Published in final edited form as:
Epidemiology. 2009 March ; 20(2): 289–294. doi:10.1097/EDE.0b013e31819642c4.

Missing Data in a Long Food Frequency Questionnaire:

Are Imputed Zeroes Correct?

Gary E. Fraser, **Ru Yan**, **Terry L. Butler**, **Karen Jaceldo-Siegl**, **W. Lawrence Beeson**, and **Jacqueline Chan**
From the Department of Epidemiology and Biostatistics, Loma Linda University, Loma Linda, CA.

Abstract

Background—Missing data are a common problem in nutritional epidemiology. Little is known of the characteristics of these missing data, which makes it difficult to conduct appropriate imputation.

Methods—We telephoned, at random, 20% of subjects (n = 2091) from the Adventist Health Study–2 cohort who had any of 80 key variables missing from a dietary questionnaire. We were able to obtain responses for 92% of the missing variables.

Results—We found a consistent excess of "zero" intakes in the filled-in data that were initially missing. However, for frequently consumed foods, most missing data were not zero, and these were usually not distinguishable from a random sample of nonzero data. Older, black, and less-well-educated subjects had more missing data. Missing data are more likely to be true zeroes in older subjects and those with more missing data. Zero imputation for missing data may create little bias except for more frequently consumed foods, in which case, zero imputation will be suboptimal if there is more than 5%–10% missing.

Conclusions—Although some missing data represent true zeroes, much of it does not, and data are usually not missing at random. Automatic imputation of zeroes for missing data will usually be incorrect, although there is a little bias unless the foods are frequently consumed. Certain identifiable subgroups have greater amounts of missing data, and require greater care in making imputations.

Missing data are a problem in food frequency questionnaires (FFQ), especially when the questionnaires are long. Yet the factors underlying missing dietary data, and the nature of those missing data, are poorly characterized. The methods used to deal with missing data in nutritional epidemiology are frequently not reported, but a common approach is to assume that participants who did not answer a question left it blank because they did not eat that food. Therefore a value of zero is imputed.[1,2] Other approaches are to impute the means of nonmissing values,[3] or to use the indicator variable method,[4,5] which is unsatisfactory in multivariate analyses.[6] Multiple imputation is a more attractive solution,[7] particularly if steps are taken to ensure that data are approximately missing at random.[8] We explore the characteristics of missing data and of subjects who omit these data.

METHODS

The Adventist Health Study (AHS)–2 is a national cohort study.[9] The data reported here pertain to the first 19,611 subjects enrolled, mainly from the western United States. Subjects are Seventh-day Adventists older than 29 years of age. About 50% of Adventists are vegetarian

Correspondence: Gary E. Fraser, Room 203, Evans Hall, Loma Linda University, Loma Linda, CA 92350. E-mail: gfraser@llu.edu.

and many do not drink coffee. A long questionnaire (48 pages) was designed for this study[9–11] and takes 1.5–3 hours to complete (30–60 min for the FFQ). The FFQ section (13 pages) from which items for these analyses were selected contains 130 specified foods, each item having between 7 and 9 possible frequency responses. The first response category, labeled "never or rarely" is specifically identified in methods described later. A standard portion size is given for each item and the subject chooses "standard," "one-half or less than standard," or "one-half or more than standard."

We identified about 80 "key foods." Based on previous pilot work,[12] these included all sets of 4–5 foods that contributed most strongly to validity correlation coefficients for each of 18 FFQ indices of nutrients, vitamins, and minerals.

A random 20% of all subjects with 1 or more of these key items missing (the "sample missing population") were contacted by telephone for the purpose of guiding multiple imputation.[8] In this sample, missing data for particular foods are also a random 20% sample of missing data for each food. Averaging across key foods, it was possible to fill in 92% of these missing data. The average time between the original questionnaire and the follow-up phone call was about 1 year. Subjects were asked to recall their diet at the time of the original questionnaire. We assume that data reported by telephone are not systematically different from those which would have originally been reported on the questionnaire if not omitted.

Data from those successfully contacted in the 20% "sample–missing" subjects are replicated 4 times and combined with data from subjects with "initially complete" information (n = 9165) to provide what we reasonably assume are approximately unbiased estimates of completed data for the total population ("estimated complete population").

