Supplemental Materials

STATISTICAL APPENDIX

Concentration model

Estimating the relationship between months of folic acid supplementation and RBC folate concentration using the Folic Acid Dosing Trial (FADT) data

The FADT study was a comparison of folic acid dosing schemes in which one arm was comprised of 317 *MTHFR* genotyped women taking daily doses of 400 µg of folic acid. Among these women, 54 (17%) had the CC allele, 148 (47%) had the CT genotype, and 115 (36%) were TT. Measurements of RBC folate concentration were obtained from the participants prior to initiation of supplementation (0 months) and at 1, 3, and 6 month intervals after they began daily dosing. The distribution of measured RBC concentrations among FADT participants was highly skewed and, as a result, the natural logarithm of the measurements was used in constructing models relating time on pills (recorded as 0, 1, 3 and 6 months) and RBC folate concentration. A generalized estimating equation (GEE) approach, using an identity working correlation matrix, was used to estimate the standard errors associated with fitted model parameters to address the potential for correlation due to the repeated measurement of RBC folate within study participants. After assessing a variety of candidate models, a quadratic model was selected as the best descriptor of the association between months taking pills and RBC folate concentration on the basis of both visual assessments of model fit and comparison of model adequacy measures.[1] *MTHFR* genotype did not appear to alter the relationship between time on pills and subsequent RBC concentration (**Fig C**). Genotype, however, did significantly impact the baseline RBC folate concentration, with the CC allele having the highest concentration followed by CT and TT. As a result, the dose model selected as best representing the observed information in the FADT data, relating time on 400 µg folic acid pills and RBC folate concentration, is given by

$$log(RBC_{ij}) = \beta_0 + \beta_1 CT_i + \beta_2 TT_i + \beta_3 Month_j + \beta_4 Month_j^2 \quad [1]$$

where $log\ (RBC_{ij})$ is the *i*th individual's measured log transformed RBC folate concentration in $Month_j$ (0, 1, 3 or 6), CT_i takes the value of one if individual *i* has genotype CT, and TT_i takes the value of one if the individual *i* has genotype TT. The estimated parameters and correlation matrix for the parameters in the model are presented in **Table D,** along with the standard deviation associated with the model fit, defined as the square root of the average squared residual. Alternative working correlation matrices were evaluated as a sensitivity assessment, but no alternative selection resulted in meaningful variation in either the estimated parameters or their standard errors.

Application of the FADT model for estimating RBC folate concentration among Community Intervention Project (CIP)participants

The model relating intake and RBC folate concentration given in equation 1 was modified to increase its applicability to members of the CIP cohort. Specifically, CIP participants from the southern region likely differ from the northern region FADT study participants in baseline RBC folate concentration.[2-4] To reflect this difference, we adjusted the FADT dose model by assuming that the RBC folate concentration for CIP study participant *i* at the time of neural tube closure can be modeled as

$$log(RBC_i) \sim N(\mu_i, \sigma^2)$$

where $log(RBC_i)$ is the natural log of the RBC folate concentration at closure for the *i*th woman in the CIP study population and μ_i has the assumed form

$$\mu_i = \beta_0 + \beta_1 CT_i + \beta_2 TT_i + \beta_3 Month_i + \beta_4 Month_i^2 + \beta_5 South_i . \quad [2]$$

In equation 2, CT_i takes value one if individual *i* has genotype CT, TT_i has value one if the individual is genotype TT, $Month_i$ is the months on pills for woman *i*, adjusted for compliance, from first clinic visit until date of NT closure, and $South_i$ takes value one if participant *i* resides in the southern region and zero if she resides in the north. Calculated months of folic acid supplement consumption is presented in **Table C.** To utilize equation 2 in the Bayesian analysis, we needed to develop prior estimates for all model parameters. Prior assumptions on $\beta_0, \beta_1, \beta_2, \beta_3$ and β_4 were that these parameters follow a multivariate Normal prior distribution (MVN) reflecting the fit of the model to the FADT study population such that

$$(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) \sim MVN(\underline{\beta}, \Sigma)$$

where $\underline{\beta}$ and Σ are the estimated coefficients and the covariance matrix derived in the fit of the FADT data corresponding to the results presented in **Table B.** The term β_5 in equation 2 reflects the difference in the mean log of RBC folate among women taking no supplemental pills due to residing in the South. Prior assumptions on the value of this parameter were based on data presented in Table 3 of Hao and colleagues[2] in which the ratio of the mean observed RBC folate concentration among southern to northern Chinese women aged 35-44 was 1.79, corresponding to a difference in mean log transformed folate concentration of 0.58. Comparison of Hao and colleagues's [2] southern Chinese women's mean estimate to that observed among northern Chinese women participating in the FADT [5 6] led to a south minus north difference in mean log RBC folate of 0.34. To reflect the uncertainty in our knowledge regarding regional differences in mean baseline log RBC folate concentration, we considered the prior distribution for the parameter reflecting the increase in baseline mean log RBC folate in the South to be

$$\beta_5 \sim N(0.45, 0.08^2)$$

leading to a prior 95% prior uncertainty interval (UI) of [0.29, 0.61] which captures both the estimate from Hao and that based on comparison of that result to the observed log concentrations observed among the northern women participating in the FADT study. The prior distribution for σ, the standard error of the assumed RBC concentrations about the modeled mean, was assumed to be uniform bounded between 0.1 and 0.5. This prior was selected both to allow substantial uncertainty concerning the adequacy of the model in the CIP population and to incorporate the estimated standard error observed in the analysis of the FADT data.

