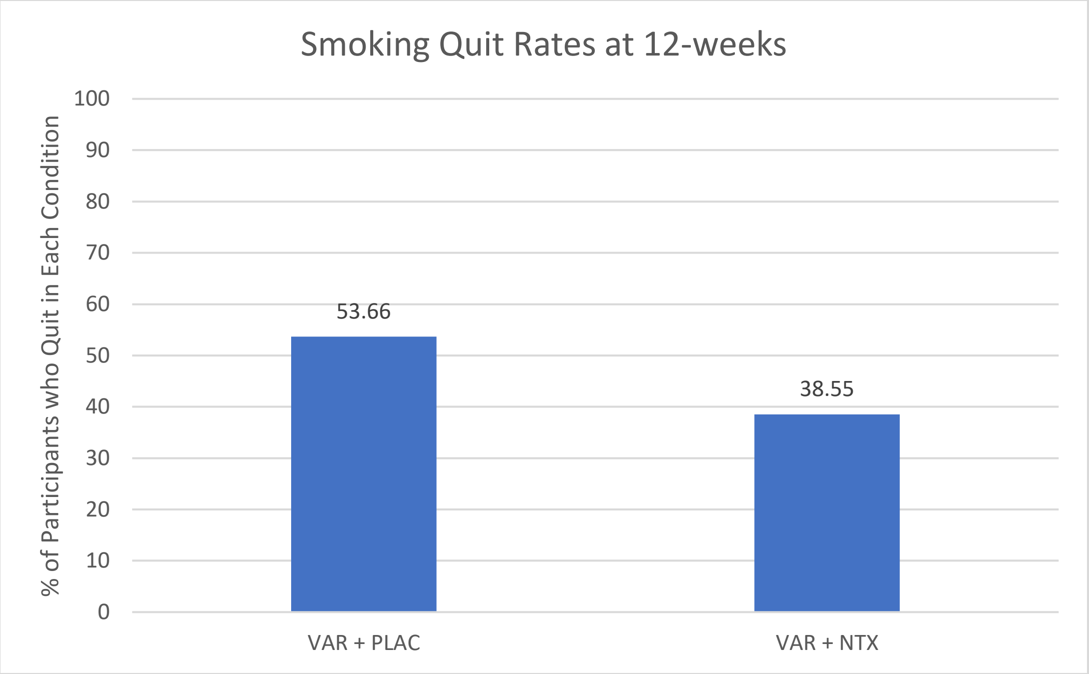
TABLE S1. Adverse Events by Medication Condition

Adverse Event (No.)	Varenicline + Placebo	Varenicline + Naltrexone	Test for Difference
Gastrointestinal (9)	**58**	**67**	**χ^2 = .83, *p* = .36**
Nausea	29	39	
Vomiting	12	9	
Constipation	4	5	
Diarrhea	3	4	
Upset stomach	9	4	
Decreased appetite	0	4	
Increased appetite	0	1	
Decreased bowel movements	0	1	
Acid reflux	1	0	
Central nervous system / psychiatric (19)	**64**	**80**	**χ^2 = 2.58, *p* = .11**
Headache	10	17	
Migraine	0	1	
Decreased concentration	0	1	
Loss of consciousness	1	0	
Tired	3	10	
Fatigue	5	4	
Dizziness	2	1	
Abnormal dreams	18	13	
Trouble sleeping	5	9	
Anxiety	3	3	
Irritability	11	6	
Jittery / Shaky	2	4	
Depression	1	6	
Apathetic	1	2	
Confusion	0	1	
Mood swings	0	1	
Tension	1	0	
Seizure	1	0	

