APPENDIX F

SURVEY WEIGHTS AND STATISTICAL METHODS

Amang S. Sukasih
Sonya L. Vartivarian
Esther M. Friedman
Donsig Jang

PAGE IS INTENTIONALLY LEFT BLANK TO ALLOW FOR DOUBLE-SIDED COPYING

The National Evaluation of the Welfare-to-Work (WtW) Grants Program uses information on individuals who participated in eleven programs that received WtW grants. The eleven programs are referred to as "sites" in this appendix, labeled according to the city or state in which they operated. All individuals who entered WtW at one of the eleven sites during a specified period should have been enrolled in the evaluation. However, comparison of WtW program records with data on individuals actually enrolled revealed that program staff failed to enroll a substantial group of WtW participants in the evaluation, resulting in undercoverage.[1] Individuals who were enrolled in the evaluation are referred to as covered cases; those who should have been enrolled but were not are identified as noncovered cases.

The process of data collection can be described as follows. At the time of enrollment, a baseline survey was administered as part of the sample enrollment process. The survey instrument was a hard copy questionnaire called the Background Information Form (BIF), administered only to covered cases. A follow-up survey (Wave 1) was conducted 12 months after the baseline survey on covered cases only. Hence, this survey had two sources of missing data: the noncovered cases, and the covered cases who did not respond to the survey. A second follow-up survey (Wave 2) was conducted two years after the baseline survey, regardless of response status to the Wave 1 survey. For Baltimore County, Chicago, Phoenix, St. Lucie County and Yakima, the undercoverage was discovered early in the sample intake period, allowing administration of the 24-month survey to the noncovered cases (Exhibit F.1 provides an illustration of the data structure for these sites). For Boston, Ft. Worth, Nashville, Philadelphia and West Virginia, the timing of undercoverage detection did not allow for the inclusion of noncovered cases in the sample. For these five sites, the data structure was the same with respect

[1] Boston and Milwaukee did not maintain electronic program records and as a result, we were unable to explore potential undercoverage there. Refer to Fraker et al. (2004), Appendix C, for detailed discussion of the covered population.

to coverage in both the 12-month and 24-month surveys. The 24-month survey also had unit nonresponse, but the set of unit nonrespondents differed for Wave 2 due to different coverage, when applicable, and response status as compared to Wave 1.

EXHIBIT F.1

DATA STRUCTURE FOR BASELINE/ADMINISTRATIVE, 12- AND 24-MONTH SURVEYS

<table>
<tr><th>Baseline/ Administrative Information</th><th>12-Month Survey (Wave 1)</th><th>24-Month Survey (Wave 2)</th></tr>
<tr><td rowspan="4">Covered Cases</td><td rowspan="2">Wave 1 Respondents</td><td>Wave 2 Nonrespondents</td></tr>
<tr><td rowspan="2">Wave 2 Respondents</td></tr>
<tr><td>Wave 1 Nonrespondents</td><td>Wave 2 Nonrespondents</td></tr>
<tr><td colspan="2" rowspan="2">Noncovered Cases</td><td>Wave 2 Respondents</td></tr>
<tr><td>Wave 2 Nonrespondents</td></tr>
</table>

Note: Shading designates subgroups of the evaluation sample for which data from the surveys indicated by the column headings were not available.

A summary of the data available is as follows:

- ***State Administrative Data.*** We requested state administrative data for all members of the population whom we wanted to characterize. We were able to code quarterly measures of TANF benefits, food stamps, and earnings for each member of the population.
- ***Baseline Survey Data.*** Virtually all covered cases completed the BIF. However, as described earlier, some individuals who should have been enrolled in the study were not, and thus did not complete a BIF.

- ***12-Month Follow-Up Survey Data.*** We attempted to interview all individuals who were enrolled in the study for the 12-month follow-up survey. While the overall response rate was high at 83 percent—particularly for a population of individuals that can be difficult to locate—outcome variables based on 12-month follow-up survey data are necessarily missing for nonrespondents.
- ***24-Month Follow-Up Survey Data.*** Similarly, outcome variables measured at the 24-month follow-up are necessarily missing for nonrespondents of the 24-month follow-up survey data. However, since we were able to target noncovered cases in five sites, the response and coverage patterns differ across the 12- and 24-month follow-up surveys.

Because the 12- and 24-month follow-up surveys have different response and coverage patterns, we could not use in the analysis based on the 24-month survey the same statistical weights developed to address the undercoverage and nonresponse problems in the 12-month survey. We therefore developed separate statistical weights and procedures for the 24-month follow-up survey. In addition to descriptive analyses, we compared the Wave 1 and Wave 2 outcomes. This was complicated, however, by the fact that the samples are not independent across two data collections, where different response and/or coverage patterns occur across waves—making the variance-covariance terms in the estimation difficult to compute. To correct this, we constructed replicate weights and used them to compute these variances.