Genotype model

Analysis of the data from the FADT study indicated that *MTHFR* genotype exerts a meaningful influence on baseline RBC folate concentration. Because genotype information was not available for CIP study participants and because the distribution of *MTHFR* has been observed to differ dramatically between residents of northern and northern China,[3 4] we developed different models for genotype distribution for northern and southern participants. To do this, each CIP participant was assigned a set of variables CC_i, CT_i and TT_i where $CC_i = 1$ if individual *i* has *MTHFR* allele CC; $CT_i=1$ if individual *i* has the *MTHFR* allele CT; and $TT_i=1$ if individual *i* has *MTHFR* allele TT. We assumed two sets of probabilities, $P_S(CC)$, $P_S(CT)$ and $P_S(TT)$, for participants from the southern region and $P_N(CC)$, $P_N(CT)$ and $P_N(TT)$ for those from the North. For individual *i*, membership in one of the *MTHFR* genotypes was modeled using a multinomial distribution (*Mul*) with a sample size of one such that

$$(CC_i, CT_i, TT_i,) \sim Mul(P_s(CC), P_s(CT), P_s(TT)\ , 1,)$$

if she is from the southern region and

$$(CC_i, CT_i, TT_i,) \sim Mul(P_N(CC), P_N(CT), P_N(TT)\ , 1,)$$

if she is from the northern region. Dirichlet (*Dir*) priors were assumed for both $P_S(CC)$, $P_S(CT)$, $P_S(TT)$ and $P_N(CC)$, $P_N(CT)$,$P_N(TT)$, and were based on one of two sources depending on the residence of study participants. For CIP study participants from the northern region, priors for the genotype probabilities were based on the observed genotype distribution in the entire cohort of the FADT study population (all participants in the study regardless of their assignment to a folic acid supplementation dosing regime). Among the total of 1194 FADT participants, the observed distribution of genotypes was CC = 196 (16%), CT = 559 (47%) and TT = 439 (37%) leading to a prior for northern region CIP participants of *Dir* (196, 559, 439). The Dirichlet prior for the southern region was based on available information from the literature. Mao and colleagues (Table 2)[4] reported probabilities of $P(CC)$ = 0.39, $P(CT)$ = 0.53 and $P(TT)$ = 0.08 among 217 female Han women from the southern region of China. In an ethnically similar population also from southern China, Wilcken and colleagues (Table 2)[3] reports identical allele probabilities among 430 genotyped females. Combining these data leads to a Dirichlet prior for the genotype probabilities among southern participants in the CIP study of *Dir*(252, 343, 52).

Risk model

A logistic regression model was used to relate the log odds of having a child or fetus with an NTD among CIP study participants to the estimated log RBC concentration at neural tube closure. The form of the model used is

$$log(O_i) = \delta_0 + \delta_1 * \log(RBC_i) \qquad [3]$$

where, if p_i is the probability of subject *i* having a child with a NTD, then

$$log(O_i) = log(p_i / (1 - p_i)).$$

The parameter δ_0 in equation 3 represents the log odds of having a child with an NTD when a woman's RBC folate concentration is equal to one nmol/L and δ_1 is the log odds ratio reflecting the increase in the

odds of an NTD for an increase in RBC folate of 2.72 (or e^1) nmol/L. Prior values for the parameters in equation 3 were selected to be non-informative with means reflecting the background level of NTD risk in the Chinese population [7] of approximately 10 NTDs per 10 000 pregnancies for δ_0 and no effect of RBC folate concentration for δ_1. To reflect this, we assumed δ_0 and δ_1 to follow a multivariate normal prior distribution such that

$$\begin{bmatrix}\delta_0\\ \delta_1\end{bmatrix} \sim MVN\left(\begin{bmatrix}-7\\ 0\end{bmatrix}, \begin{bmatrix}16 & \rho\sqrt{16*100}\\ \rho\sqrt{16*100} & 100\end{bmatrix}\right).$$

Note that the assumed prior variance for δ_0 implies a 95% prior uncertainty for the risk at 1 nmol/L ranging from 0.2 to 460 NTDs per 10 000 pregnancies. Similarly, the large prior variance of 10 for δ_1 was selected to place minimal weight on the assumed prior mean. The correlation coefficient, ρ, relating δ_0 and δ_1 was assumed to follow a uniform prior distribution bounded by -1 and 1.

Markov Chain Monte Carlo fitting algorithm

A Markov Chain Monte Carlo (MCMC) algorithm was used to derive posterior estimates of the model parameters [8] using WinBUGS 1.4.3 software. The MCMC updating process was comprised of two chains, each with differing initial values for the parameters of the risk model, with each chain run for 100,000 iterations. In one chain, initial values for all parameters were set to the means of the assumed prior distributions. Initial values for the parameters of the risk model in the second chain were set to the estimated values derived in a logistic regression relating a single fixed estimate of RBC dose for each participant to the observed collection of NTD outcomes. The prior dose estimates for this fit were predicted using equation 2 with the parameters of the dose model set to their assumed prior means and the variables CT_i and TT_i set to their prior means of 0.53 and 0.08 for CIP participants from the southern region and 0.47 and 0.37 for study participants from the northern region. Fitting this regression lead to

initial values for the risk model parameters of 18 for δ_0 and -3 for δ_1. To increase the likelihood of convergence to the true posterior, we discarded the initial 40,000 iterations for each chain as burn-in samples. The history plots for the parameters of the risk model and Gelman-Rubin statistic[8] plots for these parameters are presented in **Fig F**. Evaluation of these plots indicates that convergence appears likely after the assumed 40 000 iteration burn in for these parameters. Similar evidence indicated likely convergence for all other model parameters. After discarding the initial 40 000 samples, we retained only every 6th sample from the remaining 60 000 iterations of each chain to reduce autocorrelation. The remaining 10 000 samples from each chain were then combined leading to a final set of 20 000 posterior samples on which all summary statistics were based. We summarized the central tendency of the estimated posterior distribution using the median and the uncertainty associated with the estimates using a 95% equal tailed posterior interval defined by the 2.5th and 97.5th percentiles of the 20 000 samples. Posterior estimates for all model parameters are given in **Table F.**

Estimation of NTD risk at specified RBC folate concentrations

Posterior estimates of NTD risk presented in **Table 1, Fig 4 and Table G** were derived based on the posterior samples of the risk model parameters. To illustrate, let δ_0^j and δ_1^j be the jth , j = 1, 2, …20 000, sample from the posterior distribution for these parameters. Let *rbc* be an assumed fixed value for RBC folate concentration. Then the corresponding jth posterior sample for the NTD risk given *rbc*, $P_j|rbc$, was estimated as

$$P_j|rbc = \frac{e^{\delta_0^j + \delta_1^j \ln(rbc)}}{1 + e^{\delta_0^j + \delta_1^j \ln(rbc)}} \quad .$$

Alternatively, given a specified NTD risk of *P*, we used the equation

$$rbc_j \mid P = \frac{1}{\delta_1^j}\left[\log\left(\frac{p}{1-p}\right) - \delta_0^j\right]$$

to derive posterior estimates of the RBC folate concentration, rbc_j, associated with an NTD risk equal to *P*.