Hot flashes	0	1	
Dental and oral (3)	**3**	**0**	**Fisher’s Exact Test** ***p* = .122**
Dental pain	1	0	
Mouth pain from wisdom tooth growth	1	0	
Hypersensitivity teeth and mouth	1	0	
Ear, nose, throat (8)	**8**	**5**	**χ^2 = .75, *p* = .39**
Earache	1	0	
Ear infection	1	1	
Tinnitus	1	0	
Bloody nose	1	0	
Runny nose	1	0	
Sinus infection	0	1	
Sore throat	1	0	
Dry mouth	2	3	
Cardiopulmonary (9)	**21**	**14**	**χ^2 = 1.62, *p* = .20**
Cough	0	2	
Cold	12	4	
Flue	5	2	
Elevated blood pressure	1	1	
Rapid heartbeat	0	2	
Chest discomfort	1	1	
Difficulty breathing	1	0	
Respiratory infection	0	1	
Cardiac event	1	1	
Skin (3)	**13**	**8**	**χ^2 = 1.31, *p* = .25**
Sweating	1	0	
Night sweats	1	0	
Skin irritation	11	8	
Musculoskeletal (2)	**7**	**8**	**χ^2 = .06, *p* = .81**
Joint pain or swelling	6	8	
Fractured ribs	1	0	
Genitourinary (2)	**0**	**3**	**Fisher’s Exact Test** ***p* = .25**
UTI	0	1	
Neon green urine	0	2	

Ophthamological (3)	**2**	**1**	**Fisher's Exact Test** ***p* = .62**
Pink eye	0	1	
Eye twitch	1	0	
Light sensitivity	1	0	
Reproductive (3)	**4**	**2**	**Fisher's Exact Test** ***p* = .45**
Uterine biopsy procedure	0	1	
Bartholin duct cyst	0	1	
Yeast infection	4	0	
Endocrine (3)	**1**	**2**	**Fisher's Exact Test** ***p* = 1.0**
Low blood sugar	1	0	
Borderline diabetes	0	1	
Hyperglycemia	0	1	
General Disorders (10)	**11**	**4**	$\chi^2 = 3.48$, $p = .06$
Fever	3	1	
Feeling high	1	0	
Dehydration	1	1	
Extra thirsty	1	0	
Hair loss	1	0	
Decreased sex drive	1	1	
Weight gain	1	0	
Anemia	1	0	
Metallic taste	0	1	
Elevated cholesterol	1	0	

FIGURE S1. At the 12-week assessment, when participants were still on active medication, the quit rate was 53.66% (44/82 participants quit) in the varenicline only condition, compared to 38.55% quit rate (32/83 participants quit in the varenicline plus naltrexone condition, [χ^2(1, N=165) = 3.79, *p*=0.051].



Sensitivity Analysis for Smoking Quit Rates

A long-standing approach in the literature is to count dropouts (missing data at week 26) all as quit rate failures (e.g., Hall et al., 2001). We henceforth refer to this as *logical imputation* because it imputes the missing quit status scores with a deterministic failure. While there is nothing inherently wrong about this procedure, it is important to note that deterministic imputation assumes a particular missing data process that could be impactful on parameter estimates. To explore this issue, we performed a sensitivity analysis that applied modern missing data handling procedures with different assumptions about dropout. In particular, we used Bayesian estimation and model-based multiple imputation procedures (Du, Enders, Keller, Bradbury, & Karney, in press; Enders, Du, & Keller, 2020) to explore that possibility that missingness at week 26 is explained by (a) observed data and covariates from prior measurement occasions (e.g., treatment assignment or one's baselines scores could predict later dropout), or (b) a participant's unobserved smoking status at week 26 (e.g., participants with quit failures at the end of the study could be more likely to attrit). These processes are known as missing at random (MAR) and missing not at random (MNAR) mechanisms. These mechanisms represent very different systematic dropout processes, both of which could be plausible in this study.