This appendix describes the final disposition of the sample for the 24-month follow-up survey (Section A), the weighting methods taken into account for biases that might result from survey nonresponses and undercoverage in the 24-month follow-up survey (Section B), and the development of replicate weights and variance estimation for comparisons between Wave 1 and Wave 2 estimates (Section C).

A. FINAL DISPOSITION OF THE SAMPLE FOR THE 24-MONTH SURVEY

The overall response rate for the 24-month follow-up survey was 74 percent, ranging from 57 percent in Phoenix to 86 percent in Nashville (Exhibit F.2).[2] This rate is high, particularly when considering the late addition of noncovered cases into the target sample of five sites (Baltimore County, Chicago, Phoenix, St. Lucie County and Yakima). Since the contact information for noncovered cases was on average older than that for the covered cases, it is not surprising that these five sites were the ones that tended to exhibit the lowest response rates. For example, Phoenix exhibited the lowest response rate of all sites (57 percent) and Yakima exhibited the second lowest (64 percent). Milwaukee's response rate of 69 percent was relatively low in part because the target sample consisted mostly of noncustodial fathers with a criminal offense in their record—a group of individuals generally considered hard to locate.

The most common reason for not completing the 24-month follow-up survey was not being able to verify the contact information for the target respondent ("unlocated" in Exhibit F.2). About 3 of every 4 cases where we could not complete an interview were due to this reason. In Phoenix, almost all the non-complete cases can be attributed to this reason. The inability to locate individuals once the contact information had been verified represented about 13 percent of the non-completed interviews in all sites ("other located" in Figure F.2), and about a third of non-completed interviews in Milwaukee. Finally, the refusal rate tended to be low overall (3 percent) but was relatively high for Baltimore and Boston (7 and 8 percent, respectively).

The response rate for the 24-month follow-up survey was lower than that for the 12-month follow-up survey (74 versus 83 percent).[3] This may have been due in part to the naturally

[2] This rate is calculated as the percent of cases in the target sample for which we completed a survey interview. The target sample consisted of all covered cases for six of the sites and of both covered and noncovered cases for the other five sites.

[3] Refer to Fraker et al. (2004), Appendix Exhibit C-3, for statistics on the disposition of the sample for the 12-month follow-up survey.

EXHIBIT F.2

FINAL DISPOSITION OF THE SAMPLE FOR THE 24-MONTH FOLLOW-UP SURVEY
(Percentages)

Final Disposition	Baltimore Co.	Boston	Chicago	Ft. Worth	Milwaukee	Nashville	Philadelphia	Phoenix	St. Lucie Co., FL	West Virginia	Yakima, WA	Total
Complete (Respondent)	70.6	76.3	67.4	83.3	69.3	85.8	83.9	57.5	77.8	90.7	64.3	73.9
Not Complete (Nonrespondent)	29.4	23.7	32.6	16.7	30.7	14.2	16.1	42.5	22.2	9.3	35.7	26.1
Refusal	7.0	7.7	2.2	1.9	2.9	4.1	3.0	0.4	3.4	2.1	2.8	2.8
Deceased	0.0	0.2	0.2	0.1	0.7	0.2	0.4	0.6	0.0	0.0	0.3	0.3
Language, Physical Barrier	0.0	0.1	0.0	0.1	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.1
Other Located[1]	3.4	2.8	3.4	3.3	10.2	2.1	2.3	1.3	3.4	2.4	6.1	3.4
Unlocated[2]	19.0	12.9	26.8	11.3	16.9	7.8	10.0	40.2	15.4	4.8	26.5	19.5
Sample Size	242	807	3,246	994	274	641	1,282	497	234	289	618	9,124

Source: Tracking database for the evaluation.

[1] Sample members for whom contact information was verified, but who were incarcerated, had moved outside of the area, could not be contacted, or were unavailable for an interview.

[2] Sample members for whom contact information could not be verified.

increasing difficulty of interviewing sample members as time elapses following sample selection, but was almost certainly also due to the fact that the target sample was expanded between the two waves of the survey to include noncovered cases in five sites. In fact, these sites tended to exhibit the greatest decline in response rates. Particularly worth noting are Phoenix and Yakima, which had 12-month survey response rates of 75 and 93 percent respectively and 24-month survey response rates of 57 and 64 percent respectively. In contrast, the sites where the target sample was not expanded to include noncovered cases exhibited similar response rates in both waves of the survey.

B. WEIGHTS TO ACCOUNT FOR MISSING DATA

The 24-month follow-up survey was designed to characterize enrollees at the eleven sites two years after program entry. However, some enrollees did not respond to this follow-up survey. If the individuals who did not respond to the survey differ systematically from those who did, sample nonresponses could bias the estimates based on data from the survey. In addition, as described in Appendix C of Fraker et al. (2004), WtW program staff failed to enroll a substantial minority of WtW participants, resulting in undercoverage. If covered cases differ systematically from noncovered cases, sample undercoverage could bias the estimates from the follow-up survey. The purpose of the weighting adjustments to the respondents of the 24-month follow-up survey is therefore two-fold: to account for survey nonrespondents, and for undercoverage in the WtW enrollment. The adjusted weights are expected to reduce bias due to nonresponse and/or undercoverage.