To generate estimates of NTD prevalence before and after fortification in the U.S. the model was applied to published RBC folate concentrations. [9 10] The published RBC folate concentration centiles were normalized to the method used in papers by both the Daly and Hao, using a standardizing equation generated from Pfeiffer 2011[11] - NHANES RBC folate [nmol/L] = (Dublin RBC folate [nmol/L] * 0.7876) + 34.2802 [nmol/L]-personal communication. These adjusted RBC folate concentrations were used to generate a modeled population with a similar distribution of RBC folate concentration, and then the modeled associations in the Chinese data were used to predict NTD prevalence in that modeled U.S. population.

Sensitivity of results to prior assumptions

The sensitivity of the estimates to assumptions on the prior distributions was evaluated by comparing the estimated posteriors derived under the assumptions described above with those developed using alternative priors for the parameters of the risk model and for the variance of the log RBC folate concentrations about their assumed mean, which is the prior for σ^2 in the assumed distribution for $log(RBC_i)$. Data on observed NTD risk for participants in the study reported by Daly and colleagues[12] were obtained through personal communication with Dr. Anne Molloy (co-author). These data allowed estimation of the proportion of participants in each decile of the distribution of measured RBC folate concentrations who had a child with an NTD. Confidence intervals for the observed proportion of study participants with an NTD within each of the deciles of RBC folate concentration were derived using the methods outlined by Daly and colleagues.[12] No information from the Daly and colleagues analysis was

utilized to inform development of prior estimates for any parameters in the genotype, concentration, or risk models. This exclusion enabled comparison of modeled results with those from Daly's original analysis. However, the alternative prior for the parameters of the risk model had identical covariance as that for the primary analysis but had mean values of $\delta_0 = 1.6$ and $\delta_1 = -1.2$. These choices correspond to the estimates presented in Daly and colleagues [12] from their fit of a logistic regression model relating measured RBC folate concentrations and NTD risk in an Irish population. We stress that these priors were evaluated only after we completed our analysis with the first set of assumed priors described in the main text and that the Daly results were not utilized in any other way in this analysis other than this sensitivity assessment. This exclusion of Daly's finding was done to facilitate a comparison of our results with theirs.

In addition, we considered an alternative gamma (γ) prior for the variance of the log RBC folate concentration about the mean with parameters of 0.01 and 0.01. This highly uninformative comparison prior was selected to ensure that our use of the uniform prior for the variance over a potentially limited range did not constrict possible values for the posterior estimates of the parameters of the risk model. Use of the Daly estimates as priors for the risk model resulted in slightly higher median posterior estimates of NTD risk at RBC concentrations less than 400 nmol/L. However, these RBC concentrations were below the range of folate concentrations estimated to occur among the CIP participants. In the range of folate concentrations of primary interest in our analysis, approximately 500 to 1500 nmol/L, use of the Daly-based prior resulted in no meaningful changes in the results presented in this paper. To illustrate, **Fig E** shows the 95% posterior uncertainty intervals for NTD risk derived under the two prior assumptions across this folate concentration range. Notice that there is substantial overlap in the uncertainty intervals associated with the NTD risk posterior estimates. Results under the alternative gamma prior for the dose model standard error were virtually identical to those produced when the standard error of the RBC concentrations was assumed to follow the uniform prior.

Due to the large amount of prior evidence available for the genotype model, we did not assess implications of deviations in the prior assumptions for these parameters. In addition, due to the large number of parameters in the concentration model, the availability of the FADT data and the large uncertainty placed on the parameter relating the change in baseline RBC concentration between regions, we did not assess prior assumptions for the concentration model beyond the evaluation of the impact of the prior assumption on the standard error of individual RBC folate concentrations about the assumed mean.

ASSAY CONSIDERATIONS

When comparing the RBC folate concentration estimates presented in this paper with those of other populations, there are a number of points to consider. First, the assays used to measure RBC folate concentrations vary widely and standardization and comparison between individual assays can be impossible.[13 14] Fortuitously, the microbiological assay used in the FADT [5] utilized the same methodology as the Daly and colleagues [12] study. Recent studies have shown differences in measurement even among microbiological assays (with different calibrators and microorganisms); however, conversions are available to enable some general comparisons, with appropriate important caveats.[11 13] Using the microbiological assay calibrated with 5' methylTHF (as is currently done by NHANES), RBC folate concentrations corresponding with the 6 NTD per 10 000 births risk (1180 nmol/L; 95%UI 1050 to 1340) would be 964 nmol (95% UI 861 to 1090 nmol/L).

References:

1. Pan W. Akaike's information criterion in generalized estimating equations. Biometrics 2001;**57**(1):120-5.

2. Hao L, Ma J, Stampfer MJ, et al. Geographical, Seasonal and Gender Differences in Folate Status among Chinese Adults. J Nutr 2003;**133**(11):3630-35.

3. Wilcken B, Bamforth F, Li Z, et al. Geographical and ethnic variation of the 677C>T allele of 5,10 methylenetetrahydrofolate reductase (MTHFR): findings from over 7000 newborns from 16 areas world wide. Journal of medical genetics 2003;**40**(8):619-25.

4. Mao R, Fan Y, Chen F, et al. Methylenetetrahydrofolate reductase gene polymorphisms in 13 Chinese ethnic populations. Cell biochemistry and function 2008;**26**(3):352-8.

5. Hao L, Yang Q-H, Li Z, et al. Folate status and homocysteine response to folic acid doses and withdrawal among young Chinese women in a large-scale randomized double-blind trial. Am J Clin Nutr 2008;**88**(2):448-57.

6. Crider KS, Zhu JH, Hao L, et al. MTHFR 677C->T genotype is associated with folate and homocysteine concentrations in a large, population-based, double-blind trial of folic acid supplementation. Am J Clin Nutr 2011;**93**(6):1365-72.

