APPENDIX A.
DATA WEIGHTING METHODS

Surveys rarely obtain information from everyone within the population of interest. As a result, it is necessary to differentially weight survey respondents to account for both the sampling frame (the probability of being selected into the sample) and for nonresponse among those who are sampled. In a survey that achieves less than a 100 percent response rate there is a risk that respondents may be systematically different from nonrespondents. Such differences would imply that the respondents should not be regarded as a random subsample of the full survey sample. If the survey data are not adjusted to mitigate these differences, such as by differentially weighting the survey respondents, it may be inappropriate to draw inferences about the sampling frame from statistics computed on the basis of the survey data.

This appendix describes the procedures that were used to develop weights for the Temporary Assistance for Needy Families (TANF) Caseload Survey data from Illinois, the District of Columbia, Maryland, and South Carolina. We followed three basic steps to create the weights. First, we created sampling adjustments to account for the number of TANF recipients in the full population represented by each individual in the survey sample. Second, we adjusted for potential non-response bias. And, third, we constructed post-stratification adjustments to ensure that the sum of the weighted survey respondents equals the sampling frame counts. While the steps were similar, the specific process employed within each step was different for Illinois than for the other three sites. We first describe the weighting procedures in Illinois separately and then describe the weighting procedures in the District of Columbia, Maryland, and South Carolina together. Table A.1 provides important details on the study populations, sample sizes and designs, and response rates for each of the sites.

A. Illinois

Component 1: Sampling weight

The sample weight for each case in the survey sample accounts for the number of cases it represents in the sampling frame, based on the sample selection procedure. The sample weight is the inverse of the actual probability of selection. The Illinois survey used a stratified sample with two strata, Cook County and downstate. Cook County was sampled at a marginally higher rate than downstate. For sampled cases in Cook County and downstate, respectively, the sampling weight is 62.93 and 63.09.

Component 2: Non-response adjustment factor

This component compensates for the reduction in the sample due to cases that could not be interviewed. It is the inverse of the survey response rate. Because the

response rate was slightly lower in Cook County than downstate, the value of this component is slightly higher in Cook County (1.28) than downstate (1.25).

TABLE A.1. Summary Table for the TANF Caseload Surveys

State/Site	Description and Size of the Study Population	Sample Size and Design	Number of Completed Interviews	Survey Response Rate	Survey Fielding Period
District of Columbia	11,918 single-parent TANF cases as of August 15, 2002.	581 cases randomly selected from the full study population.	420 of the 581 sample members	72.3%	Mid-September 2002 - Mid-November 2002
Illinois	33,495 single-parent cases that were authorized to receive a TANF grant in November 2001.	532 cases randomly selected from the full study population within two strata--residence in Cook County or in the rest of the state.	416 of the 532 sample members	78.2%	November 19, 2001 - March 3, 2002
Maryland	15,867 single-adult TANF cases with at least one recipient child in June 2002.	1,146 cases drawn from the full study population: half randomly selected from Baltimore City and half randomly selected from the rest of the state.	819 of the 1,146 sample members	71.5%	August 19, 2002 - October 31, 2002
South Carolina	11,002 single-parent TANF cases in May 2002.	1,493 cases drawn from the full study population within four strata: • Households with less than 24 months of benefits and head of household subject to work requirements (randomly selected) • Households in which the head was exempt from work requirements (randomly selected) • Households where the status of the household head could not be determined (randomly selected) • Households that had exhausted the 24 months of TANF benefits, but which had been granted an extension of benefits ("take-all" stratum)	1,120 of the 1,493 sample members	75.0%	August 2002 - November 2002

Component 3: Post-stratification adjustment

This component of the survey weights is based on a post-stratification of the survey respondents into five cells as shown in Table A.2. This factor causes the sum of the weighted survey respondents to equal the number of cases in the sampling frame in each cell. The five cells were defined by three variables that were extracted from the Illinois DHS administrative data system in November 2001: residence in Cook County or downstate, the grantee's age less than or equal to 28 years or greater than 28 years, and a zero or positive TANF benefit amount. While in principle, these variables could be used to define eight cells, the infrequency of zero-benefit cases led us to consolidate them in a single cell. The values of the post-stratification adjustment factor range from 0.86 to 1.25. In general, the larger values are for cells containing cases with older grantees, who had a lower survey response rate than cases with younger grantees.

