Biostatistics (2010), **1**, 1, *pp.* 1–20
doi:10.1093/biostatistics/bio000

Supplementary material to: "Allowing for never and episodic consumers when correcting for error in food record measurements of dietary intake"

RUTH H. KEOGH*
MRC Biostatistics Unit, Robinson Way, Cambridge, CB2 0SR, UK
MRC Centre for Nutritional Epidemiology in Cancer Prevention and Survival, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK
ruth.keogh@mrc-bsu.cam.ac.uk

IAN R. WHITE
MRC Biostatistics Unit, Robinson Way, Cambridge, CB2 0SR, UK

A. Alternative methods of analysis

In the crude analysis the log odds ratio (OR) is estimated by replacing T_i in the diet-disease model with the mean of the available set of repeated observed measurements for individual i, $\bar{R}_{i\cdot} = \sum_{j=1}^{J_i} R_{ij}/J_i$, where J_i denotes the number of available repeat measurements for individual i.

*To whom correspondence should be addressed.

Under linear regression calibration the log OR is estimated by replacing T_i with $E(T_i|\bar{R}_{i.})$ for individuals with repeated measurements and with $E(T_i|R_{i1})$ for individuals with only one measurement, where the expectations are the fitted values from the models $T_i = \lambda_0 + \lambda_1 \bar{R}_{i\cdot} + e_i$ and $T_i = \lambda_0' + \lambda_1' R_{i1} + e_i$ respectively. Parameters λ_0, λ_1, λ_0', and λ_1' are estimated by

$$\hat{\lambda}_1 = \frac{\text{cov}(T_i, \bar{R}_{i\cdot})}{\text{var}(\bar{R}_{i\cdot})}, \quad \hat{\lambda}_0 = E(T_i) - \hat{\lambda}_1 E(\bar{R}_{i\cdot}),$$

$$\hat{\lambda}_1' = \frac{\text{cov}(T_i, R_{i1})}{\text{var}(R_{i1})}, \quad \hat{\lambda}_0' = E(T_i) - \hat{\lambda}_1' E(R_{i1}).$$

To estimate $\text{cov}(T_i, \bar{R}_{i\cdot})$, $\text{cov}(T_i, R_{i1})$ and $E(T_i)$ we assume that the observed measurements follow the classical measurement error model $R_{ij} = T_i + \epsilon_{ij}$, where the errors ϵ_{ij} are independent of T_i and of each other. The classical measurement error assumption is equivalent to the assumption that the observed measurements provide unbiased measures of true intake, which was also used to calculate the fitted values $\hat{T}_i(\theta)$ under the NEC model. Under this assumption we estimated $\text{cov}(T_i, \bar{R}_{i\cdot}) = \text{cov}(T_i, R_{i1}) = \text{var}(T_i)$ by $\sum_{j,k(j\neq k)} \text{cov}(R_{ij}, R_{ik})/J(J-1)$, using only individuals with repeat measurements in the unbalanced data situation. $E(T_i)$ was estimated by $\sum_{j=1}^{J} \bar{R}_{\cdot j}/J$, where $\bar{R}_{\cdot j} = \sum_{i=1}^{n_j} R_{ij}/n_j$ and n_j denotes the number of individuals with a jth measurement. These calculations would have to be modified in the situation of more unbalanced data, that is when individuals have variable numbers of measurements.

B. Simulation study

For a study population of 1000 individuals $(i = 1, \ldots, 1000)$ the random effects u_{0i} and (u_{1i}, u_{2i}) were generated randomly from a binomial distribution and a bivariate normal distribution respectively, and within-person errors ϵ_{ij} $(j = 1, \ldots, 10)$ were generated randomly from a normal distribution, using the parameters in Table 1 of the main text. Using equation (2.11), true intake T_i was obtained using the generated values for $\mathbf{u}_i$ and true parameters. As estimated in the EPIC-Norfolk alcohol data, the observed measurements R_{ij} were assumed to be Normally distributed on a Box-Cox transformed scale with $\lambda = 0.25$. The transformed observed measurements R^*_{ij} were generated from the NEC model defined in (2.1-2.3), and R_{ij} were calculated using the inverse transformation.

The simulation was repeated to give 500 simulated data sets, each containing true intake T_i and repeated observed measurements R_{ij} ($j = 1, \ldots, 10$). We repeated the simulation using $H(\gamma_0) = 0.75$ instead of $H(\gamma_0) = 0.88$.

In each simulated data set the NEC model was fitted by maximum likelihood using $J = 2, 4, 10$ repeat measurements per individual, to give estimated parameters $\hat{\theta}$. The integrals required to find the joint distribution $f(\mathbf{R}^*_i; \theta)$ of the transformed observed measurements were evaluated numerically using Gauss-Hermite quadrature. The estimated parameters from the NEC model were used to calculate the fitted values $\hat{T}_i(\hat{\theta})$ in each simulated data set using (2.11). The integrals in (2.11) were evaluated numerically using Gauss-Hermite quadrature.

The NEC model was refitted to reduced data in which only 15% of individuals had a complete set of $J = 2, 4, 10$ repeat measurements while the remaining 85% had just one measurement, R_{i1}, since 15% of EPIC-Norfolk participants currently have two 7-day diary measurements of alcohol intake.

We generated binary outcomes (disease status) according to logistic models with disease probability $\exp(\alpha+\beta T_i)/(1+\exp(\alpha+\beta T_i))$. Parameter β was chosen to give an

OR per 10grams/day increase in alcohol intake of 1.2, 1.5 or 2. Parameter α was fixed at -3, resulting in approximately 6%, 9% and 13% of individuals having the disease for ORs 1.2, 1.5 and 2 respectively. The log OR was estimated using the fitted values $\hat{T}_i(\hat{\theta})$ and also using the crude approach, linear regression calibration (RC), and the fitted values from the episodic consumers (EC) model. Coverage was estimated as the proportion of simulated data sets in which the 95% confidence interval for the log OR estimate contained the true value.

The simulation described above with $H(\gamma_0) = 0.75$ was repeated with $\sigma^2_{u1} = 2, 8$ and $\sigma^2_\epsilon = 4$ and also by increasing the sample size to 5000. We investigated the effects on results of falsely assuming that the random effects u_{1i} are normally distributed by repeating the simulations using heavy tailed and skew distributions for u_{1i}. The heavy tailed distribution was created using a mixture of two bivariate normal distributions for (u_{1i}, u_{2i}), the first as in the above simulations and the second replacing $\sigma^2_{u_1}$ by 10×4.13 while ρ and $\sigma^2_{u_2}$ remained the same. The skew distribution was again created using a mixture of normals, the first as in the main simulations and the second replacing the mean of 0 for u_{1i} by a mean of 5. In both cases (u_{1i}, u_{2i}) were sampled from the first distribution with probability 0.8 and from the second with probability 0.2. Finally, we investigated the effects on results of misspecifying the value of the Box-Cox transformation parameter as $\lambda = 0.3$.

All analyses were done in `R` and the maximum likelihood estimation was performed using the `nlm` (non-linear minimization) function. main results are given Tables 2 and 3 in the paper and additional results in Tables 1-11.

C. Measurement error correction using linear regression calibration

Here we explain why linear regression calibration for measurement error correction ought to work well when the observed measurements are subject to excess zeros in the case of a *linear* model for the association between true intake T_i and a continuous outcome Y_i. Suppose that we wish to fit the model $Y_i = a + bT_i + \epsilon_i$, but instead we fit $Y_i = a^* + b^* R_{i1} + \epsilon_i$ using the observed exposure. The effect of error in R_{i1} on the association between the true and observed diet-outcome association is summarised by the regression dilution ratio (RDR)

$$\frac{b^*}{b} = \frac{\text{cov}(Y_i, R_{i1})}{\text{var}(R_{i1})} \frac{\text{var}(T_i)}{\text{cov}(Y_i, T_i)}.$$

Under the assumption that the linear model $Y_i = a + bT_i + \epsilon_i$ is *true*, $\text{cov}(Y_i, R_{i1}) = b \times \text{cov}(T_i, R_{i1})$ and the RDR is

$$\frac{b^*}{b} = \frac{\text{cov}(T_i, R_{i1})}{\text{var}(R_{i1})}.$$

It follows that the RDR can be estimated from a linear regression of T_i on R_{i1} and that this applies even when there are excess zeros in the observed measurements, because the argument involved no assumption about the association between T_i and R_{i1}. The corrected estimate of β obtained using the RDR is identical to that obtained by replacing T_i with $E(T_i|R_{i1})$ in the diet-outcome model, where $E(T_i|R_{i1})$ are the fitted values from the linear regression of T_i on R_{i1}.