7. Berry RJ, Li Z, Erickson JD, et al. Prevention of neural-tube defects with folic acid in China. China-U.S. Collaborative Project for Neural Tube Defect Prevention. N Engl J Med 1999;**341**(20):1485-90.

8. Gelman A. *Bayesian data analysis*. 2nd ed. Boca Raton, Fla.: Chapman & Hall/CRC, 2004.

9. Pfeiffer CM, Hughes JP, Lacher DA, et al. Estimation of trends in serum and RBC folate in the U.S. population from pre- to postfortification using assay-adjusted data from the NHANES 1988-2010. The Journal of nutrition 2012;**142**(5):886-93.

10. Branum AM, Bailey R, Singer BJ. Dietary supplement use and folate status during pregnancy in the United States. The Journal of nutrition 2013;**143**(4):486-92.

11. Pfeiffer CM, Zhang M, Lacher DA, et al. Comparison of serum and red blood cell folate microbiologic assays for national population surveys. The Journal of nutrition 2011;**141**(7):1402-9.

12. Daly LE, Kirke PN, Molloy A, et al. Folate levels and neural tube defects. Implications for prevention. JAMA 1995;**274**(21):1698-702.

13. Pfeiffer C, Fazili Z, Zhang M. Folate analytical methodology. In: Bailey LB, ed. Folate in Health and Disease. 2nd ed: CRC Press, Taylor and Francis Group, 2009.

14. Fazili Z, Pfeiffer CM, Zhang M, et al. Erythrocyte Folate Extraction and Quantitative Determination by Liquid Chromatography-Tandem Mass Spectrometry: Comparison of Results with Microbiologic Assay. Clin Chem 2005;**51**(12):2318-25.

Table A Sources and types of data available to model the association of red blood cell (RBC) folate concentration and neural tube defect (NTD) risk

Sources	**Data Available**
Community Intervention Project (CIP) (individual level data)[7]	Pregnancy outcome: NTD Yes vs. No Folic acid supplement use: Yes vs. No Folic acid supplement dose: all 400 µg/d Pill taking compliance (% compliance from pill count) Date of last menstrual period Pill start and stop date Region: From northern (Hebei) or southern (Zhejiang and Jiangsu) regions
Folic Acid Dosing Trial (FADT) (individual level data)[5 6]	RBC folate concentrations at enrollment, 1, 3, and 6 months of folic acid supplementation for the 400 µg/d dose groups Region: All from northern Region (Hebei) Folic acid supplement dose: all 400 µg/d Pill taking compliance (% compliance from pill count) *MTHFR 677* genotypes northern region Baseline RBC folate concentrations in northern Chinese women
Literature	Baseline RBC folate concentrations in southern Chinese women[2]
	MTHFR 677 genotype distribution southern Chinese [3 4]

Table B Assumptions on the parameters and prior distributions for the RBC folate concentration, genotype and NTD risk models

Models Model Output	**Assumed Distribution**	**Model Parameters**	**Prior Distribution***	**Data Source**†
RBC Dose Model Natural log RBC Folate Concentration	Normal with mean reflecting the background concentration of log RBC folate, *MTHFR* genotype, change in background log concentration due in pill consumption and difference in log concentration due to region of residence with variance σ^2	Background log concentration, change in background to genotype, and change in log concentration due to folic acid supplementation	Multivariate normal with mean and covariance provided in supplemental materials	FADT[6] Analysis
		Change in log concentration due to residence in northern region	Normal (-0.45. 0.08)	Hao et al[2], FADT[6] Analysis
		σ^2	Uniform (0.1, 0.5)	FADT[6] Analysis
Genotype Model *MTHFR* Genotype Distribution	Multinomial with genotype probabilities CC_S, CT_S and TT_S for the southern region and CC_N, CT_N and TT_N for the northern region	CC_S, CT_S and TT_S	Dirichlet (252,343, 52)	Wilcken et al[3], Mao et al[4]
		CC_N, CT_N and TT_N	Dirichlet(196,559,439)	FADT[6]
Risk Model Log Odds of NTD Risk	Logistic regression model such that log(odds NTD) $= \delta_0 +$ δ_1 * log RBC	δ_0	Normal (-7, 16)	Berry et al [7]
		δ_1	Normal (0, 100)	

*The prior distribution is the assumed distribution for the unknown model parameters prior to incorporating the information on months taking pills and NTD outcome observed in the CIP. It represents initial beliefs about the values of the model parameters that are then updated in the Bayesian estimation process based on the observed information.

†The data source is the collection of information on which assumptions concerning the prior distribution are based.

Table C Adjusted months taking daily doses of 400 µg folic acid pills until closing of fetal neural tube by region in the Community Intervention Project (CIP)

	Region					
	North		South		Total	
Months Taking Supplement*	N	%	N	%	N	%
0	14,377	48	12,3382	62	137,759	60
<1	1,819	6	12,158	6	13,977	6
1-2	3,917	13	15,649	8	19,566	9
2-3	3,593	12	13,376	7	16,969	7
3-4	2,546	9	10,183	5	12,729	6
4-6	2,511	8	12,751	6	15,262	7
>6†	1,283	4	10,911	5	12,194	5
Total	30,046		198,410		228456	

* Estimated time of closure of fetal neural tube was date of initiation of last reported menstrual cycle plus 42 days. Months on pills was calculated based on pill start date until neural tube closure and were adjusted for pill taking compliance based on monthly counts of unused pills.