We used the Blimp software application (Keller & Enders, 2020) to implement model-based multiple imputation routine that produced 10 imputed data sets under four different assumptions about the missingness process. Model-based imputation tailors imputations around a regression model that predicts the incomplete week 26 outcome from other variables from previous measurements.

$$CO_{26i} = \beta_0 + \beta_1 CO_{0i} + \beta_2 TX_i + \beta_3 (TX_i)(CO_{0i}) + \text{covariates} + \varepsilon_i \tag{1}$$

Alveolar CO level is used in this case because imputing the continuous scores then categorizing imputes will maximize power. Note that this is not the analysis of substantive interest. Rather, it is a regression model that preserves important features of the data (e.g., treatment group differences). In addition to baseline CO levels and the interaction with treatment status, the imputation regression model also included gender, the number of drinks per drinking day in the last 28 days, number of cigarettes per smoking day in the last 28 days, and a measure of nicotine dependence. This imputation model assumes a missing at random process where a missing smoking status score at the final assessment is systematically related to the predictors in the imputation model.

A second set of models were fit that assumed a missing not at random process where one's unseen CO value at week 26 predicts missingness. Selection models were used for this purpose (Heckman, 1976; Heckman, 1979). A selection model for missing data pairs the regression in Equation 1 with a probit regression model with the binary missing data indicator as the dependent variable

$$M_{26i} = \gamma_0 + \gamma_1 CO_{26i} + r_i \tag{2}$$

Where $M_{26} = 0$ if the week 26 smoking status score is observed and $M_{26} = 1$ if it is missing. Gomer and Yuan (in press) refer to Equation 2 as a *focused MNAR* process because missingness depends on CO scores at the final assessment, some of which are missing. They refer to a *diffuse MNAR* process as one where additional variables influence missingness above and beyond scores at the final assessment. Diffuse models are difficult to estimate and generally require large samples (Du et al., in press), but we carefully explore this mechanism by adding individual covariates and checking model fit. We considered diffuse models that added treatment group assignment and gender to the missingness model.

Finally, it is well known that selection models rely heavily on the normality assumption for the incomplete variable. Because the CO levels are positively skewed, we performed a second set of analyses where all CO measures were log transformed in the imputation model then back-transformed to the original metric post-imputation. After creating imputations under the various assumptions about the missing data process, we categorized the CO values such that participants with values less than 6 were classified as quit successes and values of 6 or greater were quit failures. We then performed logistic regression analyses on the imputed data to determine whether quit rates differed between the two treatment regimes.

Table S2 shows the predicted probabilities of quitting (i.e., the cell proportions from a two-way contingency table) for the various missing data assumptions, along with single degree of freedom chi-square test statistics that evaluate treatment group differences in quit rates. Not surprisingly, a deterministic assignment scheme (labeled "logical imputation") produced very different cell proportions and quit rates than the missing data models, although the direction of the effects were the same (i.e., the NTX group had higher quit rates than the placebo group). The point estimates and substantive conclusions were remarkably stable across different assumptions about the missing data process. That is, the analysis that assumed systematic dropout based on earlier scores produced effectively equivalent estimates to the analyses that assumed a process where week 26 scores influenced missingness. Further, transforming the non-normal CO score prior to imputation increased estimates by about 1-2%, but the normality assumption clearly didn't play a substantial role.

TABLE S2. Smoking Quit Rates From Missing Data Sensitivity Analysis

Model	VAR + PLAC	VAR + NTX	χ^2	*p*
Logical Imputation	.45	.27	6.11	.01
Raw Metric Imputation				
MAR	.59	.38	4.53	.03
Focused MNAR	.61	.39	7.40	< .01
Diffuse MNAR + Gender	.59	.39	4.09	.04
Diffuse MNAR + Tx Condition	.60	.40	4.88	.03
Logarithmic Transformation Imputation				
MAR	.63	.43	5.24	.02
Focused MNAR	.62	.44	4.38	.04
Diffuse MNAR + Gender	.62	.42	5.86	.02
Diffuse MNAR + Tx Condition	.65	.44	5.41	.02

Importantly, there is no way of knowing which of the models is better, as all make untestable assumptions about the unseen score values. Statistical theory tells us that, if those assumptions are correct, the estimates will be accurate, and they could be biased otherwise. However, the stability of the quit rates across very different assumptions about the missing data process provides strong evidence in favor of the reported treatment effect.

