Supplemental Table 1. Sample size for composite fat soluble biomarkers for the adult US population $\geq 20 \mathrm{y}$, NHANES 2003-2006 ${ }^{1,2}$

| Composite biomarker ( $n$ ) | Components | $n^{3}$ |
| :---: | :---: | :---: |
| CAR (4387) | alpha-carotene | 4436 |
|  | beta-carotene | 4440 |
|  | cis- and trans-lycopene | 4391 |
| XAN (4416) | lutein and zeaxanthin | 4440 |
|  | beta-cryptoxanthin | 4416 |
| SFA (1705) | myristic [14:0] | 1817 |
|  | palmitic [16:0] | 1826 |
|  | stearic [18:0] | 1827 |
|  | arachidic [20:0] | 1777 |
|  | docosanoic [22:0] | 1759 |
|  | lignoceric [24:0] | 1762 |
| MUFA (1681) | myristoleic [14:1n5] | 1829 |
|  | palmitoleic [16:1n7] | 1826 |
|  | cis-vaccenic [18:1n7] | 1781 |
|  | oleic [18:1n9] | 1819 |
|  | eicosenoic [20:1n9] | 1826 |
|  | nervonic [24:1n9] | 1717 |
| PUFA (1807) | linoleic [18:2n6] | 1827 |
|  | alpha-linolenic [18:3n3] | 1822 |
|  | gamma-linolenic [18:3n6] | 1816 |
|  | eicosadienoic [20:2n6] | 1826 |
|  | homo-gamma-linolenic [20:3n6] | 1827 |
|  | arachidonic [20:4n6] | 1828 |
|  | eicosapentaenoic [20:5n3] | 1827 |
|  | docosatetraenoic [22:4n6] | 1829 |
|  | docosapentaenoic-3 [22:5n3] | 1829 |
|  | docosapentaenoic-6 [22:5n6] | 1829 |
|  | docosahexaenoic [22:6n3] | 1829 |
| tFA (1459) | SFA | 1705 |
|  | MUFA | 1681 |
|  | PUFA | 1807 |
|  | docosenoic [22:1n9] | 1621 |

${ }^{1}$ CAR, carotenes [sum of alpha-carotene, beta-carotene and cis- and trans-lycopene]; MUFA, sum of 6 monounsaturated fatty acids [docosenoic [22:1n9] was excluded]; PUFA, sum of 11 polyunsaturated fatty acids; SFA, sum of 6 saturated fatty acids; tFA, total fatty acids [sum of 24
fatty acids, including 22:1n9]; XAN, xanthophylls [sum of lutein, zeaxanthin and betacryptoxanthin]
${ }^{2}$ Carotenes and xanthophylls (NHANES 2005-2006); plasma concentrations of total (free and esterified) saturated-, monounsaturated- and polyunsaturated fatty acids (NHANES 2003-2004)
${ }^{3} n$, number of non-missing values

Supplemental Table 2. Descriptive information for the adult US population $\geq 20$ y by sociodemographic and lifestyle variables ${ }^{1}$

| Factor | Category | Full sample 2003-2004 | Full sample 2005-2006 | Fatty acids fasting sample 2003-2004 ${ }^{8}$ | Phytoestrogen 1/3 sample 2003-2006 | lodine <br> 1/3 sample <br> 2003-2006 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Age, $y$ | 20-39 | 38.8 | 38.0 | 39.2 | 38.4 | 38.4 |
|  | 40-59 | 38.5 | 39.0 | 38.3 | 38.8 | 38.8 |
|  | $\geq 60$ | 22.7 | 23.0 | 22.5 | 22.8 | 22.8 |
| Sex | Male | 48.0 | 48.1 | 48.0 | 48.0 | 48.0 |
|  | Female | 52.1 | 51.9 | 52.0 | 52.0 | 52.0 |
| Race-ethnicity | Mexican American | 7.76 | 7.96 | 7.81 | 7.87 | 7.87 |
|  | Non-Hispanic black | 11.2 | 11.5 | 11.5 | 11.4 | 11.4 |
|  | Non-Hispanic white | 72.1 | 71.8 | 71.8 | 72.0 | 70.9 |
|  | Other Hispanic | 3.59 | 3.37 | 3.05 | 3.40 | 3.83 |
|  | Other (including multiracial) | 5.37 | 5.35 | 5.84 | 5.37 | 6.03 |
| Education | $\leq$ High school | 45.5 | 42.8 | 44.3 | 42.9 | 43.5 |
|  | >High school | 54.5 | 57.2 | 55.7 | 57.1 | 56.5 |
| PIR ${ }^{2}$ | Low (0-1.85) | 31.4 | 27.4 | 31.6 | 29.1 | 28.8 |
|  | Middle (>1.85-3.5) | 27.6 | 28.4 | 27.1 | 27.1 | 28.0 |
|  | High (>3.5) | 41.1 | 44.3 | 41.3 | 43.8 | 43.3 |
| Smoking status ${ }^{3}$ | No | 70.2 | 72.1 | 70.5 | 68.9 | 72.4 |
|  | Yes | 29.8 | 27.9 | 29.5 | 31.1 | 27.6 |
| Alcohol consumption ${ }^{4}$ | No drinks | 30.6 | 28.2 | 31.1 | 28.4 | 29.2 |
|  | <1 (not 0) | 56.6 | 57.0 | 56.8 | 56.7 | 56.3 |
|  | 1-<2 | 7.6 | 8.1 | 6.8 | 8.5 | 8.1 |