†Months on pills was truncated at 9 months for 289 (1%) women from the northern region and 3711 (2%) from the South

Table D Estimated parameters for quadratic model* relating months on daily doses of 400 µg of folic acid to measured red blood cell folate concentration derived from Folic Acid Dosing Trial data

			Correlation Coefficients			
Parameters	Estimate	Std. Error	β_1	β_2	β_3	β_4
β_0	6.54	0.03	-0.72	-0.70	-0.44	0.36
β_1	-0.10	0.04		0.70	0.00	0.00
β_2	-0.24	0.04			0.00	0.00
β_3	0.19	0.01				-0.97
β_4	-0.02	0.002				
Standard Deviation (σ)		0.40				

* See equation 2

Table E Estimated distribution of RBC folate concentration (nmol/L) at neural tube closure among 228 456 Community Intervention Project (CIP) participants

RBC Folate Concentration (nmol/L)	Estimated Percentage %*	95% Uncertainty Interval†
0 - 400	0.0	0.0 to 6.5
401 - 500	6.9	1.2 to 7.5
501 - 600	2.3	1.5 to 2.9
601 - 700	2.5	1.8 to 3.4
701 - 800	1.3	0.0 to 2.7
801 - 900	54.2	0.0 to 55.2
901 - 1000	3.0	2.5 to 57.1
1001 - 1100	3.5	2.8 to 5.0
1101 - 1200	4.5	3.9 to 6.2
1201 - 1300	4.7	4.1 to 6.5
1301 - 1400	4.8	4.1 to 6.4
1401 - 1500	5.9	4.6 to 9.8
1501 +	5.2	0.0 to 8.8

The estimates were derived using the estimated parameters of the RBC folate concentration model with inputs reflecting the observed number of months of folic acid supplementation among CIP participants. Estimated time of closure of fetal neural tube was date of initiation of last reported menstrual cycle plus 42 days. Details on the model, underlying assumptions and methods used to develop thee estimates are provided in the supplemental material.

*This estimate is the median of posterior distribution of possible values for the percentage of CIP participants with RBC folate concentrations in the given range.

†The lower value of this interval is the 2.5th percentile of the posterior distribution of possible values for the percentage of CIP participants with estimated RBC folate concentrations in the given range and the upper value is the 97.5th percentile of that distribution.

Table F Posterior estimates of the parameters of the concentration, genotype and risk model based on data from the Community Intervention Projects.

	Parameter	Median	95% Uncertainty Interval
Concentration Model (natural log of RBC folate concentration nmol/L)			
	β_0	6.18	6.04 to 6.29
	β_1	-0.094	-0.14 to -0.035
	β_2	-0.24	-0.29 to -0.18
	β_3	0.20	0.16 to 0.23
	β_4	-0.019	-0.023 to -0.013
	β_South	0.69	0.60 to 0.80
	σ	0.14	0.10 to 0.21
Genotype Model (genotype frequency in populations)			
	South *CC*	0.39	0.35 to 0.43
	South *CT*	0.53	0.49 to 0.56
	South *TT*	0.08	0.06 to 0.10
	North *CC*	0.16	0.14 to 0.18
	North *CT*	0.47	0.44 to 0.50
	North *TT*	0.37	0.34 to 0.40
Risk Model (log Odds of NTD risk)			
	δ_0	4.57	2.45 to 6.64
	δ_1	-1.70	-2.01 to -1.38

Table G Estimated NTD risk per 10 000 births and 95% uncertainty intervals for the predicted NTD risk for various RBC folate concentrations among U.S. women

	Pre to Fortification		**Post to Fortification**							
	All women*		**All women***		**Pregnant Women All Trimesters Non to supplements users‡**		**Pregnant women All Trimesters supplement users‡**		**Pregnant Women 1st trimester‡§**	
Centiles	nmol/L	Median NTD risk† (95% UI)	nmol/L	Median NTD risk† (95% UI)	nmol/L	Median NTD risk† (95% UI)	nmol/L	Median NTD risk† (95% UI)	nmol/L	Median NTD risk† (95% UI)
5th	407	35.9 (28.1 to 46.2)	695	14.6 (12.4 to 17)						
10th	482	27.0 (21.9 to 33.4)	810	11.2 (9.6 to 13.1)	746	12.9 (11 to 15.1)	1229	5.6 (4.4 to 6.8)	982	8.1 (6.8 to 9.6)
25th	647	16.4 (13.9 to 19.3)	1050	7.2 (6.0 to 8.6)	1042	7.3 (6.1 to 8.8)	1683	>	1166	6.1 (4.9 to 7.4)
50th	904	9.4 (7.8 to 10.9)	1379	4.6 (3.5 to 5.8)	1278	5.2 (4.1 to 6.5)	2024	>	1550	>
75th	1277	5.2 (4.1 to 6.6)	1810	>	1682	>	2455		2029	>
90th	1785	>	2369	>	2375	>	3029	>	2561	>
Total		13.0 (10.1 to 16.4)		5.9 (4.2 to 7.8)		NC	NC	NC		NC

All RBC folate concentrations were normalized to the method used in the both the Daly and Hao paper using standardizing equations generated from Pfeiffer 2011 NHANES RBC folate [nmol/L] = (Dublin RBC folate [nmol/L] * 0.7876) + 34.2802 [nmol/L]-personal communication.

* RBC folate concentrations are from Pfeiffer et a. Journal of Nutrition 2012 –Supplemental Table 4 female participants 4 y and older during the pre-fortification period -NHANES 1988–1994 and Supplemental Table 6 Total female participants aged 4 years and older NHANES 2005-2010.[9]

† The estimates were derived using the estimated parameters of the NTD risk model. The presented estimate is the median of the posterior distribution of possible values for the NTD risk associated with the specified RBC folate concentration. Details on the model, underlying assumptions and methods used to develop the estimates are provided in the supplemental material. 95% uncertainty interval = 95% UI The lower values of this interval is the 2.5th percentile of the posterior distribution for possible values of the NTD risk associated with the specified RBC folate concentration and the upper value is the 97.5th percentile of that distribution.

‡ Adjusted RBC folate concentrations from Branum et al. 2013 –Table 5 [10]

§ Combination of all pregnant women consuming both fortified foods and ready to eat cereal in addition to dietary folate, but not supplements.

> Estimated NTD risks were not calculated for values >1500nmol/L as these are outside if the range of estimated RBC folate concentrations in the model.

