Estimated Fish Consumption Rates for the U.S. Population and Selected Subpopulations (NHANES 2003-2010)

Appendix C

Supplemental Statistical Methodology
The following provides justification for the assumption that the predicted logit ($B_{ij}$) when including
the random effect is proportional to the predicted logit when excluding the random effect ($B'_{ij}$).

Assume that $C_{ij}$ are realizations from a binomial distribution with probability $P_{ij}$:

$$C_{ij} \sim Binomial(1, P_{ij})$$

The values of $P_{ij}$ may vary among individuals and be different for the first and second recalls or $P_{ij}
may be the same for all individuals and recalls. Regardless of what assumptions are made about $P_{ij},
a basic fact about logistic regression is that when fitting a logistic regression model without random
effects or other independent predictors, i.e., fitting the model:

$$Logit(P_{ij}) = \pi'$$

the intercept parameter is equal to:

$$\pi' = Logit \left( Mean(C_{ij}) \right).$$

Since

$$E \left( Mean(C_{ij}) \right) = E(C_{ij}) = E(P_{ij}),$$

the intercept is approximately:

$$\pi' \approx Logit \left( E(P_{ij}) \right).$$

We can assume a model for the $P_{ij}$. For simplicity, assume the probability of consuming fish can be
modeled using logistic regression with an intercept, no other predictors, and a random person-
specific effect having a normal distribution on the logit scale, i.e.,

$$Logit(P_{ij}) = log \left( \frac{P_{ij}}{1 - P_{ij}} \right) = \pi_0 + \pi_i$$

$$\pi_i \sim Normal(0, \sigma^2).$$

We can define the following ratio:

$$\beta = \frac{\pi_0}{\pi'}$$

and, with an estimate of $\beta$ can calculate $\pi_0$ as:

$$\pi_0 = \beta \pi'_0$$
When the logistic regression model has additional predictors, the predicted logit $B'_{ij}$ replaces $\pi_0'$ and the EPA method assumes $\beta$ is reasonably constant over the range of values of $B'_{ij}$.

The following is a heuristic argument for why this assumption is reasonable.

First, let $P_0 = \text{logistic}(\pi_0)$. If $P_{ij}$ is the same for all 24-hour recalls (i.e., $P_{ij} = P_0$ and $\sigma_1^2 = 0$) then $\pi_0 = \pi_0'$ and $\beta = 1$ for all values of $\pi_0$.

If $\pi_0 = 0$, the expected probability of fish consumption is $E(P_{ij}) = 0.50$, regardless of whether $\sigma_1^2 = 0$, i.e., $\pi_0 = \pi_0' = 0$. If $\sigma_1^2 = 0$, all individuals have the same probability of fish consumption and $E(P_{ij}) = P_0$. If $\sigma_1^2 > 0$, some people have a higher probability and some have a lower probability of fish consumption; however, since the logistic function is symmetric around $\pi_0 = 0$, these probabilities balance out and the average probability fish consumption is $E(P_{ij}) = E(\text{logistic}(\pi_0 + \pi_i)) = E(\text{logistic}(\pi_i)) = 0.50$. In the case where $\pi_0 = 0$, $\beta = 0/0$ which is not defined. However, if $\beta$ is used to define $\pi_0$ using $\pi_0 = \beta \pi_0'$, then any value of $\beta$ can be used since $\pi_0' = 0$.

Because the logit function is nonlinear and $P_{ij}$ is limited on the high side (i.e., $P_{ij} \leq 1$), if $\pi_0 > 0$, $0.5 < E(P_{ij}) < P_0$, $0 < \pi_0' < \pi_0$, and $\beta > 1$. Since $P_{ij}$ is also limited on the low side (i.e., $P_{ij} \geq 0$), if $\pi_0 < 0$, $0.5 > E(P_{ij}) > P_0$, $0 > \pi_0' > \pi_0$, but because both $\pi_0$ and $\pi_0'$ are negative, the ratio is still positive, i.e., $\beta > 1$. $\beta$ is the same for $\pi_0$ and $-\pi_0$. As $\pi_0$ increases in absolute magnitude, the non-linearity of the logit function increases. As a result, the difference between $\pi_0$ and $\pi_0'$ increases. The EPA method assumes the ratio, $\beta$, is relatively constant.

The following provide numerical estimates of $\beta$, illustrating the $\beta$ is reasonably constant for different values of $\pi_0$ or $\pi_0'$.

Given $\pi_0$ and the variance of the random effect ($\sigma_1^2$), we used numerical integration to calculate $\beta$:

$$P_{ij} = \text{logistic}(\pi_0 + \pi_i)$$

$$\beta \approx \frac{\pi_0}{\text{Logit}(E(P_{ij}))}$$

Table C-1 shows $\beta$ as a function of $\pi_0$ and $\sigma_1^2$ calculated using numerical integration.
<table>
<thead>
<tr>
<th>$\sigma_1^2$</th>
<th>$-2$</th>
<th>-1</th>
<th>0.001</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>1.047</td>
<td>1.056</td>
<td>1.060</td>
<td>1.056</td>
<td>1.047</td>
</tr>
<tr>
<td>0.50</td>
<td>1.093</td>
<td>1.107</td>
<td>1.114</td>
<td>1.107</td>
<td>1.093</td>
</tr>
<tr>
<td>0.75</td>
<td>1.138</td>
<td>1.156</td>
<td>1.164</td>
<td>1.156</td>
<td>1.138</td>
</tr>
<tr>
<td>1.00</td>
<td>1.182</td>
<td>1.202</td>
<td>1.212</td>
<td>1.202</td>
<td>1.182</td>
</tr>
<tr>
<td>1.25</td>
<td>1.224</td>
<td>1.246</td>
<td>1.256</td>
<td>1.246</td>
<td>1.224</td>
</tr>
</tbody>
</table>

In Table C-1, $\beta$ is greater than or equal to 1.0, relatively constant across rows corresponding to different values of $\pi_0$ for the same $\sigma_1^2$ and increases within increasing $\sigma_1^2$. The EPA method does not require that $\beta$ be constant across all possible values of $\pi_0$, but reasonably constant across values of $B_{ij}'.$