Data availability, sample size and coverage problems differ for each of the eleven sites. Therefore, we developed weights for each site separately. We developed a general protocol following the weighting procedure for nonresponse adjustment and post-stratification that we used for the 12-month follow-up survey. Although consistency in weighting class formation and

collapsing (merging) of cells is generally desired, we explored alternative statistical methodologies in several sites to accommodate unique data and model situations. In this section, we outline general procedures for weighting class formation and collapsing of cells.

For the purpose of nonresponse and undercoverage adjustment, information on basic demographic characteristics are available for both covered cases (gathered from the baseline survey) and for noncovered cases from data extracted from each program's Management Information System (MIS).[4] Furthermore, administrative data obtained from the states—on employment status, earnings, and TANF and Food Stamp program participation status—were also available for both covered and noncovered cases. Since the demographic and administrative data were available for both covered and noncovered cases, we compared respondents and nonrespondents to the 24-month follow-up survey using these data.

1. Preliminary Bias Analysis

The bias due to nonresponse to the 24-month follow-up survey is a function of the nonresponse rate and the relationship between the response probability and the survey outcome of interest. The 24-month response rates are presented in Exhibit F.3.[5] The relationships between nonresponse and the outcomes analyzed in this report are unknown because the information needed to construct the survey outcome measures was collected only for respondents. However, the 24-month follow-up survey outcomes may be related to basic demographic information from the baseline survey and to program participation and employment information from state administrative records. Therefore, we assess the differences between respondents and nonrespondents in basic demographics, welfare participation, and employment.

[4] Program MIS data were collected for all sites except Milwaukee and Boston, which did not maintain electronic records that we could use.

[5] No explicit sampling was performed to select the sample for the evaluation, and the sampling weight can be set to one for all cases. Hence, the weighted and unweighted response rates are equal.

EXHIBIT F.3

SAMPLE SIZES, COVERAGE RATES, AND RESPONSE RATES FOR THE 24-MONTH FOLLOW-UP SURVEY

	Baltimore Co.	Boston	Chicago	Ft. Worth	Milwaukee	Nashville	Philadelphia	Phoenix	St. Lucie Co., FL	West Virginia	Yakima, WA	Total
All Cases[1]	242	807	3,249	3,201	276	811	2,543	497	234	337	618	12,815
Covered Cases[2]												
Number	242	807	3,249	996	276	643	1,282	497	234	290	618	9,134
Pct. of all cases	100%	100%	100%	31%	100%	79%	50%	100%	100%	86%	100%	71%
Respondents[3]												
Number	171	615	2,189	828	190	550	1,075	286	182	262	397	6,745
Pct. of covered cases	71%	76%	67%	83%	69%	86%	84%	57%	78%	90%	64%	74%

Source: Survey tracking database for the evaluation and program records provided by WtW grantees.

[1] All WtW participants who should have been enrolled in the evaluation according to program records.

[2] WtW participants who were enrolled into the evaluation via procedures described in Section A of Appendix C in Fraker et al. 2004.

[3] Covered WtW participants (enrolled in the evaluation) who respoonded to the 24-month follow-up survey.

This assessment revealed statistically significant differences at the 0.05 level between respondents and nonrespondents to the 24-month follow-up survey on several dimensions, including sex, age, race and ethnicity, marital status, sources of income, and the timing of program entry (Exhibit F.4).

- ***Sex.*** Respondents were significantly more likely than nonrespondents to be female in Baltimore County, Boston, Ft. Worth, and Yakima.

- ***Age.*** In Chicago and Ft. Worth, respondents were significantly more likely than nonrespondents to be less than 25 years old; in Ft. Worth, respondents were significantly less likely than nonrespondents to be 25 to 40 years old. In Yakima, respondents are significantly more likely than nonrespondents to be more than 40 years old.

- ***Race and Ethnicity.*** Philadelphia had a significantly higher proportion of respondents than nonrespondents who are black and non-Hispanic. In Ft. Worth, the proportion of respondents who were white and non-Hispanic was significantly less than that of nonrespondents.

- ***Marital Status.*** In Ft. Worth, a significantly higher proportion of nonrespondents than respondents were married, whereas respondents had a significantly higher proportion reporting themselves never married. In West Virginia, a significantly higher proportion of nonrespondents reported themselves never married than respondents, but a significantly lower proportion had been previously married.

- ***Sources of Income.*** In Phoenix, Nashville, and Chicago, respondents were significantly more likely than nonrespondents to have income from TANF benefits. In Philadelphia, Milwaukee, and Nashville, respondents were significantly more likely than nonrespondents to have income from TANF and food stamps.