| Factor | Category | Full sample 2003-2004 | Full sample 2005-2006 | Fatty acids fasting sample 2003-2004 ${ }^{8}$ | Phytoestrogen 1/3 sample 2003-2006 | lodine 1/3 sample 2003-2006 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathrm{BMI}^{5}$ | $\geq 2$ | 5.2 | 6.7 | 5.4 | 5.4 | 6.4 |
|  | Underweight | 1.73 | 1.81 | NR | 2.12 | NR |
|  | Normal | 32.0 | 31.2 | 31.6 | 31.2 | 31.7 |
|  | Overweight | 34.1 | 32.8 | 34.1 | 32.8 | 33.2 |
| Supplement use ${ }^{6}$ | Obese | 32.2 | 34.3 | 32.8 | 33.9 | 33.6 |
|  | No | 46.0 | 45.9 | 46.2 | 46.9 | 44.9 |
|  | Yes | 54.0 | 54.2 | 53.8 | 53.1 | 55.1 |
| Physical activity ${ }^{7}$ | None reported | 33.0 | 31.3 | 33.2 | 30.3 | 33.7 |
|  | 0-<500 | 25.0 | 23.4 | 24.4 | 24.6 | 23.1 |
|  | 500-<1000 | 13.6 | 14.3 | 14.5 | 14.6 | 14.3 |
|  | $\geq 1000$ | 28.4 | 31.1 | 27.9 | 30.6 | 29.0 |

${ }^{1}$ Values represent weighted percentage (\%) by various mobile examination center weights
${ }^{2}$ PIR, family poverty income ratio
3 "Smoker" defined by serum cotinine concentration $>10 \mu \mathrm{~g} / \mathrm{L}$
${ }^{4}$ Alcohol consumption: calculated as average daily number of "standard" drinks [(quantity $x$ frequency) / 365.25)]; 1 drink $\sim 15 \mathrm{~g}$ ethanol
${ }^{5} \mathrm{BMI}\left(\mathrm{kg} / \mathrm{m}^{2}\right.$ ) definitions: underweight: <18.5; normal weight: $18.5->25$; overweight: $25-<30$; and obese: $\geq 30$
6 "Supplement user" defined as participant who reported taking a dietary supplement within the past 30 d
${ }^{7}$ Physical activity: calculated as total metabolic equivalent task (MET)-min/wk from self-reported leisure time physical activities
${ }^{8}$ Plasma concentrations of FA were measured in fasted ( $\geq 8 \mathrm{~h}$ ) adults
NR: not reported due to small sample size ( $n<42$ )

## Supplemental Text 1

The use of a logarithmic transformation to a response variable in linear regression provides a natural interpretation of the response as a percent change. When the response has been log transformed, one can compute the percent change in the response at two different values of a covariate, while holding all others constant. Recognizing that the effects of regression model estimates can be interpreted as the difference between a pair of fitted or predicted values provides a scheme to organize and present model results that can be easily compared across nutritional biomarkers, as well as assessing the impact on the estimated associations after controlling simultaneously for many variables.

Consider the difference between a pair of fitted values of $Y_{1}$ and $Y_{2}$. It is easy to show that the difference between this pair of fitted values can be interpreted as the percent change between $Y_{1}$ and $Y_{2}$ log transformed. The formulation below assumes a natural log transformation has been used.

Start with the hypothetical model $\operatorname{In} Y=\beta_{0}+\beta_{1} X+\beta_{1} W+\beta_{1} Z$. Compute the difference for each fitted changing the value of the variable X from $\mathrm{k}_{1}$ to $\mathrm{k}_{2}$, while holding all other covariates the same:

$$
\ln Y_{2}-\ln Y_{1}=\left(\beta_{0}+\beta_{1} k_{2}+\beta_{1} W+\beta_{1} Z\right)-\left(\beta_{0}+\beta_{1} k_{1}+\beta_{1} W+\beta_{1} Z\right)
$$

After some cancellation it is easy to see the following:

$$
\begin{align*}
& \ln Y_{2}-\ln Y_{1}=\beta_{1}\left(k_{2}-k_{1}\right) \\
& \ln \frac{Y_{2}}{Y_{1}}=\beta_{1}\left(k_{2}-k_{1}\right) \\
& \frac{Y_{2}}{Y_{1}}=e^{\beta_{1}\left(k_{2}-k_{1}\right)}  \tag{1}\\
& 100 \times\left(\frac{Y_{2}}{Y_{1}}-1\right)=100 \times\left(e^{\beta_{1}\left(k_{2}-k_{1}\right)}-1\right) .
\end{align*}
$$

## Supplemental Text 2

Twenty of the analytes use a natural log transformation, so we can approximately interpret these values as the percent change in the response for 1 unit change in the covariate, or in the case of a categorical variable as a percent change in the response comparing 1 category to a selected baseline category. As an example, consider serum folate and the covariate sex. The beta coefficient, with males as the reference, from model 1 (simple linear regression) is 0.129 (95\% CI: 0.0.104-0.154). Using the approximate interpretation of a natural log transformation, this suggests that females have approximately $12.9 \%(0.129 \times 100)$ higher serum folate levels than males. This interpretation can be made exact using the formula: $100 \times\left(e^{0.129}-1\right)=13.8$.