In our simulation study we considered a logistic diet-disease model $\text{Pr}(Y_i = 1|T_i) = \exp(\alpha + \beta T_i)/(1 + \exp(\alpha + \beta T_i))$, where $Y_i = 1$ indicates disease and $Y_i = 0$ indicates no disease. There is no closed form expression for the estimate of β under this model (Rosner *and others*, 1989) and the explanation above for linear models therefore does not exactly extend to this situation. Our simulation studies have shown, however, that this argument appears to extend approximately to logistic diet-disease associations.

D. Simulation study: FFQ-adjusted NEC model

The FFQ-adjusted NEC model in (5.1-5.3) was fitted without covariate adjustment to EPIC-Norfolk data on alcohol intake. The analyses are based on 17,392 individuals with one or two 7-day diary measurements plus one or two FFQ measurements. We also repeated the original unadjusted analysis using this set of individuals. Non-zero 7-day diary and FFQ measurements were transformed using a Box-Cox transformation with $\lambda = 0.25$. The resulting parameter estimates are shown in Table 12.

We simulated data sets of 1000 individuals according to the model in (5.1-5.3) using the parameter estimates in Table 12. As in the earlier simulations the observed measurements R_{ij} were assumed to be Normally distributed on a Box-Cox transformed scale with $\lambda = 0.25$. Measurements Q_i were generated following the distribution of the mean FFQ measurements in the EPIC-Norfolk data. We assumed a single FFQ measurement per individual and Q_i were generated such that they were 0 with probability 0.21. Non-zero transformed measurements Q_i^* were generated from a normal distribution with mean 2.39 and standard deviation 1.93; these were the mean and standard deviation of measurements in the EPIC-Norfolk data after Box-Cox transformation with $\lambda = 0.25$. The simulation was repeated to give 500 simulated data sets, each containing true intake T_i, repeated observed measurements R_{ij} ($j = 1, \dots, 10$), and FFQ measurements Q_i. In each simulated data set the NEC model was fitted by maximum likelihood using

$J = 2, 4, 10$ repeat measurements per individual, with and without FFQ-adjustment, to give estimated parameters $\hat{\theta}$. These were used to calculate the fitted values $\hat{T}_i(\hat{\theta})$ from the FFQ-adjusted and unadjusted models in each simulated data set. As before, we generated binary outcomes according to logistic models with ORs 1.2, 1.5 and 2 associated with a 10grams/day increase in alcohol intake. Fitted values from the FFQ-adjusted and unadjusted NEC models were used in the diet-disease model to estimate ORs in each simulated data set.

E. Allowing systematic bias and correlated errors in R_{ij}

We have assumed that the observed measurements are unbiased for true intake in order to calculate the fitted values $\hat{T}_i(\theta)$. However, studies which have compared food record measurements with recovery biomarkers, which provide unbiased measures of true intake, suggest that errors in food record measurements depend on true intake and are correlated across repeated measures (Day *and others*, 2001; Kipnis *and others*, 2003). In this section we extend the NEC model to allow systematic bias in R_{ij}.

We first allow the errors ϵ_{ij} in part (2.3) of the NEC model to be correlated such that they have a multivariate normal distribution with means 0 and variance-covariance matrix Σ_ϵ with σ^2_ϵ on the diagonal and $\rho_\epsilon\sigma^2_\epsilon$ on the off-diagonals. The joint distribution of the R^*_{ij} measurements given $\mathbf{u}_i$ and $u_{0i} = 1$ is now

$$f(\mathbf{R}^*_i|\mathbf{u}_i, u_{0i} = 1; \theta) = \prod_{j=1}^{J} \{MVN\left(\mathbf{R}^*_i - (\gamma_2 + u_{2i})\, , \Sigma_\epsilon\right)\}^{I(R_{ij}>0)} \times \{H(\gamma_1 + u_{1i})\}^{I(R_{ij}>0)}\, \{1 - H(\gamma_1 + u_{1i})\}^{1-I(R_{ij}>0)}$$

where $MVN\left(\mathbf{R}^*_i - (\gamma_2 + u_{2i})\, , \Sigma_\epsilon\right)$ denotes the pdf of the multivariate normal distribution with mean vector $\mathbf{R}_i^* - (\gamma_2 + u_{2i})$ and variance-covariance matrix Σ_ϵ. The joint distribution $f(\mathbf{R}_i^*;\theta)$ follows, where θ now includes parameter ρ_ϵ. Error in R_{ij} which is correlated with true intake can be accommodated at the stage at which the fitted values $\hat{T}_i(\theta)$ are calculated. We suppose that the association between error in R_{ij} and true intake can be modelled as

$$E(R_{ij}|T_i, u_{0i}) = u_{0i}\{\alpha_R + \beta_R T_i\}$$

and write

$$T_i = u_{0i}\left\{\frac{g^*(u_{2i};\theta,\lambda)H(\gamma_1 + u_{2i}) - \alpha_R}{\beta_R}\right\}.$$

The fitted values are

$$\hat{T}_i(\theta) = \frac{1}{\beta_R}\left\{\frac{H(\gamma_0)\iint H(\gamma_1 + u_{1i})g^*(u_{2i};\theta,\lambda)f(\mathbf{R}_i^*|\mathbf{u}_i, u_{0i} = 1;\theta)f(u_{1i}, u_{2i};\theta)\mathrm{d}u_{1i}\mathrm{d}u_{2i}}{f(\mathbf{R}_i^*;\theta)} - \alpha_R\right\}.$$

Values $\alpha_R = 0$ and $\beta_R = 1$ correspond to the case where there is no intake-related bias in the observed measurements R_{ij}. The parameters α_R, β_R, and ρ_ϵ cannot be estimated without comparisons of food record measurements with unbiased measures of food intake. However there are currently no such measurements for any foods and we suggest that sensitivity analyses could be performed to assess the effects of correlated errors and intake-related error. In a small number of studies the structure of error in self-reported measures of nutrient intake, including systematic error in food record measurements, has been estimated using recovery biomarkers, which are available only for intakes of total energy, protein, potassium and sodium (Day *and others*, 2001; Kipnis *and others*, 2003). Suitable values for β_R could be chosen from these, though it is not clear that the structure of systematic error in food record measurements should persist in the same way across different foods and nutrients, while values for α_R could depend very much on the food in question, and the degree to which α_R affects measurement error correction is an area for further work.

REFERENCES

DAY, N.E., MCKEOWN, N., WONG, M.Y., WELCH, A. AND BINGHAM, S. (2001). Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *International Journal of Epidemiology* **30**, 309–317.

KIPNIS, V., SUBAR, A.F., MIDTHUNE, D., FREEDMAN, L.S., BALLARD-BARBASH, R., TROIANO, R.P., BINGHAM, S., SCHOELLER, D.A., SCHATZKIN, A. AND CARROLL, R.J. (2003). Structure of dietary measurement error: Results of the OPEN biomarker study. *American Journal of Epidemiology* **158**, 14–21.

ROSNER, B., WILLETT, W.C. AND SPIEGELMAN, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine* **8**, 1051–1069.