As described above, we developed 24-month follow-up survey weights for respondents to account for differences between respondents and nonrespondents and covered and noncovered cases, and used them in computing estimates based on 24-month follow-up survey data. The strategy for developing these weights involved two stages: identifying variables that are predictive of response to the 24-month follow-up survey, and developing weights to reduce bias due to survey nonresponse.

EXHIBIT F.4

COMPARISON OF SURVEY RESPONDENTS (R) WITH NONRESPONDENTS (N)
(Percentages)

	Baltimore Co.			Boston			Chicago			Ft. Worth			Milwaukee			Nashville		
Characteristic	R	N	Sig.	R	N	Sig.	R	N	Sig.	R	N	Sig.	R	N	Sig.	R	N	Sig.
Female	94	86	*	94	88	***	97	96		95	89	***	b	b		99	99	
Age at WtW Enrollment																		
Less than 25 years old	22	23		40	45		36	32	**	44	36	*	37	41		32	35	
25 - 40 years old	61	51		52	51		52	55		46	57	**	55	53		58	57	
More than 40 years old	17	26		7	5		12	13		10	8		8	6		10	8	
Race and Ethnicity																		
White, non-Hispanic	19	13		7	6		2	2		21	29	**	4	7		11	11	
Black, non-Hispanic	80	82		60	62		93	89	***	51	45		86	86		87	88	
Hispanic	b	b		33	29		5	8	***	27	25		10	7		b	b	
Other	b	b		b	3		1	0		b	b		0	0		b	b	
Marital Status																		
Currently married	c	c		5	6		c	c		8	15	**	b	b		b	b	
Never married	c	c		76	75		c	c		57	46	**	62	71		67	68	
Previously married	c	c		20	18		c	c		35	39		32	25		30	27	
Education																		
Less than H.S. diploma	c	c		60	58		59	60		67	74	***	81	86		61	64	
High school diploma	c	c		29	31		28	27		23	15	**	14	10		27	23	
More than H.S. diploma	c	c		11	11		13	13		11	10		5	5		12	13	
Sources of Income [1]																		
Earnings	72	70		d	d		37	40		47	50		48	51		d	d	
TANF benefits	34	25		d	d		88	85	**	93	90		b	b		97	92	**
Food Stamp benefits	61	54		d	d		89	90		92	91		33	20	**	98	95	**
Early Enrollment	50	46		15	16		0	0		12	11		43	19	***	40	38	
Sample Size [2]	171	71		615	192		2,189	1,060		828	168		190	86		550	93	

EXHIBIT F.4 *(continued)*

	Philadelphia			Phoenix			St. Lucie Co., FL			West Virginia			Yakima, WA		
Characteristic	R	N	Sig.	R	N	Sig.	R	N	Sig.	R	N	Sig.	R	N	Sig.
Female	99	99		97	95		95	92		84	82		79	73	*
Age at WtW Enrollment															
Less than 25 years old	30	34		49	c		42	43		29	25		30	34	
25 - 40 years old	58	56		43	c		50	47		57	64		54	55	
More than 40 years old	12	10		7	c		8	10		b	b		16	10	*
Race and Ethnicity															
White, non-Hispanic	1	3		9	11		39	49		85	79		d	d	
Black, non-Hispanic	91	86	*	34	40		57	22	***	13	18		d	d	
Hispanic	7	7		50	41	*	3	27	***	b	b		d	d	
Other	1	3	***	7	8		b	b		b	b		d	d	
Marital Status															
Currently married	3	3		8	c		10	c		27	29		c	c	
Never married	80	81		64	c		57	c		30	46	*	c	c	
Previously married	17	17		29	c		33	c		43	25	*	c	c	
Education															
Less than H.S. diploma	72	75		79	79		c	60		60	68		c	c	
High school diploma	23	19		16	17		c	27		32	21		c	c	
More than H.S. diploma	5	7		4	4		c	13		b	b		c	c	
Sources of Income [1]															
Earnings	72	62	***	49	50		77	83		15	7		d	d	
TANF benefits	98	98		91	83	**	76	69		94	96		d	d	
Food Stamp benefits	97	94	**	92	89		91	92		95	96		d	d	
Early Enrollment	75	71		34	33		38	35		68	68		63	62	
Sample Size [2]	1,075	207		286	211		182	52		262	28		397	221	

Source: 1999-2002 Baseline Survey of WtW enrollees, program MIS data provided by WtW grantees, and state administrative data.

[1] Quarterly income approximately one year after WtW enrollment.

[2] Sample sizes do not reflect item missing data. Statistics are reported when the rate of item missing data is 20 percent or less.

a: No MIS data on uncovered cases; b: Less than five observations for the group of interest; c: More than 20 percent of the values are missing;

d: Data not available

*/**/*** Difference between survey respondents and nonrespondents is statistically significant at .10/.05/.01 level.