NC= not calculated

Figure Legends

Fig A Sample selection Community Intervention Project (CIP)

Fig B Sample selection – Folic Acid Dosing Trial (FADT)

Fig C Observed and predicted RBC folate concentration (nmol/L) based on months consuming daily doses of 400 µg of folic acid by *MTHFR* 677 C to T genotype based on Folic Acid Dosing Trial (FADT) data

Fig D Markov Chain Monte Carlo history plots (red = chain 1, black = chain2) and Gelman Rubin Statistic plots (red = Gelman Rubin Statistic, blue = between chain variance, green = within chain variance) for the intercept, δ_0 and log odds ratio, δ_1, parameters in the assumed logistic regression relating estimated RBC folate concentration and NTD risk using data from the Community Intervention Project (CIP)

Fig E Impact of Alternative Priors Ninety-five percent posterior uncertainty intervals for estimated NTD risk (per 10 000) associated with folate concentrations between 500 and 1500 nmol/L

Fig F RBC folate mean concentrations from controlled trials with subjects receiving ~400 µg folic acid per day

Fig A Sample selection Community Intervention Project (CIP)

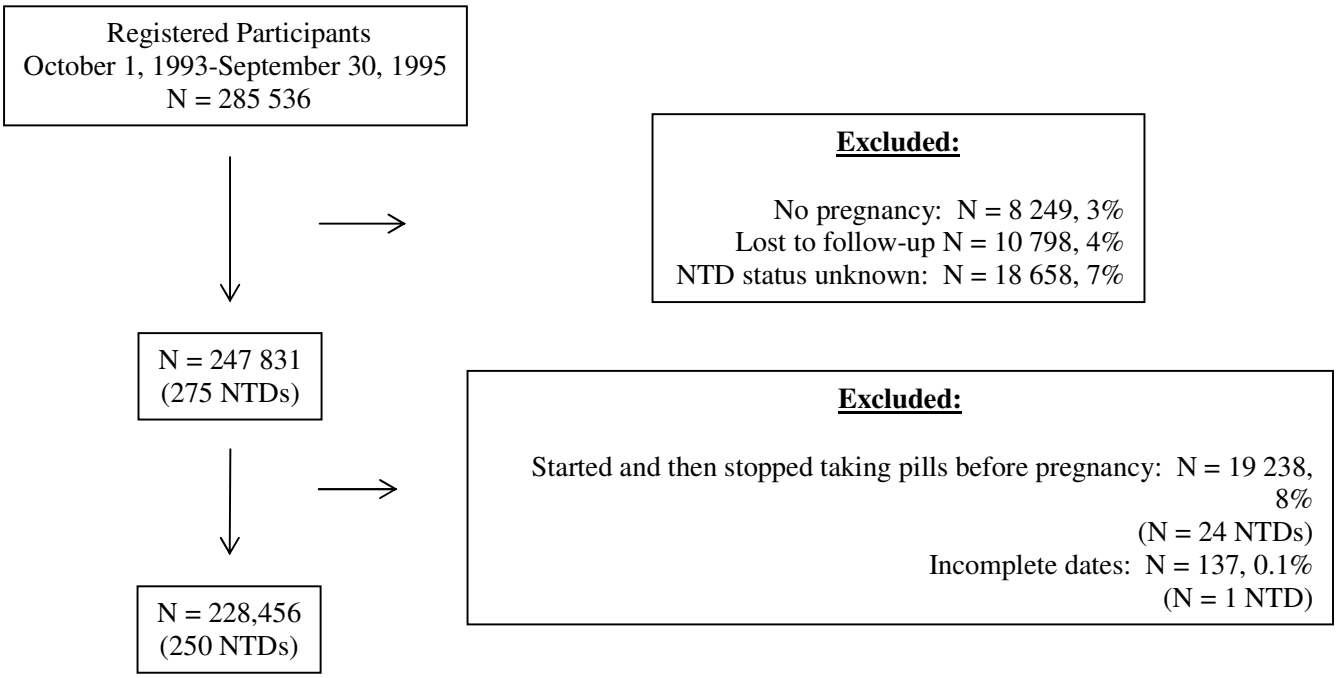
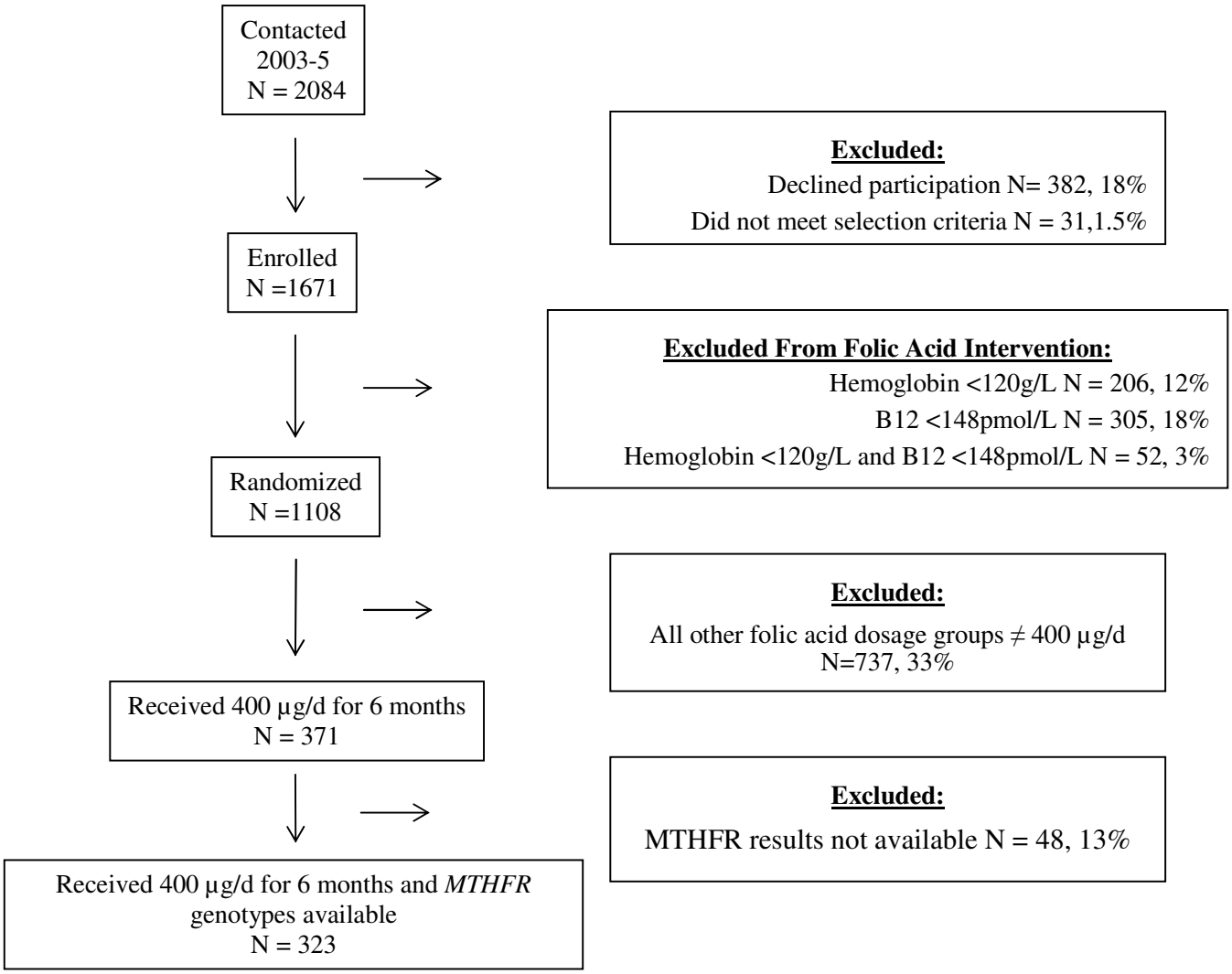