Table 1. *Sensitivity of parameter estimation to changes in sample size and values of $\sigma^2_{u_1}$ and σ^2_{ϵ}: Mean (empirical standard deviation) of maximum likelihood estimates of parameters from the NEC model across 500 simulated data sets using $J = 2, 4, 10$ repeat measurements, when there are 25% never consumers*

		Complete repeats			Incomplete repeats		
Parameter	True value	$J = 2$	$J = 4$	$J = 10$	$J = 2$	$J = 4$	$J = 10$
Increase sample size from 1000 to 5000							
γ_1	2.13	2.04 (0.18)	2.15 (0.05)	2.14 (0.04)	1.94 (0.38)	2.16 (0.11)	2.15 (0.09)
γ_2	2.67	2.53 (0.14)	2.67 (0.05)	2.68 (0.04)	2.48 (0.23)	2.68 (0.08)	2.69 (0.06)
$\sigma^2_{u_1}$	4.13	7.10 (2.82)	4.31 (0.37)	4.15 (0.19)	8.90 (5.20)	4.53 (0.88)	4.13 (0.42)
$\sigma^2_{u_2}$	4.45	4.70 (0.30)	4.45 (0.14)	4.45 (0.13)	4.73 (0.43)	4.43 (0.21)	4.40 (0.17)
ρ	0.91	0.87 (0.02)	0.90 (0.01)	0.90 (0.01)	0.85 (0.03)	0.88 (0.01)	0.89 (0.01)
σ^2_{ϵ}	1.17	1.17 (0.03)	1.17 (0.02)	1.17 (0.01)	1.17 (0.09)	1.17 (0.05)	1.17 (0.03)
$H(\gamma_0)$	0.75	0.80 (0.05)	0.75 (0.01)	0.75 (0.01)	0.82 (0.08)	0.75 (0.02)	0.75 (0.01)
Change $\sigma^2_{u_1}$ to 8							
γ_1	2.13	1.80 (0.54)	2.13 (0.18)	2.14 (0.13)	1.82 (0.85)	2.11 (0.48)	2.15 (0.29)
γ_2	2.67	2.49 (0.29)	2.67 (0.12)	2.68 (0.09)	2.51 (0.36)	2.68 (0.24)	2.69 (0.16)
$\sigma^2_{u_1}$	8	12.98 (6.98)	8.28 (1.71)	8.04 (0.89)	15.53 (11.03)	9.60 (5.72)	8.27 (2.63)
$\sigma^2_{u_2}$	4.45	4.71 (0.60)	4.43 (0.35)	4.44 (0.29)	4.63 (0.71)	4.38 (0.56)	4.37 (0.46)
ρ	0.91	0.89 (0.02)	0.90 (0.01)	0.90 (0.01)	0.87 (0.04)	0.88 (0.03)	0.90 (0.02)
σ^2_{ϵ}	1.17	1.17 (0.08)	1.17 (0.04)	1.17 (0.02)	1.16 (0.18)	1.17 (0.11)	1.17 (0.06)
$H(\gamma_0)$	0.75	0.82 (0.10)	0.75 (0.03)	0.75 (0.02)	0.83 (0.13)	0.76 (0.07)	0.75 (0.04)
Change $\sigma^2_{u_1}$ to 2							
γ_1	2.13	2.00 (0.32)	2.15 (0.10)	2.13 (0.07)	1.97 (0.46)	2.19 (0.23)	2.14 (0.14)
γ_2	2.67	2.47 (0.26)	2.67 (0.08)	2.67 (0.08)	2.42 (0.32)	2.66 (0.13)	2.67 (0.10)
$\sigma^2_{u_1}$	2	6.32 (5.91)	2.14 (0.40)	2.03 (0.20)	8.81 (9.29)	2.60 (1.71)	2.08 (0.50)
$\sigma^2_{u_2}$	4.45	4.78 (0.55)	4.46 (0.27)	4.46 (0.25)	4.87 (0.71)	4.47 (0.40)	4.43 (0.35)
ρ	0.91	0.83 (0.05)	0.89 (0.02)	0.90 (0.01)	0.81 (0.08)	0.86 (0.05)	0.89 (0.06)
σ^2_{ϵ}	1.17	1.17 (0.07)	1.17 (0.04)	1.17 (0.02)	1.17 (0.19)	1.17 (0.10)	1.17 (0.06)
$H(\gamma_0)$	0.75	0.82 (0.09)	0.75 (0.02)	0.75 (0.01)	0.84 (0.11)	0.76 (0.03)	0.75 (0.02)
Change σ^2_{ϵ} to 4							
γ_1	2.13	1.74 (0.46)	2.13 (0.12)	2.13 (0.09)	1.78 (0.59)	2.13 (0.29)	2.15 (0.19)
γ_2	2.67	2.36 (0.31)	2.67 (0.10)	2.67 (0.08)	2.43 (0.38)	2.67 (0.21)	2.69 (0.13)
$\sigma^2_{u_1}$	4.13	10.86 (6.60)	4.37 (0.87)	4.16 (0.42)	11.63 (9.64)	5.14 (3.44)	4.19 (1.06)
$\sigma^2_{u_2}$	4.45	4.97 (0.70)	4.45 (0.37)	4.44 (0.29)	4.90 (1.12)	4.45 (0.72)	4.37 (0.50)
ρ	0.91	0.86 (0.04)	0.89 (0.02)	0.90 (0.01)	0.84 (0.07)	0.86 (0.05)	0.89 (0.50)
σ^2_{ϵ}	4	3.99 (0.25)	3.99 (0.14)	4.00 (0.08)	3.93 (0.58)	3.97 (0.34)	4.00 (0.19)
$H(\gamma_0)$	0.75	0.86 (0.10)	0.75 (0.02)	0.75 (0.01)	0.85 (0.12)	0.76 (0.05)	0.75 (0.03)

Table 2. *Sensitivity of parameter estimation to falsely assuming a Normal distribution for the random effects* u_{1i}*: Mean (empirical standard deviation) of maximum likelihood estimates of parameters from the NEC model across 500 simulated data sets using* $J = 2, 4, 10$ *repeat measurements, when there are 25% never consumers*

		Complete repeats			Incomplete repeats		
Parameter	True value	$J = 2$	$J = 4$	$J = 10$	$J = 2$	$J = 4$	$J = 10$
True distribution for u_{1i}: Heavy tailed							
γ_1	2.13	1.99 (0.49)	2.26 (0.14)	2.26 (0.11)	1.91 (0.76)	2.20 (0.43)	2.25 (0.24)
γ_2	2.67	2.56 (0.31)	2.75 (0.11)	2.74 (0.09)	2.54 (0.39)	2.72 (0.24)	2.74 (0.14)
$\sigma^2_{u_1}$	11.56	10.27 (7.65)	5.48 (1.12)	5.43 (0.59)	13.47 (14.53)	6.76 (4.99)	5.50 (1.51)
$\sigma^2_{u_2}$	4.45	4.69 (0.61)	4.38 (0.31)	4.37 (0.28)	4.67 (0.78)	4.41 (0.54)	4.36 (0.41)
ρ	0.91	0.87 (0.03)	0.88 (0.02)	0.89 (0.01)	0.85 (0.06)	0.87 (0.04)	0.88 (0.03)
σ^2_ϵ	1.17	1.17 (0.08)	1.17 (0.04)	1.17 (0.02)	1.17 (0.19)	1.16 (0.11)	1.17 (0.06)
$H(\gamma_0)$	0.75	0.79 (0.10)	0.72 (0.02)	0.72 (0.02)	0.81 (0.13)	0.74 (0.07)	0.73 (0.03)
True distribution for u_{1i}: Skew							
γ_1	2.13	2.43 (0.48)	2.75 (0.16)	2.73 (0.13)	2.52 (0.70)	2.64 (0.48)	2.67 (0.28)
γ_2	2.67	2.40 (0.31)	2.66 (0.10)	2.65 (0.11)	2.52 (0.35)	2.60 (0.25)	2.61 (0.16)
$\sigma^2_{u_1}$	8.13	16.34 (11.88)	5.84 (1.28)	5.51 (0.65)	16.10 (38.46)	8.21 (8.39)	5.98 (3.05)
$\sigma^2_{u_2}$	4.45	4.89 (0.63)	4.47 (0.29)	4.44 (0.27)	4.69 (0.72)	4.58 (0.57)	4.57 (0.48)
ρ	0.91	0.78 (0.06)	0.79 (0.03)	0.81 (0.03)	0.77 (0.11)	0.78 (0.08)	0.82 (0.05)
σ^2_ϵ	1.17	1.17 (0.07)	1.17 (0.04)	1.17 (0.02)	1.17 (0.18)	1.17 (0.10)	1.17 (0.06)
$H(\gamma_0)$	0.75	0.85 (0.11)	0.76 (0.02)	0.75 (0.01)	0.82 (0.12)	0.78 (0.07)	0.76 (0.03)

