Attachment I – Method to Adjust for Non-Response Bias

We will use a logistic regression model of the probability that an SFA or CN agency responded to the survey to estimate the probability of survey response and to construct appropriate weights for each respondent. The initial logistic regression model will be estimated using all SFA characteristics. The model can be expressed by the following equation:

P¿

The dependent variable in this model (S) is the likelihood that the SFA responds to the survey. S equals 1 if the SFA responds to the survey and 0 otherwise. The initial model includes all available SFA or CN agency characteristics (X) in the VSR database. Inclusion of all characteristics in the model is important for identifying which variables are strong predictors of non-response. To identify the final set of covariates for the non-response model (i.e., the variables that are strong predictors of non-response), we will use the following measures of predictive ability and goodness of fit:

- ***McFadden's Pseudo R-squared statistic*** – This measure captures the percent of the variation in the likelihood of responding to the survey that is explained by SFA and state CN agency characteristics. This statistic is constructed as follows:[1]

$$R^2 = 1 - \frac{\ln\left(\hat{L}\left(M_{Full}\right)\right)}{\ln\left(\hat{L}\left(M_{Intercept}\right)\right)}$$

- ***Akaike's Information Criterion (Akaike 1974)*** – This statistic measures the efficiency of the model in predicting the outcome based on the number of covariates. This statistic is constructed as follows:[2]

1 Note: $\hat{L}$ is the full model prediction using all control variables; M_{Full} is the full model prediction without predictors; and $M_{Intercept}$ is the estimated likelihood of response based on the logit model.

2 Note: k is the number of parameters in the estimated model; and $\hat{L}_k$ is the maximum value of the likelihood function based on the estimated model.

$$AIC = 2k - 2\ln\left(\hat{L}_k\right)$$

We will use these statistics to assess which set of observable characteristics are strong predictors of survey non-response. After identifying this set of characteristics, we will estimate the final version of the model using these characteristics. After estimating this model, we will use the results to calculate the predicted likelihood of survey response based on each SFA/State CN agency characteristic, as follows:

$$w_i = \frac{1}{f\left(X \cdot \hat{b}\right)}$$

In words, the survey non-response weight for each SFA is the inverse of the predicted probability of response, $f\left(X \cdot \hat{b}\right)$. We use these weights in the analyses to make the sample representative of all SFAs in the seven FNS regions. This method is a widely accepted practice in program evaluations for controlling for survey non-response and for making estimation results representative of all program applicants.[3]

We will also conduct sensitivity analyses using alternative methods to address non-response bias, such as case wise deletion and dummy adjustment. The former involves deleting the entire survey response record if dependent variables have missing data. The latter requires creating a dummy variable to indicate the missing status (0 if not missing data; 1 if missing data) when independent variables in the analytic model have missing data. We will then incorporate the dummy missing indicator and the independent variables in the analytic model as predictors.

3 Source: McConnell et al., 2006; Trenholm et al., 2007; Benus et al., 2009)