2. Nonresponse Adjustment

For the 24-month follow-up survey, we formed response propensity weighting cells to adjust for nonresponse. The response status was modeled via a logistic regression with covariates observed for all sample individuals. Such covariates generally consisted of demographic variables (e.g., sex) and administrative variables (e.g., earnings). Exhibit F.4 provides a complete list of demographic and administrative variables. Model selection included quality checks for missing data on such covariates. Variables with substantial missing data were excluded from the analysis; those with few missing values were imputed solely for the purpose of obtaining weights. All covariates are used in a categorical form, thus imputation with the mode value and with a new category value are compared.

For each site, we estimated the response propensity via a logistic regression of survey response status on variables describing basic demographics, the timing of program entry, welfare participation, and employment. In addition, we included the 12-month interview status as a potential predictor for response status in the 24-month survey. We then used the same stepwise procedure for selecting variables for weighting cell construction.

The potential variables included in the weighting algorithm varied across sites according to data availability and quality. The variables selected for the largest number of sites were 12-month interview status (11 sites), receipt of TANF benefits (7 sites), and receipt of food stamps (6 sites) (Exhibit A.3). Education was selected only for Ft. Worth. The 12-month coverage status was important in the model in all sites where noncovered cases were added to the 24-month follow-up target sample, with the exception of Baltimore County.

We then considered the covariates included in the response propensity in our univariate logistic regressions and jointly in a full model and in a stepwise selected model. Further model selection was carried out to improve model fit and to allow for a parsimonious final model.

Once the final model was selected, the quintiles of the distribution of the predicted response propensity scores formed the weighting classes used in nonresponse adjustments. The nonresponse adjustment factor is the inverse of the response rate in a cell, where the response rate is estimated as the number of respondents in a cell divided by the number of sampled individuals in that cell. The number of cells initially formed ranges from five to eight, depending on the sample size for a site and the resulting adjustment weights (Cochran 1968; Eltinge and Yansaneh 1997).

When we encountered a small sample size in a particular cell, extreme weights, or sparseness, we considered using a cell collapsing procedure. Before implementing the procedure, we compared the predicted response propensities for each covariate pattern for each set of the potential adjustment cells to explore any other natural regrouping of subjects. We then determined whether the current weight in a questionable cell was actually large with respect to the range of weights within that cell that would result from using the inverse of the predicted response propensity itself as the nonresponse adjustment factor. We made final adjustments to cells based on the considerations noted above.

When collapsing was determined to be necessary and the data were assumed to be missing at random (MAR) given the response propensity, we grouped observations with similar propensity scores—often collapsing a fringe cell with its adjacent cell. In the case where a cell was not on the fringe, we compared the median response propensity for observations within adjacent cells to the median response propensity of the problematic cell, and collapsed the adjacent cell with the smallest absolute difference in median response propensity. The final decision on whether and how to collapse was site-specific, though the collapsing of adjacent cells was generally preferred.

3. Coverage Adjustment

To account for undercoverage, the sample was post-stratified (cross-classified) by site and the variables used in the 12-month post-stratification cells. With the exception of West Virginia, where cells that contained relatively few sample members were combined with other cells (with similar coverage rates if possible), we used the post-stratification cells defined in the 12-month analysis and determined the 12-month and 24-month values of certain variables. We computed a coverage adjustment factor for each post-stratification cell, using the nonresponse adjusted weights computed earlier to calculate the coverage adjustment factors. We calculated the final 24-month follow-up survey weights by multiplying the coverage adjustment factors by the nonresponse-adjusted weights for all sample members in the same cell. Hence, the final 24-month follow-up survey weights account for both undercoverage and survey nonresponse.

4. Design Effects

The precision of our estimates depends in part on the variability in the 24-month follow-up survey weights. Unequal response rates and coverage probabilities across adjustment cells used in explaining response propensity or coverage models justify variability in the 24-month follow-up survey weights to reduce bias. We did not have access to survey outcomes when developing the 24-month follow-up survey nonresponse weights, thus design effects due to weight variation can be computed as one plus the square of the coefficient of variation in the weights (Kish 1987). This design effect, presented in the first row of Exhibit F.5, ranges from 1.053 in West Virginia to 1.468 in Philadelphia.

Once the survey outcomes were made available, we evaluated the design effects across different estimates. The design effect is defined as the ratio of variance of a point estimator based on a nontrivial weighting adjustment for nonresponse to variance of a point estimator based on a trivial weighting adjustment, or equivalently, giving all responding units the weight

EXHIBIT F.5

CHARACTERISTICS USED IN CONSTRUCTING NONRESPONSE WEIGHTS FOR 24-MONTH FOLLOW-UP SURVEY

Variable	Baltimore Co.	Boston	Chicago	Ft. Worth	Milwaukee	Nashville	Philadelphia	Phoenix	St. Lucie Co., FL	West Virginia	Yakima, WA
Female		√		√							
Age at Enrollment	√			√							√
Race and Ethnicity		√	√		√						
Marital Status				√						√	
Education				√							
Sources of Income [1]											
Earnings			√	√	√		√	√	√	√	
TANF						√		√		√	
Food Stamps	√		√		√	√	√			√	
Early Enrollment		√			√				√		√
Covered Status			√					√	√		√
12-Month Intervew	√	√	√	√	√	√	√	√	√	√	√

Source: 1999-2002 Baseline Survey of WtW enrollees, program MIS data provided by WtW grantees, and state administrative data.