Table 3. *Sensitivity of parameter estimation to incorrect specification of* λ*, the Box-Cox transformation parameter: Mean (empirical standard deviation) of maximum likelihood estimates of parameters from the NEC model across 500 simulated data sets using* $J = 2, 4, 10$ *repeat measurements, when there are 25% never consumers*

		Complete repeats			Incomplete repeats		
Parameter	True value	$J = 2$	$J = 4$	$J = 10$	$J = 2$	$J = 4$	$J = 10$
True value $\lambda = 0.25$, specified value $\lambda = 0.3$							
γ_1	2.13	1.37 (0.43)	2.15 (0.12)	2.16 (0.09)	1.35 (0.88)	2.10 (0.35)	2.17 (0.19)
γ_2	2.67	2.24 (0.31)	2.87 (0.11)	2.92 (0.09)	2.15 (0.27)	2.74 (0.28)	2.89 (0.14)
$\sigma^2_{u_1}$	4.13	16.74 (6.32)	4.95 (0.99)	4.42 (0.46)	21.87 (10.41)	7.74 (5.54)	4.80 (1.25)
$\sigma^2_{u_2}$	4.45	6.84 (0.80)	5.57 (0.41)	5.47 (0.34)	7.11 (0.89)	5.93 (0.80)	5.58 (0.49)
ρ	0.91	0.88 (0.02)	0.90 (0.02)	0.90 (0.02)	0.88 (0.03)	0.89 (0.03)	0.90 (0.02)
σ^2_ϵ	1.17	1.48 (0.10)	1.47 (0.05)	1.47 (0.03)	1.45 (0.23)	1.46 (0.13)	1.46 (0.07)
$H(\gamma_0)$	0.75	0.94 (0.09)	0.76 (0.02)	0.75 (0.01)	0.97 (0.07)	0.79 (0.07)	0.75 (0.03)

Table 4. *Mean (empirical standard deviation [SD]) of log OR estimates and coverage of 95% confidence intervals across 500 simulated data sets using different correction methods when there are $J = 2, 4, 10$ repeat measurements per person and 12% of individuals are never-consumers*

		Method				
True β		Using T_i	NEC model	Crude	Linear RC	EC model
Complete repeats						
$J = 2$						
0.182	Mean (SD)	0.185 (0.064)	0.187 (0.071)	0.157 (0.060)	0.184 (0.070)	0.186 (0.071)
	Coverage	0.96	0.96	0.95	0.95	0.96
0.405	Mean (SD)	0.409 (0.061)	0.411 (0.067)	0.346 (0.056)	0.405 (0.067)	0.409 (0.066)
	Coverage	0.94	0.94	0.77	0.93	0.93
0.693	Mean (SD)	0.695 (0.062)	0.675 (0.067)	0.577 (0.058)	0.676 (0.068)	0.672 (0.067)
	Coverage	0.97	0.92	0.48	0.92	0.91
$J = 4$						
0.182	Mean (SD)	0.185 (0.064)	0.187 (0.067)	0.170 (0.061)	0.184 (0.066)	0.184 (0.066)
	Coverage	0.96	0.96	0.95	0.96	0.96
0.405	Mean (SD)	0.409 (0.061)	0.411 (0.062)	0.375 (0.057)	0.406 (0.062)	0.405 (0.062)
	Coverage	0.94	0.94	0.90	0.94	0.93
0.693	Mean (SD)	0.695 (0.062)	0.686 (0.065)	0.631 (0.060)	0.685 (0.065)	0.678 (0.064)
	Coverage	0.97	0.95	0.79	0.95	0.94
$J = 10$						
0.182	Mean (SD)	0.185 (0.064)	0.186 (0.064)	0.179 (0.062)	0.185 (0.064)	0.184 (0.064)
	Coverage	0.96	0.96	0.96	0.96	0.96
0.405	Mean (SD)	0.409 (0.061)	0.409 (0.062)	0.394 (0.060)	0.408 (0.062)	0.405 (0.061)
	Coverage	0.94	0.94	0.92	0.94	0.93
0.693	Mean (SD)	0.695 (0.062)	0.692 (0.063)	0.669 (0.061)	0.691 (0.064)	0.685 (0.063)
	Coverage	0.97	0.97	0.92	0.96	0.96
Incomplete repeats						
$J = 2$						
0.182	Mean (SD)	0.185 (0.064)	0.191 (0.078)	0.139 (0.056)	0.201 (0.099)	0.190 (0.078)
	Coverage	0.96	0.96	0.92	0.92	0.96
0.405	Mean (SD)	0.409 (0.061)	0.414 (0.073)	0.305 (0.052)	0.438 (0.140)	0.412 (0.073)
	Coverage	0.94	0.92	0.44	0.69	0.92
0.693	Mean (SD)	0.695 (0.062)	0.666 (0.078)	0.505 (0.056)	0.722 (0.212)	0.664 (0.078)
	Coverage	0.97	0.87	0.08	0.54	0.86
$J = 4$						
0.182	Mean (SD)	0.185 (0.064)	0.191 (0.077)	0.141 (0.056)	0.198 (0.094)	0.187 (0.075)
	Coverage	0.96	0.96	0.92	0.90	0.96
0.405	Mean (SD)	0.409 (0.061)	0.415 (0.071)	0.308 (0.052)	0.432 (0.130)	0.406 (0.069)
	Coverage	0.94	0.93	0.47	0.71	0.92
0.693	Mean (SD)	0.695 (0.062)	0.670 (0.073)	0.511 (0.057)	0.713 (0.193)	0.656 (0.071)
	Coverage	0.97	0.90	0.10	0.56	0.87
$J = 10$						
0.182	Mean (SD)	0.185 (0.064)	0.191 (0.076)	0.141 (0.056)	0.196 (0.090)	0.186 (0.074)
	Coverage	0.96	0.96	0.93	0.92	0.96
0.405	Mean (SD)	0.409 (0.061)	0.415 (0.069)	0.310 (0.052)	0.429 (0.125)	0.405 (0.068)
	Coverage	0.94	0.93	0.49	0.71	0.92
0.693	Mean (SD)	0.695 (0.062)	0.671 (0.070)	0.515 (0.057)	0.708 (0.182)	0.657 (0.070)
	Coverage	0.97	0.93	0.12	0.59	0.87

Table 5. *Change sample size to 5000: Mean (empirical standard deviation [SD]) of log OR estimates and coverage of 95% confidence intervals across 500 simulated data sets using different correction methods when there are* $J = 2, 4, 10$ *repeat measurements per person and 25% of individuals are never-consumers*