Du, H., Enders, C. K., Keller, B. T., Bradbury, T., & Karney, B. (in press). A Bayesian latent variable selection model for nonignorable missingness. *Multivariate Behavioral Research*.

Enders, C. K., Du, H., & Keller, B. T. (2020). A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and other nonlinear terms. *Psychological Methods, 25*, 88-112. doi:10.1037/met0000228

Gomer, K., & Yuan, K.-H. (in press). Subtypes of the Missing Not at Random Missing Data Mechanism. *Psychological Methods*.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement, 5*, 475-492.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47*(1), 153-161. doi:Doi 10.2307/1912352

Keller, B. T., & Enders, C. K. (2020). Blimp User's Guide (Version 2.2).

Missing Data Sensitivity Analyses for Primary Drinking Outcome

We conducted a series of sensitivity analyses to assess the impact of missing data handling assumptions on the primary analysis results. Linear mixed models such as those employed here assume a so-called missing at random (MAR) process whereby missingness is a function of observed scores but not the underlying drinking scores themselves[1]. In the context of a clinical trial, it is also reasonable to expect that one's underlying drinking outcome could itself determine missingness. To explore this possibility, we estimated the linear mixed model in conjunction with two alternate not missing at random (NMAR) processes[2]. The first, labeled "Slope (Growth) Causes Missingness" in the table below, allows missingness to depend on one's individual change trajectory (i.e., a random coefficient selection model, or shared parameter model)[3]; and the second, labeled "Unobserved Score Causes Missingness" allows missingness at a particular wave to depend on one's unseen alcohol outcome score at that wave (i.e., a selection mode)[4].

In additional to considering different assumptions about the dropout process, we also invoked different distributional assumptions for the analyses. Not surprisingly, the alcohol outcome was positively skewed with a point mass at zero. Linear mixed models assume a normal error distribution for the repeated measurements, and missing data handling procedures explicitly invoke the same assumption. We implemented two alternate approaches. The first was to log transform the alcohol outcome (after adding 1 point to each score to avoid applying the transformation to zero values). The second approach viewed the outcome as a mixture of two distinct subgroups: non-drinkers with scores of zero and drinkers with a continuous distribution comprised of continuous scores. Schafer and Olsen (1999) described a two-part imputation model that (a) first imputes whether the missing score is zero or non-zero, then (b) applies a continuous imputation model to cases designated as having non-zero scores at the first step[5]. The second stage can be applied with or without the log transformation. After generating imputations with the two-stage approach, the linear mixed model was fit to the imputed data, and estimates were summarized using standard pooling rules from Rubin (2004)[6].

The sensitivity analyses included seven different combinations of missing data process and distributional assumption. Table S3 gives the mean difference for the two medications at the 4-week follow-up under each combination of assumptions (the log transformed estimates in the bottom half of the table reflect the difference on a different metric). The bolded rows give the average estimates across the untransformed and transformed estimates. The results in the table suggest that conclusions were robust to different distributional and missingness assumptions, as the estimates were quite stable. For example, the untransformed estimates based on an MAR process and the corresponding NMAR process where one's rate of change predicted missingness was equivalent to about one-tenth of a standard error unit. The estimates were similarly stable on the log transformed metric.