For each food, 2 statistical tests were performed to evaluate the distribution of filled in data that were initially missing. The first analysis tests the hypothesis that this distribution does not differ from that of estimated complete data for the total population. The test actually compared 2 independent data sets for each food: the initially complete data for that food, and the initially missing for that food. The second test excluded subjects who responded never or rarely, thus evaluating whether the distributions among remaining response categories differ. The dietary variables in this table were chosen a priori to include a representative range of foods eaten less or more commonly by this population.

To evaluate whether covariates could predict the proportion of filled in initially missing data that were actually zero, we used a logistic regression with a binomial rather than Bernoulli distribution function, among only those in the sample–missing population. The *i*th subject contributed 1 vector observation of length N_i, of zeroes/nonzeroes indicating final disposition of each data point, where N_i is the number of initially missed key variables. The link function was logistic and the error distribution overdispersed "binomial" with a dispersion coefficient (σ^2) of 1.92, which was incorporated into statistical tests.

A log-linear analysis was used to identify demographic factors mat predict the number of missing variables (N_i) among subjects in the total population as initially observed, excluding any filled in data. All 2-way interactions and 3-way interactions that do not include the variable *N*, and age*education**N*, are included, and these provided a satisfactory model fit with deviance $X^2_{138}=139.1$ ($P=0.46$).

We now derive a formula to describe the effects of zero imputation of missing data. Let missing data among nonzero values of dietary variables, X_j, be completely at random, as is approximately the case in our data. Note that the pattern of missing data among zero–valued data does not affect the calculations, as they are automatically again replaced by zeroes in the

imputation. Let P_{nzmj} be the proportion of missing data among nonzero values of dietary variable X_j, μ_j be the overall mean of X_j, and σ_j^2 its variance. After imputing zeroes for missing data we relabel the X_j as X'_j. It can be shown that the regression of some dependent variable Y on the sum of several X'_j, ($j = 1,\ldots J$), has a beta coefficient β' that measures the slope of this regression given by

$$\beta\prime=\sum_{j=1}^{J}(1-P_{\mathrm{nzmj}})\mathrm{Cov}(Y,X_j)/\sum_{j=1}^{J}\Big[(1-P_{\mathrm{nzmj}})(\sigma_j^2 +\mu_j^2\cdot P_{\mathrm{nzmj}})+2\sum_{k=j+1}^{J}(1-P_{\mathrm{nzmj}})(1-P_{\mathrm{nzmk}})Cov(X_j,X_k)\Big|. \quad (1)$$

To sum dietary variables, X_j, is a common procedure when estimating intake from a food group or a particular nutrient.

A ratio of interest is β'/β, where β is the "true" coefficient obtained when $P_{nzmj} = 0$. Equation 1 assumes that missing data are assorted independently between the X_j. While this is unlikely, it is a conservative assumption because a positive correlation between missing data from different X variables will further decrease the ratio β'/β. However, in practical situations the covariance terms are generally much less likely than the μ_j^2 terms to create bias.

Where there is only 1 X in the independent variable (the interest is say in 1 food), then $\beta\prime/\beta=1/[1+(\mu_X{}^2/\sigma_X{}^2)P_{\mathrm{nzmX}}]$. This demonstrates the importance of $\mu_X{}^2$ in relation to $\sigma_X{}^2$ when predicting bias after zero imputation. A similar equation for the ratio of biased to true squared correlation coefficients (ρ^2) between Y and X is $\rho\prime^2/\rho^2=(1-P_{\mathrm{nzmX}})/[1+(\mu_X{}^2/\sigma_X{}^2)P_{\mathrm{nzmX}}]=(1-P_{\mathrm{nzmX}})(\beta\prime/\beta)$.

RESULTS

From the total cohort of 19,611 subjects there were 2091 in the sample missing population, and we were able to contact 1928 of these (92%). The demographic characteristics of the total and sample missing populations are shown in Table 1. As expected, when those with no missing key variables in the total population are excluded (note values in parentheses), the proportions in categories of missing key variables approximately agree between the 2 populations.