		Method				
True β		Using T_i	NEC model	Crude	Linear RC	EC model
Complete repeats						
$J = 2$						
0.182	Mean (SD)	0.183 (0.030)	0.186 (0.033)	0.157 (0.027)	0.182 (0.032)	0.183 (0.032)
	Coverage	0.94	0.94	0.88	0.95	0.95
0.405	Mean (SD)	0.406 (0.026)	0.408 (0.030)	0.347 (0.026)	0.402 (0.030)	0.404 (0.029)
	Coverage	0.95	0.93	0.32	0.93	0.93
0.693	Mean (SD)	0.698 (0.029)	0.678 (0.031)	0.585 (0.028)	0.677 (0.033)	0.671 (0.031)
	Coverage	0.96	0.91	0.02	0.90	0.87
$J = 4$						
0.182	Mean (SD)	0.183 (0.030)	0.185 (0.031)	0.169 (0.029)	0.182 (0.031)	0.181 (0.031)
	Coverage	0.94	0.94	0.94	0.95	0.95
0.405	Mean (SD)	0.406 (0.026)	0.408 (0.028)	0.374 (0.026)	0.403 (0.028)	0.400 (0.027)
	Coverage	0.95	0.94	0.76	0.94	0.94
0.693	Mean (SD)	0.698 (0.029)	0.688 (0.031)	0.636 (0.029)	0.686 (0.031)	0.676 (0.030)
	Coverage	0.96	0.94	0.48	0.93	0.90
$J = 10$						
0.182	Mean (SD)	0.183 (0.030)	0.184 (0.027)	0.177 (0.029)	0.182 (0.030)	0.181 (0.030)
	Coverage	0.94	0.95	0.96	0.95	0.95
0.405	Mean (SD)	0.406 (0.026)	0.407 (0.027)	0.393 (0.026)	0.405 (0.027)	0.401 (0.026)
	Coverage	0.95	0.94	0.92	0.94	0.94
0.693	Mean (SD)	0.698 (0.029)	0.693 (0.030)	0.672 (0.029)	0.693 (0.030)	0.685 (0.030)
	Coverage	0.96	0.96	0.86	0.96	0.95
Incomplete repeats						
$J = 2$						
0.182	Mean (SD)	0.183 (0.030)	0.189 (0.035)	0.140 (0.026)	0.186 (0.044)	0.187 (0.035)
	Coverage	0.94	0.93	0.67	0.86	0.94
0.405	Mean (SD)	0.406 (0.026)	0.411 (0.033)	0.309 (0.024)	0.406 (0.063)	0.407 (0.032)
	Coverage	0.95	0.93	0.02	0.64	0.94
0.693	Mean (SD)	0.698 (0.029)	0.670 (0.035)	0.518 (0.027)	0.680 (0.099)	0.665 (0.035)
	Coverage	0.96	0.84	0	0.49	0.81
$J = 4$						
0.182	Mean (SD)	0.183 (0.030)	0.189 (0.035)	0.142 (0.026)	0.184 (0.042)	0.183 (0.034)
	Coverage	0.94	0.93	0.69	0.88	0.95
0.405	Mean (SD)	0.406 (0.026)	0.413 (0.031)	0.312 (0.024)	0.403 (0.058)	0.400 (0.030)
	Coverage	0.95	0.94	0.02	0.67	0.93
0.693	Mean (SD)	0.698 (0.029)	0.674 (0.033)	0.523 (0.027)	0.676 (0.090)	0.656 (0.033)
	Coverage	0.96	0.87	0	0.50	0.77
$J = 10$						
0.182	Mean (SD)	0.183 (0.030)	0.189 (0.035)	0.142 (0.026)	0.184 (0.041)	0.180 (0.033)
	Coverage	0.94	0.93	0.70	0.89	0.95
0.405	Mean (SD)	0.406 (0.026)	0.413 (0.031)	0.314 (0.025)	0.403 (0.055)	0.394 (0.030)
	Coverage	0.95	0.94	0.03	0.70	0.92
0.693	Mean (SD)	0.698 (0.029)	0.675 (0.032)	0.527 (0.027)	0.677 (0.084)	0.649 (0.032)
	Coverage	0.96	0.88	0	0.53	0.66

Table 6. *Change $\sigma^2_{u_1}$ to 8: Mean (empirical standard deviation [SD]) of log OR estimates and coverage of 95% confidence intervals across 500 simulated data sets using different correction methods when there are $J = 2, 4, 10$ repeat measurements per person and 25% of individuals are never-consumers*

		Method				
True β		Using T_i	NEC model	Crude	Linear RC	EC model
Complete repeats						
$J = 2$						
0.182	Mean (SD)	0.175 (0.078)	0.177 (0.085)	0.150 (0.072)	0.173 (0.083)	0.175 (0.084)
	Coverage	0.93	0.95	0.94	0.95	0.95
0.405	Mean (SD)	0.409 (0.061)	0.412 (0.068)	0.351 (0.058)	0.405 (0.068)	0.408 (0.067)
	Coverage	0.95	0.95	0.79	0.94	0.95
0.693	Mean (SD)	0.701 (0.068)	0.685 (0.073)	0.594 (0.064)	0.685 (0.075)	0.680 (0.072)
	Coverage	0.94	0.93	0.60	0.93	0.92
$J = 4$						
0.182	Mean (SD)	0.175 (0.078)	0.177 (0.082)	0.162 (0.075)	0.174 (0.080)	0.174 (0.080)
	Coverage	0.93	0.94	0.95	0.94	0.94
0.405	Mean (SD)	0.409 (0.061)	0.412 (0.064)	0.378 (0.059)	0.407 (0.064)	0.406 (0.063)
	Coverage	0.95	0.94	0.91	0.95	0.94
0.693	Mean (SD)	0.701 (0.068)	0.693 (0.070)	0.643 (0.066)	0.693 (0.071)	0.685 (0.070)
	Coverage	0.94	0.94	0.85	0.93	0.94
$J = 10$						
0.182	Mean (SD)	0.175 (0.078)	0.175 (0.079)	0.169 (0.076)	0.175 (0.079)	0.174 (0.079)
	Coverage	0.93	0.93	0.94	0.93	0.93
0.405	Mean (SD)	0.409 (0.061)	0.409 (0.062)	0.395 (0.060)	0.407 (0.062)	0.405 (0.061)
	Coverage	0.95	0.94	0.95	0.94	0.95
0.693	Mean (SD)	0.701 (0.068)	0.697 (0.068)	0.675 (0.067)	0.696 (0.069)	0.691 (0.068)
	Coverage	0.94	0.95	0.94	0.95	0.94
Incomplete repeats						
$J = 2$						
0.182	Mean (SD)	0.175 (0.078)	0.179 (0.091)	0.134 (0.069)	0.184 (0.113)	0.178 (0.090)
	Coverage	0.93	0.95	0.92	0.88	0.95
0.405	Mean (SD)	0.409 (0.061)	0.417 (0.075)	0.312 (0.055)	0.428 (0.143)	0.414 (0.074)
	Coverage	0.95	0.93	0.53	0.69	0.93
0.693	Mean (SD)	0.701 (0.068)	0.681 (0.085)	0.526 (0.062)	0.715 (0.209)	0.678 (0.085)
	Coverage	0.94	0.90	0.20	0.53	0.90
$J = 4$						
0.182	Mean (SD)	0.175 (0.078)	0.179 (0.091)	0.135 (0.069)	0.184 (0.110)	0.176 (0.089)
	Coverage	0.93	0.95	0.93	0.89	0.95
0.405	Mean (SD)	0.409 (0.061)	0.419 (0.071)	0.316 (0.055)	0.428 (0.139)	0.409 (0.070)
	Coverage	0.95	0.94	0.56	0.69	0.93
0.693	Mean (SD)	0.701 (0.068)	0.685 (0.078)	0.533 (0.062)	0.715 (0.200)	0.672 (0.077)
	Coverage	0.94	0.92	0.24	0.51	0.90
$J = 10$						
0.182	Mean (SD)	0.175 (0.078)	0.179 (0.090)	0.136 (0.069)	0.182 (0.107)	0.173 (0.087)
	Coverage	0.93	0.95	0.93	0.89	0.94
0.405	Mean (SD)	0.409 (0.061)	0.418 (0.069)	0.318 (0.055)	0.423 (0.134)	0.404 (0.067)
	Coverage	0.95	0.94	0.58	0.71	0.94
0.693	Mean (SD)	0.701 (0.068)	0.684 (0.076)	0.536 (0.062)	0.709 (0.193)	0.666 (0.075)
	Coverage	0.94	0.93	0.25	0.54	0.91