[1] Quarterly income approximately one year after WtW enrollment.

n/r (where n = original sample size and r = number of respondents). The design effects vary across different variables and sites, but they range from about 0.76 in Yakima for the proportion employed two years after enrollment to about 1.9 in Phoenix for the proportion with children living outside the household (Exhibit F.6). As the design effects reveal, the variance of the weighting class mean can have a lower variance than that of the trivially weighted mean. This reduction in variance may be due to post-stratification and is often evidenced when variables forming post-strata are predictive of the survey outcomes (Holt and Smith 1979; Little 1993). Similarly, nonresponse adjustment cells that are associated with the survey outcomes may result in a reduction in variance when using a weighting class mean.

5. Bias Reduction Due to Weighting

Earlier in the appendix, we displayed differences between the average characteristics of respondents and nonrespondents in terms of basic demographics, TANF and food stamp participation, and employment outcomes. We now examine how representative the respondents are when weighted by the 24-month follow-up survey weights. Our analysis reveals that for these variables, the weighted proportions for respondents are very close to the population proportions (Exhibit F.7). As a consequence, the error rates tend to be low; the exceptions tend to involve very sparsely populated cells.

C. ESTIMATION METHODS AND MEASURES OF PRECISION

1. General Procedures

For this evaluation, we did not perform explicit sampling in our selection of the sample. We therefore constructed weights solely to address missing data (as described in Appendix C of Fraker et al. 2004). In computing estimates for the report, the choice of weights depended on the data source from which the analysis variable was computed. Since variables constructed from state administrative data rarely contained missing values, no weighting was used to compute

EXHIBIT F.6

DESIGN EFFECTS DUE TO VARIABILITY IN SURVEY WEIGHTS:
SELECTED VARIABLES CONSTRUCTED FROM THE 24-MONTH FOLLOW-UP SURVEY

Outcome Measure (App. B Exhibit Number)	Baltimore Co.	Boston	Chicago	Ft. Worth	Milwaukee	Nashville	Philadelphia	Phoenix	St. Lucie Co., FL	West Virginia	Yakima, WA
Over All	1.0580	1.1050	1.0830	1.2310	1.1290	1.0880	1.4680	1.2620	1.0670	1.0530	1.1230
Proportion of Enrollees Receiving TANF Two Years After Entry (B.11)	0.9746	1.1242	1.0447	1.1915	0.8559	1.0883	1.4652	1.2362	0.9927	0.9943	1.0843
Mean TANF Benefit Two Years After Entry, for Enrollees Who Received TANF (B. 11)	0.9293	1.1797	0.9727	1.1185	0.9058	1.1660	1.2584	1.1453	0.8960	0.9891	0.9899
Proportion of Enrollees Employed During the Second Year After Entry (B.1)	1.0538	1.1102	1.0765	1.1125	1.2018	1.0999	1.5482	1.2644	1.0760	1.0605	1.1408
Proportion of Enrollees Employed Two Years After Entry (B.3)	1.0193	0.9883	1.1333	1.0023	0.7865	1.1334	1.0773	1.8080	0.9703	1.4084	0.7629
Mean Hourly Wage on the Principal Job Held Two Years After Entry (B.8)	1.1601	0.9697	1.1866	1.4780	0.9706	0.9737	1.6468	1.6711	0.9808	1.1563	1.0022
Mean Total Monthly Household Income Two Years After Entry (B.16)	1.0241	1.1594	1.0753	1.3436	1.2883	1.1105	1.4080	1.3809	1.0347	1.0753	1.1428
Mean Index of Material Distress During the Second Year After Entry (B.18)	1.2563	1.0611	1.2264	1.4108	1.1993	1.3349	1.4652	1.4817	1.0670	1.2621	1.1719
Prop. of Enrollees w/ Children Living Outside Household Two Years After Entry (B.24)	1.4642	1.3474	1.3595	1.5292	0.9830	1.3298	1.2302	1.8883	1.2847	1.3632	1.2132

Source: 2001-03 24-month follow-up survey of Welfare-to-Work enrollees.