Table 2 shows that true values of initially missing data for frequently eaten foods are mostly nonzero. Thus an "automatic" zero imputation here may be problematic. The third and fourth columns of Table 2 are the estimated proportions of zeroes in the initial missing and estimated complete data, respectively. For every food tested, the distribution of values in the initially missing data is always different from that for estimated complete data (see $\chi^2_{(1)}$ in fifth column of Table 2). When the zero intake category is excluded, however, the distribution of initially missing data among other categories is not clearly different from expected values for most foods. Nevertheless, the evidence of excess zeroes among the initially missing data is clear for every food (tested by the difference between $\chi^2_{(1)}$ and $\chi^2_{(2)}$ as a χ^2_1 statistic).

Those participants with the largest numbers of missing items (Table 3) were more likely to be older, black, and less well educated. BMI (not shown) had only a trivial association with the number of missing items. On average, older subjects have a higher proportion of zeroes when missing values are filled in (Table 4). The greater the number of initially missing foods, the higher the proportion of these that are true zeroes, although beyond 25 initially missing key

variables, the probability that a completed value is a zero declines. The statistical evidence for this nonlinearity is strong ($t = 7.62$).

Whether a zero imputation for missing data will create biases of practical importance depends on the variable. As is clear from equation 1 and the results described previously, if a food is eaten infrequently (or more precisely, does not have a mean intake that when squared is much greater than its variance), then a zero imputation will have relatively less influence on a regression coefficient.

The foods selected for Table 2 have an average missing frequency of 4% in the estimated complete nonzero data (P_{nzm}). As we have previously shown,[8] when missing proportions are very small then the form of the imputation for missing data, within reason, has very little effect on any outcome. However, using the observed μ^2/σ^2 for these same variables but setting $P_{nzm} = 0.10$, the values of β'/β for apples, bananas, tomatoes, and broccoli (fairly commonly eaten foods) would be 0.84, 0.78, 0.81, and 0.81. Corresponding values of ρ'^2/ρ^2 are 0.76, 0.70, 0.73, and 0.73. These clearly are nontrivial changes resulting from zero imputation. On the other hand, coffee (an uncommonly consumed food in this population) has a small value of μ^2/σ^2; and even with P_{nzm} of 0.10, the value of $\beta'/\beta = 0.985$ and $\rho'^2/\rho^2 = 0.887$.

In another example, we simulated a situation in which the independent variable is the sum of 2 dietary variables (X_1 and X_2), and the nonzero values of these variables (before missing data are added) were generated using a lognormal distribution such that correlation (X_1, X_2) = 0.75. X_1, with the addition of 10% true zeroes, had a mean frequency of 2.77 servings eaten per week and variance 3.14, with 99th percentile of 8.48. The second variable, with the addition of the 10% true zeroes, had mean = 7.24, variance = 15.63, and 99th percentile = 18.4. The final correlation (X_1, X_2) was 0.50. The second food was thus more commonly consumed and has a higher variance of intake. With 10% of the nonzero data for each of these variables independently missing, β'/β is 0.81 and ρ'^2/ρ^2 is 0.70.

DISCUSSION

There is an excess proportion of zeroes among missing data (Table 2). Even so averaging across all key foods, zero would be an accurate imputation only about 60% of the time. For foods eaten more frequently, a zero imputation will usually be incorrect as most initially missing values are nonzero and typically have a distribution that is similar to nonzero data that were not initially missing. Whether this will affect estimates of relative risk also depends on the proportion of data initially missing,[8] how commonly this food is eaten, and the importance of the missing variable to that nutrient or food group.

Our data are consistent with those of others[12,13] who note that a zero imputation will often (although not always) produce data that have a correlation with original (no missing) data exceeding 0.90. However, as we have shown, the influence on a β coefficient measuring the slope of a regression or the squared correlation coefficient of that regression may be somewhat greater. We have used a linear disease regression as the example around which to explore the effects of zero imputation. Theoretically the situation should be at least qualitatively similar for a logistic disease regression,[14–16] although more work remains to be done.