Table 7. *Change $\sigma^2_{u_1}$ to 2: Mean (empirical standard deviation [SD]) of log OR estimates and coverage of 95% confidence intervals across 500 simulated data sets using different correction methods when there are $J = 2, 4, 10$ repeat measurements per person and 25% of individuals are never-consumers*

		Method				
True β		Using T_i	NEC model	Crude	Linear RC	EC model
Complete repeats						
$J = 2$						
0.182	Mean (SD)	0.181 (0.070)	0.183 (0.078)	0.154 (0.066)	0.179 (0.077)	0.181 (0.077)
	Coverage	0.95	0.96	0.95	0.96	0.95
0.405	Mean (SD)	0.408 (0.064)	0.410 (0.070)	0.346 (0.059)	0.403 (0.070)	0.405 (0.069)
	Coverage	0.94	0.92	0.77	0.93	0.93
0.693	Mean (SD)	0.696 (0.066)	0.676 (0.069)	0.580 (0.060)	0.676 (0.071)	0.669 (0.068)
	Coverage	0.96	0.94	0.53	0.93	0.92
$J = 4$						
0.182	Mean (SD)	0.181 (0.070)	0.182 (0.075)	0.166 (0.068)	0.180 (0.074)	0.178 (0.073)
	Coverage	0.95	0.95	0.96	0.96	0.96
0.405	Mean (SD)	0.408 (0.064)	0.410 (0.066)	0.374 (0.060)	0.405 (0.066)	0.400 (0.064)
	Coverage	0.94	0.94	0.90	0.94	0.94
0.693	Mean (SD)	0.696 (0.066)	0.687 (0.066)	0.633 (0.061)	0.685 (0.067)	0.672 (0.064)
	Coverage	0.96	0.95	0.82	0.95	0.94
$J = 10$						
0.182	Mean (SD)	0.181 (0.070)	0.182 (0.071)	0.175 (0.069)	0.181 (0.071)	0.178 (0.070)
	Coverage	0.95	0.96	0.95	0.96	0.95
0.405	Mean (SD)	0.408 (0.064)	0.409 (0.065)	0.394 (0.063)	0.407 (0.065)	0.400 (0.063)
	Coverage	0.94	0.94	0.94	0.94	0.95
0.693	Mean (SD)	0.696 (0.066)	0.692 (0.066)	0.669 (0.064)	0.691 (0.066)	0.678 (0.064)
	Coverage	0.96	0.95	0.93	0.95	0.94
Incomplete repeats						
$J = 2$						
0.182	Mean (SD)	0.181 (0.070)	0.185 (0.084)	0.137 (0.062)	0.197 (0.110)	0.182 (0.083)
	Coverage	0.95	0.97	0.93	0.91	0.97
0.405	Mean (SD)	0.408 (0.064)	0.410 (0.076)	0.306 (0.054)	0.440 (0.150)	0.405 (0.074)
	Coverage	0.94	0.93	0.49	0.72	0.93
0.693	Mean (SD)	0.696 (0.066)	0.665 (0.077)	0.510 (0.057)	0.730 (0.228)	0.658 (0.076)
	Coverage	0.96	0.89	0.12	0.53	0.87
$J = 4$						
0.182	Mean (SD)	0.181 (0.070)	0.186 (0.084)	0.138 (0.062)	0.195 (0.104)	0.179 (0.081)
	Coverage	0.95	0.95	0.93	0.91	0.96
0.405	Mean (SD)	0.408 (0.064)	0.413 (0.073)	0.308 (0.055)	0.434 (0.136)	0.398 (0.071)
	Coverage	0.94	0.94	0.51	0.74	0.94
0.693	Mean (SD)	0.696 (0.066)	0.670 (0.071)	0.515 (0.057)	0.721 (0.207)	0.648 (0.069)
	Coverage	0.96	0.93	0.14	0.55	0.87
$J = 10$						
0.182	Mean (SD)	0.181 (0.070)	0.186 (0.083)	0.139 (0.062)	0.192 (0.100)	0.174 (0.077)
	Coverage	0.95	0.96	0.94	0.92	0.97
0.405	Mean (SD)	0.408 (0.064)	0.414 (0.072)	0.310 (0.056)	0.431 (0.131)	0.388 (0.068)
	Coverage	0.94	0.93	0.53	0.73	0.93
0.693	Mean (SD)	0.696 (0.066)	0.672 (0.069)	0.519 (0.057)	0.715 (0.194)	0.634 (0.066)
	Coverage	0.96	0.94	0.16	0.59	0.83

Table 8. *Change σ_ϵ^2 to 4: Mean (empirical standard deviation [SD]) of log OR estimates and coverage of 95% confidence intervals across 500 simulated data sets using different correction methods when there are $J = 2, 4, 10$ repeat measurements per person and 25% of individuals are never-consumers*

		Method				
True β		Using T_i	NEC model	Crude	Linear RC	EC model
Complete repeats						
$J = 2$						
0.182	Mean (SD)	0.182 (0.058)	0.187 (0.074)	0.115 (0.047)	0.175 (0.073)	0.184 (0.073)
	Coverage	0.96	0.96	0.73	0.97	0.97
0.405	Mean (SD)	0.407 (0.057)	0.412 (0.073)	0.253 (0.046)	0.389 (0.076)	0.405 (0.072)
	Coverage	0.94	0.90	0.08	0.87	0.90
0.693	Mean (SD)	0.700 (0.060)	0.651 (0.072)	0.422 (0.049)	0.646 (0.082)	0.642 (0.070)
	Coverage	0.96	0.86	0	0.79	0.83
$J = 4$						
0.182	Mean (SD)	0.182 (0.058)	0.187 (0.067)	0.141 (0.051)	0.178 (0.065)	0.180 (0.064)
	Coverage	0.96	0.96	0.93	0.97	0.97
0.405	Mean (SD)	0.407 (0.057)	0.412 (0.063)	0.312 (0.049)	0.395 (0.063)	0.395 (0.061)
	Coverage	0.94	0.93	0.44	0.92	0.93
0.693	Mean (SD)	0.700 (0.060)	0.671 (0.064)	0.524 (0.053)	0.663 (0.069)	0.649 (0.063)
	Coverage	0.96	0.93	0.12	0.88	0.87
$J = 10$						
0.182	Mean (SD)	0.182 (0.058)	0.184 (0.062)	0.163 (0.055)	0.181 (0.061)	0.178 (0.059)
	Coverage	0.96	0.95	0.97	0.96	0.95
0.405	Mean (SD)	0.407 (0.057)	0.408 (0.059)	0.362 (0.053)	0.401 (0.059)	0.395 (0.057)
	Coverage	0.94	0.94	0.82	0.93	0.92
0.693	Mean (SD)	0.700 (0.060)	0.684 (0.062)	0.616 (0.058)	0.681 (0.064)	0.667 (0.061)
	Coverage	0.96	0.94	0.68	0.92	0.91
Incomplete repeats						
$J = 2$						
0.182	Mean (SD)	0.182 (0.058)	0.193 (0.090)	0.087 (0.041)	0.199 (0.123)	0.191 (0.089)
	Coverage	0.96	0.96	0.35	0.90	0.96
0.405	Mean (SD)	0.407 (0.057)	0.424 (0.099)	0.193 (0.039)	0.439 (0.178)	0.420 (0.099)
	Coverage	0.94	0.85	0	0.66	0.84
0.693	Mean (SD)	0.700 (0.060)	0.654 (0.115)	0.319 (0.045)	0.720 (0.277)	0.650 (0.115)
	Coverage	0.96	0.70	0	0.49	0.68
$J = 4$						
0.182	Mean (SD)	0.182 (0.058)	0.195 (0.088)	0.089 (0.041)	0.189 (0.104)	0.185 (0.083)
	Coverage	0.96	0.95	0.38	0.92	0.95
0.405	Mean (SD)	0.407 (0.057)	0.428 (0.085)	0.196 (0.041)	0.416 (0.138)	0.405 (0.081)
	Coverage	0.94	0.89	0	0.69	0.89
0.693	Mean (SD)	0.700 (0.060)	0.662 (0.091)	0.327 (0.046)	0.685 (0.211)	0.632 (0.087)
	Coverage	0.96	0.80	0	0.56	0.74
$J = 10$						
0.182	Mean (SD)	0.182 (0.058)	0.194 (0.084)	0.090 (0.042)	0.183 (0.094)	0.179 (0.077)
	Coverage	0.96	0.95	0.39	0.93	0.96
0.405	Mean (SD)	0.407 (0.057)	0.428 (0.079)	0.199 (0.041)	0.407 (0.121)	0.392 (0.073)
	Coverage	0.94	0.91	0	0.73	0.90
0.693	Mean (SD)	0.700 (0.060)	0.666 (0.077)	0.332 (0.047)	0.670 (0.175)	0.614 (0.072)
	Coverage	0.96	0.88	0	0.63	0.70

