Appendix D: Sampling Estimators Used for the Hispanic Farm Operator Childhood Agricultural Injury Survey

1. Equations to derive the non-benchmarked estimates, y, and variances, $v(y)$, for injury, asthma, or youth for Hispanic farms.

$$y = \sum_{h=1}^{9}\left[\sum_{i=1}^{n_h}\frac{N_h}{n_h}y_{hi}\right] = \sum_{h=1}^{9} y_h \qquad \text{(D.1)}$$

$$v(y) = \sum_{h=1}^{9}\frac{N_h - n_h}{N_h}\sum_{i=1}^{n_h}\frac{\left(N_h y_{hi} - y_h\right)^2}{n_h(n_h - 1)} = \sum_{h=1}^{9} v(y_h) \qquad \text{(D.2)}$$

where:

h = regional strata used in post-stratification;

N_h = number of farms in region h from the NASS sampling list;

n_h = number of respondents in region h from the survey;

y_{hi} = value of the variable of interest (i.e., injury, asthma, youth) on farm i in region h from the survey;

y_h = estimate of variable of interest (i.e., injury, asthma, youth) in region h;

$v(y_h)$ = sampling variance for the variable of interest (i.e., injury, asthma, youth) in region h.

2. Equations to derive the benchmarked national estimates, $y_{(bm)}$, and variances, $v(y_{(bm)})$, for injury, asthma, or youth for Hispanic farms.

$$y_{(bm)} = \sum_{h=1}^{9}\frac{N_{(bm)h}}{N_h}y_h \qquad \text{(D.3)}$$

$$v(y_{(bm)}) = \sum_{h=1}^{9}\left(\frac{N_{(bm)h}}{N_h}\right)^2 v(y_h) \qquad \text{(D.4)}$$

where non-benchmarked values are as previously defined in D.1 and D.2, and:

$N_{(bm)\,h}$ = number of farms in region h from the published 1997 Census of Agriculture.

Note: For equations D.3 and D.4, each independent outcome of h provides the corresponding benchmarked regional estimate for the variable of interest and its variance.

3. Equations to derive the benchmarked national injury incidence or asthma prevalence rate estimates, R, and variances, $v(R)$, for Hispanic farms.

$$R = 1000\left(\frac{y_{(bm)}}{x_{(bm)}}\right) \tag{D.5}$$

$$v(R) = \frac{1000^2}{n}\left(\frac{1}{\bar{x}_{(bm)}}\right)^2\left[v(y_{(bm)}) + R^2 v(x_{(bm)}) - 2R\,\mathrm{cov}(y_{(bm)}, x_{(bm)})\right] \tag{D.6}$$

where:

$y_{(bm)}$ = benchmarked national estimate for injury or asthma from the survey;

$x_{(bm)}$ = benchmarked national estimate for youth at risk from the survey;

$\bar{x}_{(bm)}$ = benchmarked national average of youth at risk per farm from the survey;

$v(y_{(bm)})$ = benchmarked variance for the national injury or asthma estimate from the survey;

$v(x_{(bm)})$ = benchmarked variance for the national estimate of youth at risk from the survey;

$\mathrm{cov}(y_{(bm)},x_{(bm)})$ = covariance between the benchmarked injury or asthma estimate and the benchmarked estimate of youth at risk from the survey;

n = the number of farms from the NASS sampling frame used to derive the covariance between $y_{(bm)}$ and $x_{(bm)}$.

Alternatively, the variance for R can be determined by:

$$v(R) = 1000^2 R^2\left[\left(\frac{\sqrt{v(y_{(bm)})}}{y_{(bm)}}\right)^2 + \left(\frac{\sqrt{v(x_{(bm)})}}{x_{(bm)}}\right)^2 - 2\left(\frac{\mathrm{cov}(y_{(bm)}, x_{(bm)})}{\bar{y}_{(bm)}\bar{x}_{(bm)}n}\right)\right] \tag{D.6.1}$$

where:

$\frac{\sqrt{v(y_{(bm)})}}{y_{(bm)}}$ = relative standard error for $y_{(bm)}$ (i.e., injury or asthma);

$\frac{\sqrt{v(x_{(bm)})}}{x_{(bm)}}$ = relative standard error for $x_{(bm)}$ (i.e., youth at risk);

$\frac{\mathrm{cov}(y_{(bm)}, x_{(bm)})}{\bar{y}_{(bm)}\bar{x}_{(bm)}n}$ = relative covariance between $y_{(bm)}$ and $x_{(bm)}$;

$\bar{y}_{(bm)}$ = benchmarked national average for injury or asthma per farm from the survey;

n = the number of farms from the NASS sampling frame used to derive the covariance between $y_{(bm)}$ and $x_{(bm)}$.

Because the relative covariance between the estimated number of injury or asthma cases and the number of youth at risk is typically negligible, the estimate $v(R)$ may be approximated as:

$$v(R) = 1000^2 R^2 \left[\left(\frac{\sqrt{v(y_{(bm)})}}{y_{(bm)}} \right)^2 + \left(\frac{\sqrt{v(x_{(bm)})}}{x_{(bm)}} \right)^2 \right] \quad \text{(D.6.2)}$$

Equation D.6.2 was the method used to derive the standard errors for all rates in this document.