Particular situations may make zero imputation an especially bad choice. Our recent need to identify vegans in this study was based on data from more than 20 variables. A zero imputation for missing data would have exaggerated the proportion of vegans, who consume no meat, dairy, or eggs. Further, if in equation 1 T (true intake of X) is substituted for Y, this regression is that related to the validity correlation for X. A further attenuation of even 10% because of zero imputation in this validity correlation coefficient would usually be considered undesirable, as it is usually already markedly attenuated by measurement error.[17]

Although there is a substantial literature about the characteristics of missing subjects[18] in surveys, there are surprisingly few reports about the characteristics of missing data from those enrolled in nutritional epidemiologic studies. One report[13] found that 76% had initially left at least 1 item blank. On follow-up (89% complete), 55% of such foods were consumed never or less than once per month—similar to our observations.

In a study from Sweden[19] using a 56-item food frequency questionnaire, investigators were able to contact by telephone more than half of subjects with missing data, thus leaving smaller percentages of missing data. They also demonstrated that, for commonly consumed foods, initially missing data were infrequently true zeroes. Whether they imputed a zero or the observed median usually made little difference to estimates, though in some circumstances differences were larger.

Caan et al[12] found that the probability of a perfect questionnaire response was inversely related to age, and differed by race. The correlation between nutrients calculated from questionnaires with and without missing data was usually high but, as expected, was more adversely affected when questionnaires with higher numbers of missing items were included. We previously found[8] that zero imputation of missing data in a real data set often resulted in point estimates of regression coefficients that differed from the best estimator by 12%–18%.

Using guided multiple imputation[8] to handle missing data will usually be a good choice if it is practical to contact a subsample to fill in missing data. This allows the missing-at-random assumption to be approximately satisfied, thus practically eliminating further biases and attenuation due to imputation.

It is clear that different subgroups of the population have different amounts of missing data. Indeed, in less–well–educated black subjects older than 80 years of age the model predicts that about 5 times as many would be placed in the “more than 20 missing” category as compared with young college-educated white subjects. Many known barriers interfere with the participation of black subjects in research.[20]

Missing data in older subjects are a little more likely to represent true zeroes. Perhaps their greater susceptibility to fatigue leads to skipping foods not eaten. Those with about 25 missing items had the greatest proportion that are true zeroes, perhaps suggesting that those who systematically skip items not eaten, on average do not eat about 25 of these key foods. When there are more than 25 missing items the proportion of true zeroes decreases again, suggesting that these tend to be participants who are less committed or less able to complete the questionnaire carefully.

In summary, a systematic zero imputation will produce biased results for commonly consumed foods although the magnitude of that bias will depend on the particular situation. Deleterious effects on bias and validity will often be quite small, but may be more severe. Given the challenges to validity before considering missing data, even modest additional error due to inaccurate imputation should be avoided if possible. The elderly, less–well–educated, and black subjects tend to have more missing data, and subsamples used to guide multiple imputation[8] should ensure good representation from these groups.

Acknowledgments

Supported by NIH grant R01 CA094594.