Table 9. *Results when the u_{1i} have a non-normal symmetric distribution with heavy tails: Mean (empirical standard deviation [SD]) of log OR estimates and coverage of 95% confidence intervals across 500 simulated data sets using different correction methods when there are $J = 2, 4, 10$ repeat measurements per person and 25% of individuals are never-consumers*

		Method				
True β		Using T_i	NEC model	Crude	Linear RC	EC model
Complete repeats						
$J = 2$						
0.182	Mean (SD)	0.176 (0.071)	0.180 (0.080)	0.153 (0.067)	0.177 (0.078)	0.178 (0.079)
	Coverage	0.96	0.95	0.96	0.96	0.95
0.405	Mean (SD)	0.412 (0.061)	0.414 (0.066)	0.353 (0.057)	0.408 (0.066)	0.409 (0.065)
	Coverage	0.94	0.95	0.83	0.94	0.95
0.693	Mean (SD)	0.702 (0.066)	0.684 (0.069)	0.593 (0.061)	0.685 (0.071)	0.677 (0.068)
	Coverage	0.95	0.94	0.59	0.94	0.93
$J = 4$						
0.182	Mean (SD)	0.176 (0.071)	0.180 (0.075)	0.165 (0.069)	0.178 (0.074)	0.177 (0.074)
	Coverage	0.96	0.95	0.97	0.96	0.96
0.405	Mean (SD)	0.412 (0.061)	0.414 (0.063)	0.380 (0.058)	0.410 (0.062)	0.407 (0.062)
	Coverage	0.94	0.95	0.92	0.94	0.94
0.693	Mean (SD)	0.702 (0.066)	0.692 (0.067)	0.642 (0.063)	0.692 (0.068)	0.682 (0.066)
	Coverage	0.95	0.96	0.84	0.96	0.96
$J = 10$						
0.182	Mean (SD)	0.176 (0.071)	0.178 (0.072)	0.172 (0.070)	0.177 (0.072)	0.176 (0.071)
	Coverage	0.96	0.96	0.96	0.96	0.96
0.405	Mean (SD)	0.412 (0.061)	0.413 (0.062)	0.399 (0.060)	0.411 (0.062)	0.408 (0.061)
	Coverage	0.94	0.95	0.94	0.95	0.95
0.693	Mean (SD)	0.702 (0.066)	0.698 (0.066)	0.676 (0.064)	0.697 (0.066)	0.690 (0.065)
	Coverage	0.95	0.96	0.94	0.96	0.96
Incomplete repeats						
$J = 2$						
0.182	Mean (SD)	0.176 (0.071)	0.184 (0.086)	0.137 (0.063)	0.188 (0.105)	0.183 (0.086)
	Coverage	0.96	0.95	0.92	0.87	0.95
0.405	Mean (SD)	0.412 (0.061)	0.420 (0.072)	0.315 (0.053)	0.433 (0.150)	0.416 (0.072)
	Coverage	0.94	0.93	0.57	0.67	0.93
0.693	Mean (SD)	0.702 (0.066)	0.678 (0.076)	0.524 (0.059)	0.717 (0.221)	0.675 (0.076)
	Coverage	0.95	0.91	0.16	0.48	0.90
$J = 4$						
0.182	Mean (SD)	0.176 (0.071)	0.184 (0.085)	0.138 (0.063)	0.186 (0.102)	0.179 (0.082)
	Coverage	0.96	0.95	0.93	0.89	0.96
0.405	Mean (SD)	0.412 (0.061)	0.419 (0.069)	0.318 (0.054)	0.427 (0.139)	0.407 (0.067)
	Coverage	0.94	0.95	0.60	0.71	0.95
0.693	Mean (SD)	0.702 (0.066)	0.678 (0.072)	0.530 (0.059)	0.707 (0.207)	0.663 (0.071)
	Coverage	0.95	0.93	0.19	0.51	0.89
$J = 10$						
0.182	Mean (SD)	0.176 (0.071)	0.184 (0.084)	0.139 (0.064)	0.185 (0.099)	0.176 (0.080)
	Coverage	0.96	0.95	0.93	0.90	0.95
0.405	Mean (SD)	0.412 (0.061)	0.419 (0.068)	0.320 (0.054)	0.424 (0.136)	0.401 (0.065)
	Coverage	0.94	0.95	0.63	0.70	0.96
0.693	Mean (SD)	0.702 (0.066)	0.679 (0.071)	0.534 (0.059)	0.703 (0.196)	0.654 (0.070)
	Coverage	0.95	0.92	0.21	0.54	0.88

Table 10. *Results when the u_{1i} have a skew distribution: Mean (empirical standard deviation [SD]) of log OR estimates and coverage of 95% confidence intervals across 500 simulated data sets using different correction methods when there are $J = 2, 4, 10$ repeat measurements per person and 25% of individuals are never-consumers*