EXHIBIT F.7

COMPARISON OF WEIGHTED SURVEY RESPONDENTS (R) WITH ALL CASES (A)
(Percentages)

	Baltimore Co.			Boston			Chicago			Ft. Worth			Milwaukee			Nashville		
Characteristic	R	A	Error Rate	R	A	Error Rate	R	A	Error Rate	R	A	Error Rate	R	A	Error Rate	R	A	Error Rate
Female	92	91	1.2	94	a	---	97	97	-0.1	92	96	-3.8	5	a	---	99	99	0.2
Age at WtW Enrollment																		
Less than 25 years old	22	23	-1.2	41	a	---	36	35	3.7	43	a	---	38	a	---	33	a	---
25 - 40 years old	58	58	0.6	52	a	---	52	53	-2	47	a	---	55	a	---	58	a	---
More than 40 years old	19	20	-0.4	7	a	---	12	12	-1.7	10	a	---	7	a	---	10	a	---
Race and Ethnicity																		
White, non-Hispanic	18	17	4	7	a	---	2	2	0.1	22	a	---	4	a	---	11	a	---
Black, non-Hispanic	81	80	1.1	60	a	---	91	91	-0.1	50	a	---	87	a	---	87	a	---
Hispanic	b	b	---	32	a	---	6	6	0.9	26	a	---	10	a	---	b	a	---
Other	b	b	---	b	a	---	1	1	7.1	b	a	---	0	a	---	b	a	---
Married at WtW Enrollment	c	c	---	4	a	---	c	c	---	10	a	---	5	a	---	3	a	---
Education																		
Less than H.S. diploma	c	c	---	60	a	---	59	59	-0.4	67	a	---	80	a	---	61	65	-5.9
High school diploma	c	c	---	28	a	---	28	28	0.6	22	a	---	14	a	---	27	24	9.1
More than H.S. diploma	c	c	---	11	a	---	13	13	0.4	10	a	---	6	a	---	12	10	15.5
Sources of Income [1]																		
Earnings	71	71	-1.1	d	d	---	38	38	-0	55	55	-0	45	49	-9.2	d	d	---
TANF benefits	33	31	4.1	d	d	---	87	87	-0	93	92	0.7	b	b	---	97	96	0.7
Food Stamp benefits	59	59	-0.9	d	d	---	88	90	-1.3	92	92	0.1	29	29	1.2	98	97	0.9
Early Enrollment	49	49	0	14	a	---	0	0	0	21	21	-0	38	36	8.3	42	42	-0
Sample Size [2]	171	242		615	807		2,189	3,249		828	3,201		190	276		550	811	

EXHIBIT F.7 *(continued)*

	Philadelphia			Phoenix			St. Lucie Co., FL			West Virginia			Yakima, WA		
Characteristic	R	A	Error Rate	R	A	Error Rate	R	A	Error Rate	R	A	Error Rate	R	A	Error Rate
Female	99	99	0	95	c	---	95	94	0.6	78	78	0	77	77	-0
Age at WtW Enrollment															
Less than 25 years old	29	c	---	50	c	---	42	42	-1.4	28	29	-3.5	32	32	0
25 - 40 years old	58	c	---	43	c	---	51	49	2.6	57	57	-0.3	54	54	-0
More than 40 years old	13	c	---	7	c	---	7	8	-8.1	15	14	8.6	14	14	0
Race and Ethnicity															
White, non-Hispanic	1	2	-17.5	11	c	---	41	c	---	85	83	1.8	d	d	---
Black, non-Hispanic	90	89	1	30	c	---	56	c	---	13	14	-10.5	d	d	---
Hispanic	7	8	-4	51	c	---	3	c	---	b	b	---	d	d	---
Other	1	1	-25.3	8	c	---	b	c	---	b	b	---	d	d	---
Married at WtW Enrollment	3	4	-30.5	9	c	---	12	c	---	29	28	2.3	c	c	---
Education															
Less than H.S. diploma	71	a	---	81	c	---	c	c	---	59	59	0.1	c	c	---
High school diploma	23	a	---	15	c	---	c	c	---	32	31	1.4	c	c	---
More than H.S. diploma	5	a	---	4	c	---	c	c	---	10	10	-5.1	c	c	---
Sources of Income [1]															
Earnings	79	79	0	52	50	4.4	77	79	-1.9	14	14	2.6	d	d	---
TANF benefits	98	98	0	86	88	-2	75	74	1.5	92	94	-2.2	d	d	---
Food Stamp benefits	95	97	-1.1	89	91	-2.3	90	91	-0.7	93	95	-1.9	d	d	---
Early Enrollment	47	47	-0	33	33	0	37	37	0	66	66	-0	63	63	0
Sample Size [2]	1,075	2,543		286	497		182	234		262	337		397	618	

Source: 1999-2002 Baseline Survey of WtW enrollees, program MIS data provided by WtW grantees, and state administrative data.

Error rate: 100 * (R - A) / A

[1] Income in the quarter of Welfare-to-Work enrollment.

[2] Sample sizes do not reflect item missing data. Statistics are reported when the rate of item missing data is 20 percent or less.

a: No MIS data on uncovered cases; b: Less than five observations for the group of interest; c: More than 20 percent of the values are missing; d: Data not available

means and proportions from these data. The coverage weights were constructed to address the problem of undercoverage and applied to variables computed from BIF data in computing means and proportions.[6] The 12-month follow-up survey weights, which were constructed to address both undercoverage and nonresponse to this survey, were applied to 12-month follow-up survey outcome variables in computing means and proportions.[7] Weights to account for nonrespondents of the 24-month follow-up survey data were constructed separately from those for the 12-month survey, as the set of covered and responding individuals differed for Wave 2. These Wave 2 weights were applied to the 24-month follow-up survey outcome variables in computing means and proportions.