REFERENCES

1. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol 1992;135:1114–1126. [PubMed: 1632423]
2. Willett, W. Nutritional Epidemiology. Vol. 2nd ed. New York: Oxford University Press; 1998. p. 322
3. Verkasalo PK, Appleby PN, Davey GK, Key TJ. Soy milk intake and plasma sex hormones: A cross sectional study in pre- and postmenopausal women (EPIC-Oxford). Nutr Cancer 2001;40:79–86. [PubMed: 11962259]
4. Singh PN, Fraser GE. Dietary risk factors for colon cancer in a low risk population. Am J Epidemiol 1998;148:761–764. [PubMed: 9786231]
5. Smith-Warner SA, Spiegelman D, Yaun SS, et al. Fruits, vegetables, and lung cancer a pooled analysis of cohort studies. Int J Cancer 2003;107:1001–1011. [PubMed: 14601062]
6. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am J Epidemiol 1995;142:1255–1264. [PubMed: 7503045]
7. Rubin D. Inference and missing data. Biometrika 1976;63:581–592.
8. Fraser GE, Yan R. Guided multiple imputation of missing data: using a sub-sample to strengthen the missing at random assumption. Epidemiology 2007;18:246–252. [PubMed: 17259903]
9. Butler T, Fraser GE, Beeson WL, et al. Cohort profile: the Adventist Health Study-2. Int J Epidemiol 2008;37:260–265. [PubMed: 17726038]
10. Knutsen SF, Fraser GE, Beeson WL, Lindsted KD, Shavlik DJ. Comparison of adipose tissue fatty acids with dietary fatty acids as measured by 24-hour recall and food frequency questionnaire in black and white Adventists: the Advcntist health study. Ann Epidemiol 2003;13:119–127. [PubMed: 12559671]
11. Knutsen SF, Fraser GE, Lindsted KD, Beeson WL, Shavlik DJ. Comparing biological measurements of vtamin C, folate, alpha-tocopherol, and carotene with 24 hour dietary recall information in nonhispanic blacks and whites. Ann Epidemiol 2001;11:406–416. [PubMed: 11454500]
12. Caan B, Hiatt RA, Owen AM. Mailed dietary surveys: response rates, error rates, and the effect of omitted foods on nutrient values. Epidemiology 1991;2:430–436. [PubMed: 1790195]
13. Michels KB, Willett WC. Predictors and interpretation of missing data on a self–administered food frequency questionnaire [abstract]. Am J Epidemiol 2004;159:S56.
14. Carroll, RJ.; Ruppert, D.; Stefanski, LA. Measurement Error in Nonlinear Models. New York: Chapman and Hall; 1996. p. 51-54.
15. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Stat Med 1989;8:1051–1069. [PubMed: 2799131]
16. Carroll, RJ.; Ruppert, D.; Stefanski, LA. Measurement Error in Nonlinear Models. New York: Chapman and Hall; 1996. p. 65-66.
17. Fraser GE, Shavlik DJ. Correlations between estimated and true dietary intakes. Ann Epidemiol 2004;14:287–295. [PubMed: 15066609]
18. Fowler, F. Bias associated with non-response. In: Fowler, F., editor. Survey Research Methods. Thousand Oaks, CA: Sage Publications Inc; 2002. p. 41-45.
19. Hansson LM, Galanti MR. Diet-associated risks of disease and self-reported food consumption: how shall we treat partial non-response in a food frequency questionnaire? Nutr Cancer 2000;36:1–6. [PubMed: 10798209]
20. Herring P, Montgomery S, Yancey A, Williams D, Fraser G. Understanding the challenges in recruiting blacks to a longitudinal cohort study: the Adventist Health Study. Ethn Dis 2004;14:423–430. [PubMed: 15328945]

TABLE 1

Characteristics of the Analytic Populations

Variable	Value	Total Population (n = 19,611) (%)	20% Sample of Population With ≥ Key Variable Missing Initially[a] (%)
Male		37	39
Age (years)	<50	20	16
	50–64	36	30
	65–79	29	34
	≥80	14	19
No. missing key variables	0	45	—
	1	20.7 (37)[b]	37
	2–5	18.9 (34)	36
	6–10	5.6 (10)	10
	11–20	6.1 (11)	10
	>20	4.1 (7)	6
Race	Black	15	17
	Other	85	83
Education	Grade school or some high school	7	9
	High school graduate or some college	43	46
	College degree	50	45

[a] Sample missing population.

[b] Proportions when first category is excluded.

TABLE 2

Distribution of True Values Among Data That Were Initially Missing and Reconstituted Complete Data for the Total Population

Variable	No With Initially Missing Data[a]	Proportion of Zeroes		$\chi^2_{(1)}$ [c]	*df*	$\chi^2_{(2)}$ [c]	*df*	*P*
		Sample Missing Data	Estimated Complete Data[b]					
Apples	104	0.269	0.058	98.7	4	3.04	3	0.39
Oranges	135	0.452	0.318	131.1	4	5.31	3	0.15
Bananas	62	0.161	0.058	16.1	3	3.55	2	0.17
Orange juice	209	0.589	0.225	192.3	4	11.57	3	0.01
Tomatoes	166	0.229	0.065	97.7	4	6.22	3	0.10
Carrots	181	0.392	0.141	112.5	4	3.58	3	0.31
Refried beans	233	0.691	0.297	207.2	3	6.39	2	0.04
Broccoli	127	0.173	0.065	28.2	3	0.99	2	0.61
Cooked carrots	171	0.415	0.235	38.7	3	2.36	2	0.31
Peanuts	309	0.566	0.406	44.5	4	8.14	3	0.04
Low calorie mayonnaise	275	0.735	0.575	31.9	4	0.63	3	0.89
Low fat milk	606	0.743	0.681	40.1	6	6.87	5	0.23
Poultry	224	0.674	0.540	22.0	3	7.05	2	0.03
Beef/hamburger	237	0.895	0.695	48.0	2	3.04	1	0.08
Regular coffee	349	0.883	0.765	30.4	3	5.02	2	0.08

[a] From the 20% sample of those with missing data for at least 1 key variable (sample missing population).