		Method				
True β		Using T_i	NEC model	Crude	Linear RC	EC model
Complete repeats						
$J = 2$						
0.182	Mean (SD)	0.177 (0.068)	0.179 (0.077)	0.153 (0.065)	0.176 (0.076)	0.178 (0.076)
	Coverage	0.97	0.96	0.96	0.96	0.96
0.405	Mean (SD)	0.409 (0.062)	0.411 (0.066)	0.351 (0.057)	0.405 (0.067)	0.408 (0.066)
	Coverage	0.94	0.94	0.81	0.94	0.94
0.693	Mean (SD)	0.704 (0.067)	0.685 (0.068)	0.593 (0.060)	0.686 (0.071)	0.680 (0.067)
	Coverage	0.95	0.95	0.60	0.95	0.94
$J = 4$						
0.182	Mean (SD)	0.177 (0.068)	0.180 (0.073)	0.165 (0.067)	0.177 (0.072)	0.177 (0.072)
	Coverage	0.97	0.96	0.97	0.96	0.96
0.405	Mean (SD)	0.409 (0.062)	0.411 (0.064)	0.378 (0.059)	0.407 (0.063)	0.406 (0.063)
	Coverage	0.94	0.94	0.91	0.95	0.94
0.693	Mean (SD)	0.704 (0.067)	0.694 (0.067)	0.643 (0.063)	0.693 (0.068)	0.686 (0.066)
	Coverage	0.95	0.96	0.87	0.96	0.96
$J = 10$						
0.182	Mean (SD)	0.177 (0.068)	0.179 (0.069)	0.172 (0.067)	0.178 (0.069)	0.177 (0.068)
	Coverage	0.97	0.96	0.97	0.97	0.97
0.405	Mean (SD)	0.409 (0.062)	0.410 (0.063)	0.396 (0.061)	0.409 (0.063)	0.405 (0.063)
	Coverage	0.94	0.94	0.95	0.94	0.94
0.693	Mean (SD)	0.704 (0.067)	0.700 (0.066)	0.678 (0.064)	0.699 (0.066)	0.692 (0.065)
	Coverage	0.95	0.96	0.96	0.96	0.96
Incomplete repeats						
$J = 2$						
0.182	Mean (SD)	0.177 (0.068)	0.185 (0.084)	0.137 (0.062)	0.192 (0.109)	0.183 (0.083)
	Coverage	0.97	0.95	0.94	0.90	0.96
0.405	Mean (SD)	0.409 (0.062)	0.418 (0.074)	0.312 (0.053)	0.433 (0.149)	0.414 (0.073)
	Coverage	0.94	0.93	0.56	0.67	0.93
0.693	Mean (SD)	0.704 (0.067)	0.682 (0.078)	0.525 (0.059)	0.724 (0.223)	0.677 (0.077)
	Coverage	0.95	0.91	0.17	0.51	0.91
$J = 4$						
0.182	Mean (SD)	0.177 (0.068)	0.184 (0.083)	0.138 (0.062)	0.189 (0.106)	0.180 (0.081)
	Coverage	0.97	0.95	0.95	0.90	0.96
0.405	Mean (SD)	0.409 (0.062)	0.415 (0.070)	0.315 (0.053)	0.425 (0.137)	0.406 (0.069)
	Coverage	0.94	0.95	0.57	0.68	0.94
0.693	Mean (SD)	0.704 (0.067)	0.679 (0.074)	0.530 (0.060)	0.711 (0.207)	0.668 (0.073)
	Coverage	0.95	0.92	0.19	0.51	0.90
$J = 10$						
0.182	Mean (SD)	0.177 (0.068)	0.183 (0.081)	0.138 (0.062)	0.188 (0.101)	0.176 (0.078)
	Coverage	0.97	0.96	0.95	0.92	0.96
0.405	Mean (SD)	0.409 (0.062)	0.414 (0.068)	0.317 (0.053)	0.422 (0.133)	0.397 (0.065)
	Coverage	0.94	0.95	0.59	0.71	0.94
0.693	Mean (SD)	0.704 (0.067)	0.679 (0.072)	0.534 (0.060)	0.707 (0.198)	0.655 (0.071)
	Coverage	0.95	0.94	0.22	0.53	0.88

Table 11. *Results when we incorrectly specify the Box-Cox transformation parameter* λ *(true value* $\lambda = 0.25$*, specified value* $\lambda = 0.3$*): Mean (empirical standard deviation [SD]) of log OR estimates and coverage of 95% confidence intervals across 500 simulated data sets using different correction methods when there are* $J = 2, 4, 10$ *repeat measurements per person and 25% of individuals are never-consumers*

		Method				
True β		Using T_i	NEC model	Crude	Linear RC	EC model
Complete repeats						
$J = 2$						
0.181	Mean (SD)	0.181 (0.070)	0.182 (0.076)	0.155 (0.065)	0.179 (0.075)	0.182 (0.076)
	Coverage	0.95	0.96	0.95	0.96	0.96
0.405	Mean (SD)	0.409 (0.065)	0.410 (0.071)	0.349 (0.060)	0.404 (0.071)	0.408 (0.070)
	Coverage	0.93	0.93	0.78	0.92	0.93
0.693	Mean (SD)	0.695 (0.065)	0.677 (0.069)	0.585 (0.060)	0.677 (0.070)	0.676 (0.068)
	Coverage	0.97	0.94	0.53	0.94	0.94
$J = 4$						
0.182	Mean (SD)	0.181 (0.070)	0.183 (0.074)	0.167 (0.067)	0.180 (0.072)	0.180 (0.072)
	Coverage	0.95	0.95	0.96	0.95	0.95
0.405	Mean (SD)	0.409 (0.065)	0.414 (0.067)	0.376 (0.061)	0.406 (0.066)	0.406 (0.066)
	Coverage	0.93	0.94	0.90	0.94	0.94
0.693	Mean (SD)	0.695 (0.065)	0.692 (0.067)	0.635 (0.062)	0.685 (0.067)	0.682 (0.066)
	Coverage	0.97	0.96	0.85	0.95	0.95
$J = 10$						
0.182	Mean (SD)	0.181 (0.070)	0.183 (0.071)	0.175 (0.068)	0.181 (0.070)	0.181 (0.070)
	Coverage	0.95	0.95	0.96	0.95	0.95
0.405	Mean (SD)	0.409 (0.065)	0.413 (0.066)	0.395 (0.063)	0.407 (0.066)	0.408 (0.065)
	Coverage	0.93	0.93	0.92	0.93	0.93
0.693	Mean (SD)	0.695 (0.065)	0.697 (0.066)	0.670 (0.064)	0.691 (0.066)	0.690 (0.066)
	Coverage	0.97	0.97	0.92	0.96	0.96
Incomplete repeats						
$J = 2$						
0.182	Mean (SD)	0.181 (0.070)	0.181 (0.081)	0.138 (0.061)	0.195 (0.104)	0.181 (0.081)
	Coverage	0.95	0.96	0.94	0.91	0.96
0.405	Mean (SD)	0.409 (0.065)	0.404 (0.075)	0.310 (0.055)	0.438 (0.144)	0.404 (0.074)
	Coverage	0.93	0.91	0.52	0.70	0.92
0.693	Mean (SD)	0.695 (0.065)	0.659 (0.077)	0.517 (0.058)	0.728 (0.221)	0.658 (0.077)
	Coverage	0.97	0.86	0.16	0.52	0.86
$J = 4$						
0.182	Mean (SD)	0.181 (0.070)	0.185 (0.082)	0.139 (0.062)	0.193 (0.100)	0.180 (0.080)
	Coverage	0.95	0.95	0.94	0.90	0.96
0.405	Mean (SD)	0.409 (0.065)	0.412 (0.073)	0.312 (0.055)	0.433 (0.134)	0.401 (0.071)
	Coverage	0.93	0.93	0.55	0.72	0.92
0.693	Mean (SD)	0.695 (0.065)	0.671 (0.074)	0.522 (0.058)	0.721 (0.203)	0.657 (0.072)
	Coverage	0.97	0.91	0.17	0.57	0.89
$J = 10$						
0.182	Mean (SD)	0.181 (0.070)	0.186 (0.081)	0.140 (0.062)	0.191 (0.096)	0.178 (0.078)
	Coverage	0.95	0.96	0.94	0.90	0.96
0.405	Mean (SD)	0.409 (0.065)	0.416 (0.073)	0.314 (0.056)	0.430 (0.130)	0.398 (0.070)
	Coverage	0.93	0.92	0.55	0.72	0.93
0.693	Mean (SD)	0.695 (0.065)	0.677 (0.072)	0.525 (0.059)	0.714 (0.190)	0.653 (0.070)
	Coverage	0.97	0.94	0.17	0.60	0.88

Table 12. *Parameter estimates (standard error [SE]) from fitting the unadjusted and FFQ-adjusted NEC model using maximum likelihood to one or two 7-day diary measurements of alcohol intake in EPIC-Norfolk*

Parameter	Estimate (SE)	
	Unadjusted	FFQ adjusted
γ_1	2.15 (0.09)	0.13 (0.03)
γ_2	2.69 (0.06)	0.92 (0.03)
$\sigma^2_{u_1}$	4.03 (0.77)	0.03 (0.04)
$\sigma^2_{u_2}$	4.40 (0.15)	0.61 (0.05)
ρ	0.91 (0.01)	0.77 (0.36)
σ^2_{ϵ}	1.17 (0.04)	1.28 (0.05)
$H(\gamma_0)$	0.88 (0.02)	0.37 (0.01)
ξ_1	-	0.90 (0.02)
ξ_2	-	0.88 (0.01)