Standard errors of our estimates are presented in Appendix C. These estimates provide the reader with the information necessary to assess the precision of the means and proportions presented in this report. Standard errors of survey-based estimates were computed using standard survey procedures in SUDAAN version 8 and SAS version 8. These methods are often referred to as robust variance estimation or variance estimation via Taylor series linearization methods, and we used them to account for the variability in the coverage weights, the 12-month follow-up survey weights, and the 24-month follow-up survey weights.

2. Replication Methods and Variance Estimation for Cross-Wave Comparisons

While the variance estimation procedures described in the previous subsection can be applied to both Wave 1 or Wave 2 survey data, they cannot be applied when computing differences between Wave 1 and Wave 2 outcomes because the two samples are not independent. Furthermore, response and coverage patterns differ across the two waves of data. Hence, basing

[6] These weights were also applied to variables from state administrative data in West Virginia, where we were unable to obtain administrative data for noncovered cases due to legal issues involving consent to participate in the study.

[7] Refer to Fraker et al. (2004), Appendix C, for details

the statistical tests on only those who responded to both surveys would decrease the estimates' efficiency, as we would lose a substantial number of observations.

Computing variance that accounts for different response and/or coverage patterns across surveys is complex. As an alternative, we generated replicate weights to obtain variance estimates of the outcome differences between Wave 1 and Wave 2 with a replication method. We used the jackknife method to generate these replicate weights, calculated a series of replicate weights, and attached them to each record in the dataset. Fifty replicate weights were created separately for Wave 1 and Wave 2 data, and independently for each of the sites. The number of replicate weights is the same for all sites and across waves.

When using the jackknife replication method, deleting selected cases from the full sample generates the prescribed number of replicates. Prior to computing the replicate weights, we created fifty replicates based on cases in the baseline survey. First, within each site, we sorted the file by person ID. Within this sorted file, we identified 50 mutually exclusive and exhaustive systematic subsamples of the full sample. A jackknife replicate was then obtained by dropping one subsample from the full sample.

Next, as each replicate was constructed, we applied the entire weighting process—including the Wave 1 nonresponse adjustment and post-stratification adjustment, and the Wave 2 nonresponse adjustment and post-stratification adjustment as applied to the full sample—separately to each of the jackknife replicates to produce a set of replicate weights for each record. In addition, the weights were adjusted by multiplying by a factor of 50/49 to account for dropping one subsample in the creation of replicates. Finally, the series of jackknife replicate weights (JKW1_1-JKW1_50 for Wave 1; JKW2_1-JKW2_50 for Wave 2) was attached to the final data in order to construct jackknife replication variance estimates. These replicate weights

were used to estimate the variances of the estimates of the differences between Wave 1 and Wave 2 statistics.

Estimation of the variance through a jackknife method is performed by taking differences between the point estimates computed based on replicate samples and that using the full sample. For a stratified sample, the general formula for the jackknife variance estimator in SUDAAN (RTI 2001) can be expressed as:

$$v_{Jack}\left(\hat{\theta}\right) = \sum_h \frac{N_h - D_h}{D_h S_h} \sum_i \left(\hat{\theta}_{hi} - \hat{\theta}\right)^2$$

where:

- N_h is the number of primary sampling units (PSUs) or clusters within the stratum *h*
- D_h is the number of PSUs or clusters deleted in creating the replicate
- S_h is the number of replicates selected
- $\hat{\theta}_{hi}$ is the estimate of the parameter θ from the *i*-th replicate of the *h*-th stratum
- $\hat{\theta}$ is the estimate based on the entire/full sample

In this case, $\hat{\theta}$ is defined as the difference between Wave 1 and Wave 2 estimates, $\hat{\theta} = \hat{\theta}_1 - \hat{\theta}_2$, where:

- $\hat{\theta}_1$ is the Wave 1 estimate based on the entire/full sample
- $\hat{\theta}_2$ is the Wave 2 estimate based on the entire/full sample

Note that $\hat{\theta}_{hi}$ should be calculated in the same fashion as $\hat{\theta}$.

As described in the previous paragraphs, jackknife replicate weights were constructed without stratification and based on fifty random groups. In this case, we view the sample as if it came from a single big stratum containing fifty clusters. One cluster was randomly deleted to construct a replicate, and all fifty possible replicates were selected. Consequently, the multiplier for jackknife variance estimation can be

$$\frac{N_h - D_h}{D_h S_h} = \frac{50-1}{1\times 50} = \frac{49}{50}$$

for this stratum. The simplified jackknife variance estimator can thus be expressed as:

$$v_{Jack}\left(\hat{\theta}\right) = \frac{49}{50}\sum_{r=1}^{50}\left(\hat{\theta}_r - \hat{\theta}\right)^2$$

where this formula can be computed easily using a macro that loops 50 times.