TABLE A.2. Post-Stratification Adjustment					
	Zero TANF Benefit	Positive TANF Benefit			
		Cook County		Downstate	
		≤28 Years Old	>28 Years Old	≤28 Years Old	>28 Years Old
Number of Survey Respondents	30	184	129	48	25
Weighted Number of Survey Respondents	2,406	14,859	10,417	3,776	1,967
Number of Cases in Sampling Frame	2,859	12,728	12,074	3,378	2,456
Adjustment Factor	1.19	0.86	1.16	0.89	1.25

Final survey weights

The final weights for the survey respondents are the product of the three components discussed above. There is a unique weight for each of six cells, ranging in value from 69.17 to 98.24.[1]

B. District of Columbia, Maryland, and South Carolina

Component 1: Sampling weight

The sample weight for each case in the survey sample accounts for the number of cases it represents in the sampling frame, based on the sample selection procedure. The sample weight is the inverse of the actual probability of selection. DC had one strata; Maryland had two strata, Baltimore and the rest of the state; South Carolina had four strata--those subject to work requirements and under the 24-month time limit on benefits (mandated population), those exempt from work requirements (exempt population), those who have been granted an extension to the 24-month time limit (extended population), and those whose status could not be determined (unknown). Table A.3 presents the sampling weights for each site by strata.

[1] The six cells and their associated survey weights are: (1) zero TANF benefit, Cook County, 95.96; (2) zero TANF benefit, downstate, 93.48; (3) positive TANF benefit, Cook County, less than or equal to age 28, 69.17; (4) positive TANF benefit, Cook County, greater than age 28, 93.60; (5) positive TANF benefit, downstate, less than or equal to age 28, 70.38; (6) positive TANF benefit, downstate greater than age 28, 98.24.

TABLE A.3. Sampling Weights	
	Sampling Weight
District of Columbia	20.48
Maryland	
Baltimore	17.86
Rest of State	9.83
South Carolina	
Mandated Population	14.01
Exempt Population	3.77
Extended Population	1.00
Unknown	3.72

Component 2: Non-response adjustment factor (NAF)

A model predicting response was estimated, and respondents and nonrespondents were grouped by their predicted probability of being a respondent. The respondents were then weighted within these groups ("cells"), by the inverse of the overall response probability estimated in each cell. This process was conducted separately for each of the three sites (the District of Columbia, Maryland, South Carolina).

The first step was to estimate a model of response as a function of the covariates observed for all individuals in the sample. For each site, a model selection process was used to determine the variables that are the most predictive of whether someone was a respondent or nonrespondent. Variables were chosen using a model selection process involving univariate logistic regressions and a backward stepwise selection model.[2]

For the District of Columbia, the model of response included indicators for sanction at time of sample, receiving food stamps at time of sample, youngest child between 1 and 5 years of age, age younger than 25, and age between 25 and 34 years. The probability of response in Maryland was modeled using age, total amount of TANF grant received in June 2002, and an indicator for having two unique Maryland UI-covered employments in the second quarter of 2002. The South Carolina model of response included indicators for those between 28 and 34 years of age, food stamp receipt in December of 2002, and TANF receipt in October of 2002.

After the final model predicting response was selected for each site, the response cells were formed using the quantiles of the distribution of the predicted probabilities of response; the first cell contains individuals with the lowest predicted probabilities of response, and the last cell contains individuals with the highest predicted probabilities of response. The number of cells formed for each site was five or six, depending on the sample size for the site and the resulting adjustment weights. The non-response adjustment factor (NAF) is the inverse of the response rate in a cell, where the response

[2] Variables with substantial missing values were excluded from the analysis; missing values for variables with just a few missing values were imputed solely for the purpose of constructing these non-response weighting adjustments.

rate is estimated as the number of respondents in that cell divided by the number of sampled individuals in that cell. Table A.4 presents the NAF for each cell by site.

Component 3: Post-stratification adjustment

A final poststratification adjustment was completed to ensure that the sum of the weighted survey respondents equaled the frame counts within the cells defined by covariates for each site (as determined in component 2). Unfortunately, full frame data were not available for all sites; limited frame data were available for the District of Columbia, and only the sampling strata counts and proportions were available for Maryland and South Carolina. In addition, the sampling strata for non-respondents were not available for use in constructing the South Carolina non-response weighting adjustments. Therefore, the post-stratification adjustments for the District of Columbia, South Carolina and Maryland are based on the sampling strata counts and proportions. These adjustments, presented in Table A.5, ensure that the sum of weighted respondents in each sampling strata equaled the sampling strata frame counts.