[b] Proportion of zeroes in estimated complete data representing the total population. This equals $P_M Z_M (1 - P_M) Z_C$ where Z_M and Z_C are proportions of zeroes in those with initially missing or complete data, respectively, and P_M is the proportion of subjects with initially missing data.

[c] $\chi^2_{(1)}$ tests the hypothesis, food by food, that the distribution of true values for data initially missing does not differ from distribution of true values in the estimated complete population. All *P* for $\chi^2_{(1)}$ are <0.0001 except bananas <0.001; $\chi^2_{(2)}$ tests the same hypothesis but excluding the zero intake category, and the corresponding *P* is in the final column.

TABLE 3

[a]A Log-linear Analysis Predicting the Number of Missing Key Items in the AHS-2 Data: Adjusted Relative Frequencies

Variable	No Missing	Missing 1	Missing 2–5	Missing 6–20	Missing >20
Age (years)					
<50	1.0	0.38	0.23	0.09	0.02
50–64	1.0	0.40	0.29	0.15	0.03
65–79	1.0	0.53	0.54	0.36	0.10
≥80	1.0	0.58	0.92	0.98	0.50
Sex					
Male	1.0	0.59	0.68	0.40	0.12
Female	1.0	0.53	0.54	0.36	0.10
Race					
Black	1.0	0.87	1.34	0.97	0.51
Nonblack	1.0	0.53	0.54	0.36	0.10
Education and age <50 years					
Grade school/some high school	1.0	0.41	0.38	0.31	—
High school diploma/some college	1.0	0.38	0.23	0.09	0.02
College degree	1.0	0.35	0.15	0.06	0.01
Education and age ≥80 years					
Grade school/some high school	1.0	0.64	0.89	1.47	0.91
High school diploma/some college	1.0	0.58	0.92	0.98	0.50
College degree	1.0	0.71	0.97	0.76	0.25

[a]Adjusted relative frequencies are adjusted for all covariates shown, with the reference being age 50–64, female sex, nonblack race, and education high school diploma/some college. All variables have significant associations with number of missing key items, $P < 0.0001$, thus rejecting the null hypothesis that different levels of the variable have the same distribution of missing-ness. The lost 2 categories show the interaction between age and education with the 2 extreme age categories shown only for economy of space. Education has less effect in old age.

– indicates no data observed in this cell.

TABLE 4

Predictors of the Proportion of Missing Data That are True Zeroes—Logistic Regression With Binomial Distribution Function

Variable	Value/Unit	Coefficient[a]	(SE)[b]	Exposure Value	pr (Missing Item is Zero)[c]
Intercept		−0.080	(.197)		
	Male	—	—	Male	0.620
	Female	0.042	(.061)	Female	0.630
Age[d]	(Per 10 years)	0.068	(.021)	30 years	0.589
				80 years	0.676
Race	Nonblack	—	—	Nonblack	0.630
	Black	0.019	(0.76)	Black	0.634
Education	Grade school	—	—	Grade school	0.614
	High school	0.069	(0.084)	High school	0.630
	College	−0.021	(0.090)	College	0.609
No. missing key items[e]	1	—	—	1	0.592
	2–5	−0.087	(0.127)	2–5	0.571
	6–10	0.161	(0.134)	6–10	0.630
	11–15	0.430	(0.134)	11–15	0.690
	16–20	0.468	(0.136)	16–20	0.698
	21–25	0.619	(0.143)	21–25	0.729
	26–30	0.415	(0.148)	26–30	0.687
	>30	0.037	(0.150)	>30	0.601

[a] — is reference category

[b] Incorporates σ, the over dispersion coefficient.

[c] Variables other than the I of interest are set at female, age = 50, race = nonblack, education = high school, and number of missing items = 10.

[d] $P = 0.001$.

[e] $P < 0.0001$.

SE indicates standard error.