Fig B Sample selection – Folic Acid Dosing Trial (FADT)



Note: Baseline RBC folate concentrations (N = 1675) and *MTHFR 677* genotypes (N = 1194) for enrolled participants (including those later excluded and referred for treatment of anemia/B12 deficiency) were available and used as priors in some aspects of the modeling.

Fig C Observed and predicted RBC folate concentration (nmol/L) based on months consuming daily doses of 400 µg of folic acid by *MTHFR* 677 C to T genotype based on Folic Acid Dosing Trial (FADT) data

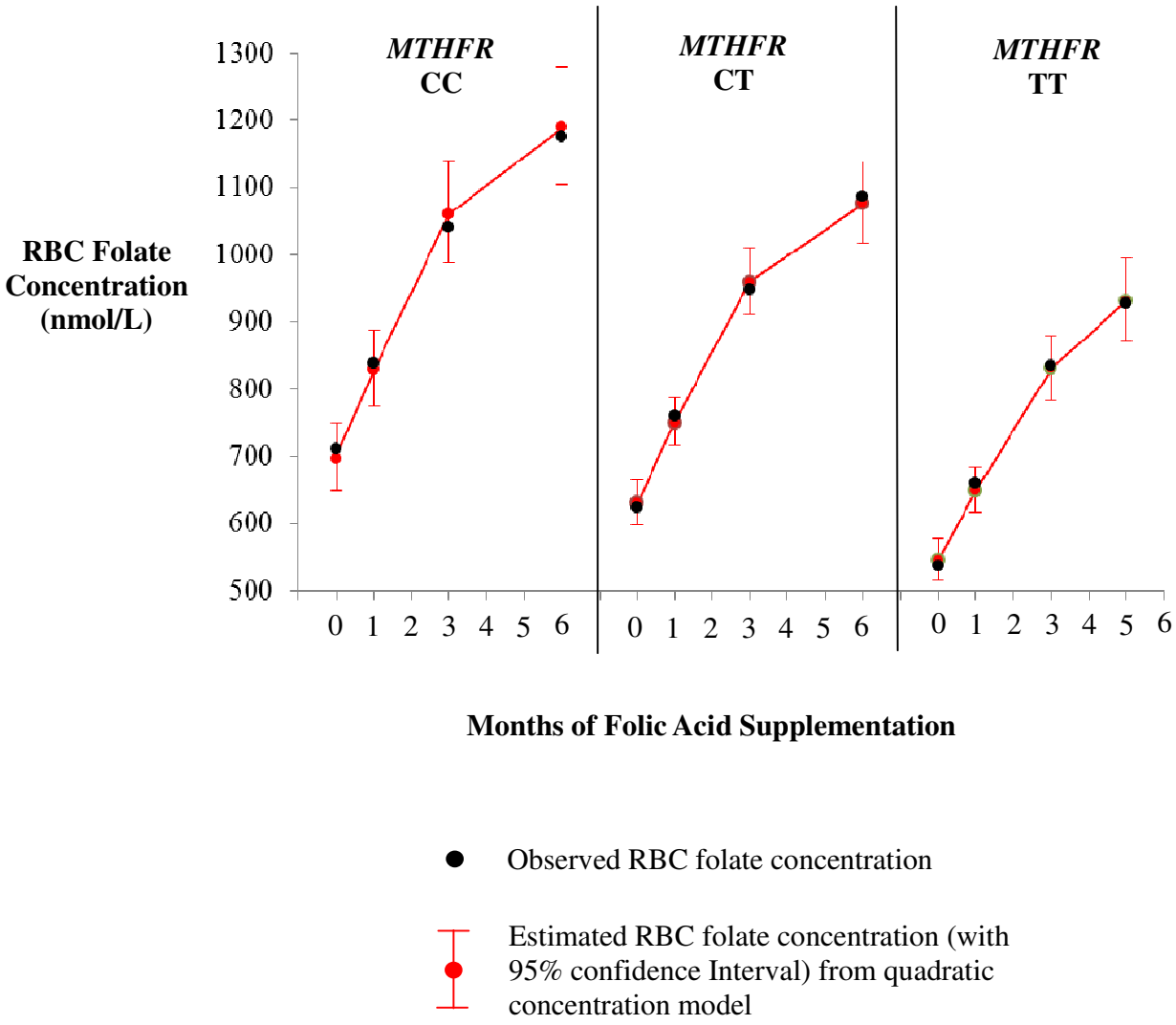


Fig D Markov Chain Monte Carlo history plots (red = chain 1, black = chain2) and Gelman Rubin Statistic plots (red = Gelman Rubin Statistic, blue = between chain variance, green = within chain variance) for the intercept, δ_0 and log odds ratio, δ_1, parameters in the assumed logistic regression relating estimated RBC folate concentration and NTD risk using data from the Community Intervention Project (CIP)