TABLE A.4. Non-response Adjustment Factors

	Response Cell	Respondents	Sampled	Response Rate	NAF
District of Columbia	1	52	88	0.59	1.69
	2	95	140	0.68	1.47
	3	49	67	0.73	1.37
	4	106	142	0.75	1.34
	5	118	145	0.81	1.23
Maryland	1	142	227	0.63	1.60
	2	156	224	0.70	1.44
	3	169	236	0.72	1.40
	4	174	229	0.76	1.32
	5	178	230	0.77	1.29
South Carolina	1	86	155	0.55	1.80
	2	78	116	0.67	1.49
	3	177	250	0.71	1.41
	4	100	131	0.76	1.31
	5	460	579	0.79	1.26
	6	219	262	0.84	1.20

TABLE A.5. Post-stratification Adjustments

	Post-stratification adjustment
District of Columbia	1.00
Maryland	
Baltimore	1.01
Rest of State	0.99
South Carolina	
Mandated Population	1.00
Exempt Population	1.00
Extended Population	1.03
Unknown	0.93

TABLE A.6. Final Weights						
	Strata	Sampling Weight	Response Cell	NAF	Post-stratification adjustment	Final Weight
District of Columbia	1	20.48	1	1.69	1.00	34.66
	1	20.48	2	1.47	1.00	30.18
	1	20.48	3	1.37	1.00	28.00
	1	20.48	4	1.34	1.00	27.43
	1	20.48	5	1.23	1.00	25.16
Maryland	1	17.86	1	1.60	1.01	28.96
	1	17.86	2	1.44	1.01	26.01
	1	17.86	3	1.40	1.01	25.30
	1	17.86	4	1.32	1.01	23.84
	1	17.86	5	1.29	1.01	23.41
	2	9.83	1	1.60	0.99	15.50
	2	9.83	2	1.44	0.99	13.92
	2	9.83	3	1.40	0.99	13.54
	2	9.83	4	1.32	0.99	12.76
	2	9.83	5	1.29	0.99	12.53
South Carolina	1	14.01	1	1.80	1.00	25.12
	1	14.01	2	1.49	1.00	20.73
	1	14.01	3	1.41	1.00	19.69
	1	14.01	4	1.31	1.00	18.26
	1	14.01	5	1.26	1.00	17.55
	1	14.01	6	1.20	1.00	16.68
	2	3.77	1	1.80	1.00	6.81
	2	3.77	2	1.49	1.00	5.62
	2	3.77	3	1.41	1.00	5.34
	2	3.77	4	1.31	1.00	4.95
	2	3.77	5	1.26	1.00	4.76
	2	3.77	6	1.20	1.00	4.52
	3	1.00	1	1.80	1.03	1.86
	3	1.00	2	1.49	1.03	1.53
	3	1.00	3	1.41	1.03	1.46
	3	1.00	4	1.31	1.03	1.35
	3	1.00	5	1.26	1.03	1.30
	3	1.00	6	1.20	1.03	1.23
	4	3.72	1	1.80	0.93	6.26
	4	3.72	2	1.49	0.93	5.16
	4	3.72	3	1.41	0.93	4.90
	4	3.72	4	1.31	0.93	4.55
	4	3.72	5	1.26	0.93	4.37
	4	3.72	6	1.20	0.93	4.16

NOTES: Strata values are as follows: Maryland: 1 = Baltimore, 2= rest of state; South Carolina: 1=mandated population, 2=exempt population, 3=extended population, 4=unknown.

Final weights may vary from the product of the three component weights shown due to rounding. Component weights were rounded to two decimal places throughout this appendix for presentational purposes but were not rounded when calculating the final weight.

Final survey weights

The final weights for the survey respondents in the District of Columbia, Maryland, and South Carolina are the product of the three components discussed above and are presented in Table A.6. There is a unique weight for each final cell, the number of which varies for each site based on the number of quantiles determined in component 2 multiplied by the number of sampling strata. There are 5 cells in the District of Columbia, 10 in Maryland, and 24 in South Carolina.

TANF RECIPIENTS AS POTENTIAL LONG-TERM CARE WORKERS: An Assessment of the Prospects in the District of Columbia, Illinois, Maryland and South Carolina

PDF Files Available for This Report

Main Report http://aspe.hhs.gov/daltcp/reports/TANFltc.pdf

APPENDIX A: Summary Table for the TANF Caseload Surveys
http://aspe.hhs.gov/daltcp/reports/TANFltcA.pdf

APPENDIX B: Measures of Personal Liabilities
http://aspe.hhs.gov/daltcp/reports/TANFltcB.pdf

APPENDIX C: State-Specific Data on Employment, Demographic Characteristics and Employment Liabilities
http://aspe.hhs.gov/daltcp/reports/TANFltcC.pdf

APPENDIX D: State-Specific Predicted Employment Probabilities
http://aspe.hhs.gov/daltcp/reports/TANFltcD.pdf