TABLE S3. Estimates of Medication Differences at 4-Week Follow-up From Sensitivity Analysis

Raw Score (Untransformed) Analysis Results	Est.	*SE*	*t*	*p*
Standard Estimation, Observed Data Causes Missingness	-0.856	0.442	-1.94	0.054
Two-Stage Imputation, Observed Data Causes Missingness	-0.782	0.478	-1.640	0.102
Slope (Growth) Causes Missingness	-0.974	0.457	-2.131	0.034
Unobserved Score Causes Missingness	NA	NA	NA	NA
Averaged Effects Over Different Assumptions	-0.871	0.459	-1.897	0.063

Log Transformed Analysis Results	Est.	*SE*	*t*	*p*
Standard Estimation, Observed Data Causes Missingness	-0.226	0.117	-1.930	0.055
Two-Stage Imputation, Observed Data Causes Missingness	-0.213	0.123	-1.740	0.082
Slope (Growth) Causes Missingness	-0.235	0.101	-2.327	0.020
Unobserved Score Causes Missingness	-0.201	0.144	-1.396	0.178
Averaged Effects Over Different Assumptions	-0.213	0.117	-1.826	0.068

TABLE S4. Results for Secondary Drinking Outcomes Across the 12-weeks of Active Medication

	Estimate	***SE***	***DF***	***t***	***p***
Percent Days Abstinent					
Intercept	0.652	0.036	161	18.23	<.0001
Time 1 (BA-4wk)	0.078	0.009	139	8.63	<.0001
Time 2 (4-12wk)	0.006	0.003	484	2.47	0.014
Time 3 (12-26wk)	-0.009	0.004	484	-2.52	0.012
Medication	0.047	0.052	484	0.90	0.368
AUD Severity	-0.006	0.018	484	-0.34	0.735
Medication × Time 1	0.020	0.013	484	1.57	0.118
Medication × Time 2	-0.007	0.004	484	-1.77	0.077
Medication × Time 3	0.009	0.005	484	1.77	0.078
Percent Heavy Drinking Days					
Intercept	0.370	0.041	161	9.14	<.0001
Time 1 (BA-4wk)	-0.066	0.012	139	-5.50	<.0001
Time 2 (4-12wk)	-0.001	0.005	484	-0.24	0.814
Time 3 (12-26wk)	-0.002	0.007	484	-0.34	0.732
Medication	-0.054	0.059	484	-0.92	0.357
AUD Severity	0.080	0.019	484	4.28	<.0001
Medication × Time 1	-0.017	0.017	484	-0.97	0.334
Medication × Time 2	<-0.001	0.007	484	-0.01	0.993
Medication × Time 3	0.004	0.010	484	0.42	0.673
Drinking Days					
Intercept	9.508	0.971	161	9.80	<.0001
Time 1 (BA-4wk)	-2.225	0.247	139	-9.00	<.0001
Time 2 (4-12wk)	-0.159	0.072	484	-2.21	0.028
Time 3 (12-26wk)	0.155	0.101	484	1.54	0.123
Medication	-1.593	1.403	484	-1.14	0.257
AUD Severity	0.181	0.497	484	0.36	0.716
Medication × Time 1	-0.646	0.357	484	-1.81	0.072
Medication × Time 2	0.187	0.106	484	1.77	0.078
Medication × Time 3	-0.260	0.149	484	-1.75	0.081

Models included outcomes of percent days abstinent, percent heavy drinking days, and drinking days across the entire trial, including the active medication period (weeks 4, 8, 12) and the follow-up period (weeks 16 and 26). As with the primary models, these analyses account for the effects of time, medication × time, and AUD severity.

References for Supplementary Materials

1. Little RJ, Rubin DB. *Statistical analysis with missing data.* Vol 793: John Wiley & Sons; 2019.

2. Enders CK. Missing not at random models for latent growth curve analyses. *Psychological methods.* 2011;16(1):1.

3. Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics.* 1989:939-955.

4. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics).* 1994;43(1):49-73.

5. Schafer JL, Olsen MK. Modeling and imputation of semicontinuous survey variables. Paper presented at: Proceedings of the Federal Committee on statistical methodology research conference1999.

6. Rubin DB. *Multiple imputation for nonresponse in surveys.* Vol 81: John Wiley & Sons; 2004.