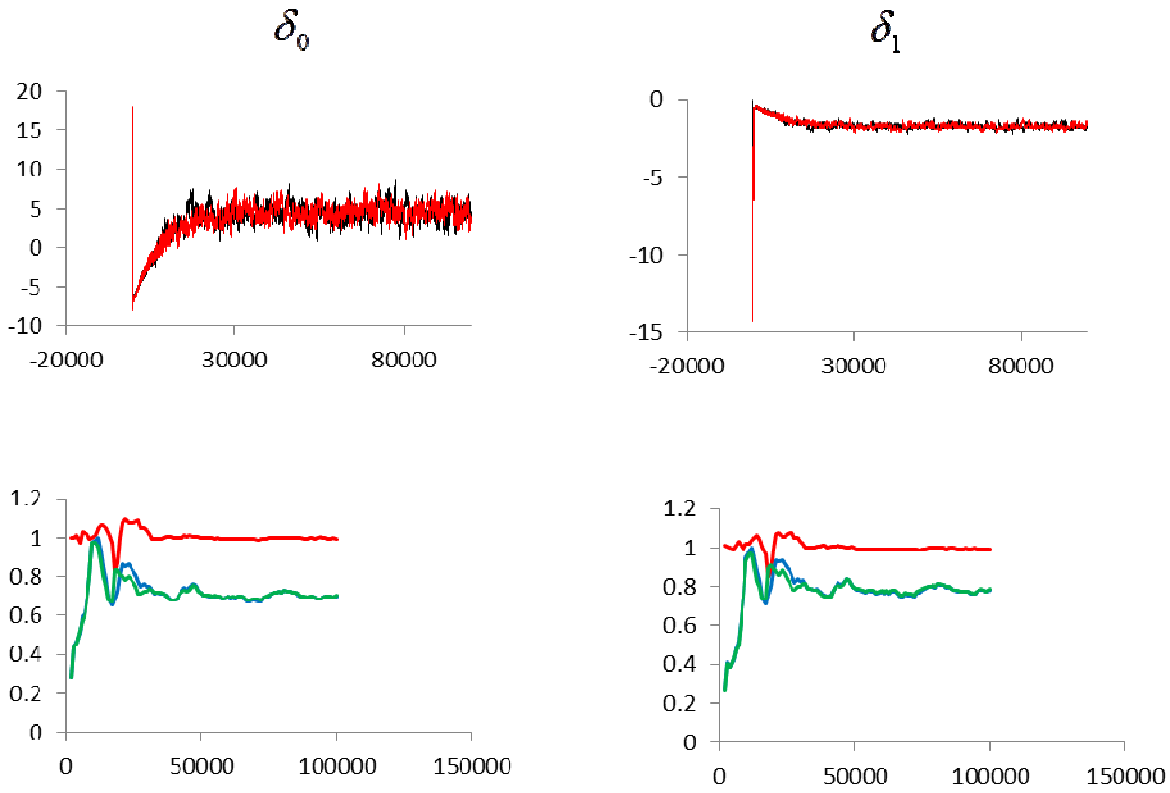


Fig E Impact of Alternative Priors Ninety-five percent posterior uncertainty intervals for estimated NTD risk (per 10 000) associated with folate concentrations between 500 and 1500 nmol/L

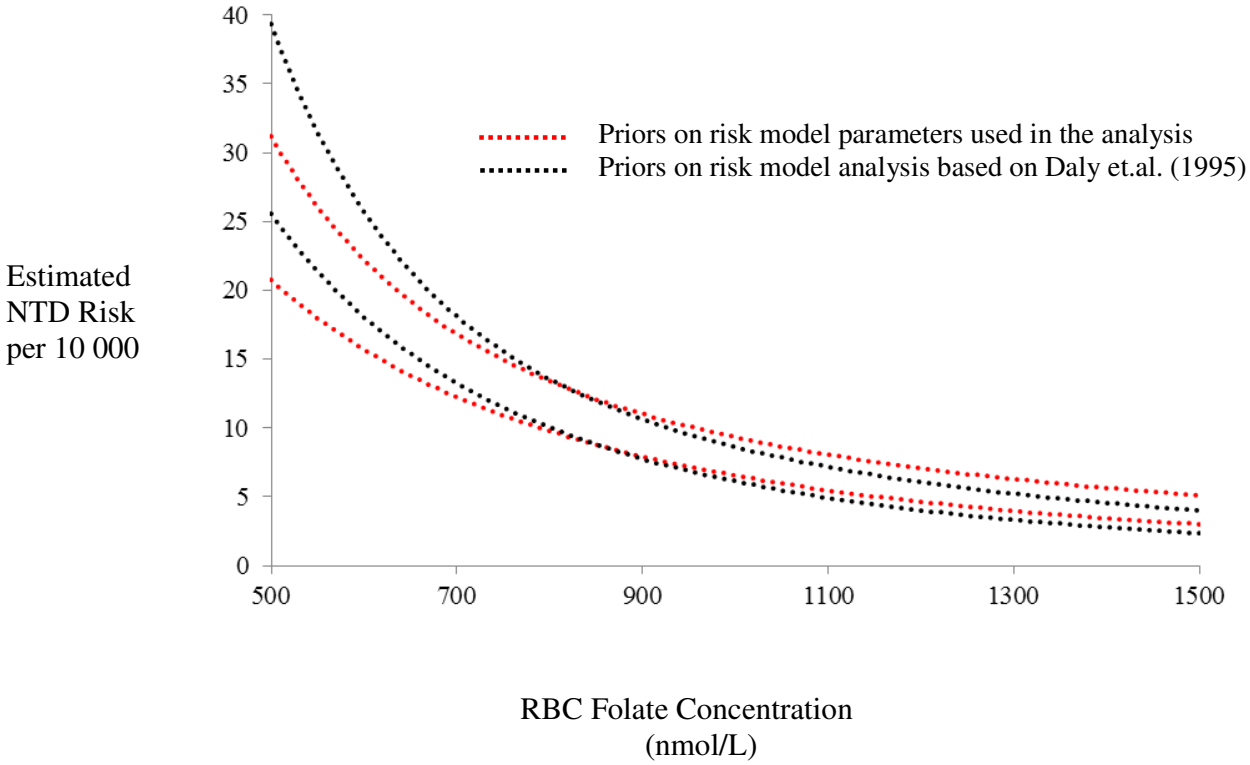


Fig F RBC folate mean concentrations from controlled trials with subjects receiving ~400 µg folic acid per day