Using Predictive Marginals to Produce Standardized Estimates

Michael B. Witt[a] and Kathryn E. Spagnola[b]
[a,b]RTI International, P.O. Box 12194, Research Triangle Park, NC 27709

Abstract[1]
Researchers are often naturally interested in comparing the prevalence of an attribute between subgroups of a population or between time periods. To account for distributional differences between the subgroups, one solution is to examine the difference between standardized estimates. This paper presents a discussion of standardized estimates created using three different approaches: direct standardization, creating standardized estimates by adjusting the sample weight (in effect, creating standardized weights) and computing predictive marginals. Predictive marginals, in particular, is an appealing method of computing standardized estimates because it is a generalization of the direct standardization approach, it shares many of the same advantages of the weight adjustment approach and predictive marginals, contrasts of predictive marginals and the associated standard errors can be estimated using some standard statistical software products. Examples will be provided using survey data from SAMHSA and DoD studies.

Keywords: Standardization, Predictive Marginals, Military Data, Survey Analysis

1. Overview

Social scientists and researchers are often interested in comparing the prevalence of some attribute between two or more subgroups of a population or between different time periods. For example, a researcher might be interested in comparing cigarette use among different service branches of the active duty military personnel (e.g. Army, Navy, Marine Corps and Air Force.) Similarly, a researcher might be interested in comparing the trends in marijuana use among 12-17 year-olds between successive years. Often times, significant differences are observed between groups or between time periods due to distributional differences across variables that are highly correlated with the measure of interest. For example, the use of marijuana among 12-17 year-olds might be increasing across time because the drug is more prevalent in certain geographic areas and the percent of 12-17 years-olds that reside in these areas has increased across the same time period. To account for these distributional differences, one solution is to examine the difference between standardized estimates.

In this paper, the predictive marginals approach will be discussed as an appealing alternative to creating standardized estimates. The advantages and disadvantages associated with using predictive marginals to create standardized estimates will be discussed and these will be compared to two more commonly used methods of standardization: the direct standardization technique and creating standardizes estimates by using an adjusted, standardized sample weight. It will be shown that the predictive

[1]Mr. Michael Witt and Ms. Kathryn Spagnola are from the Statistics and Epidemiology Division at RTI International, Research Triangle Park, NC. The opinions or assertions herein are those of the authors and do not necessarily reflect the views of the United States Department of Defense or the views of the United States Department of Health and Human Services, Substance Abuse and Mental Health Services Administration.

marginal approach is a generalization of the direct standardization technique and can be useful when one is interested in controlling for a large number of effects during the standardization process. And it will also be shown that the predictive marginal approach shares some of the same advantages of the weight adjustment approach but is appealing because variance estimates associated with the predictive marginals are typically easier to obtain from standard statistical software packages. These methods will be illustrated using data from three studies including the National Survey on Drug Use and Health and the Department of Defense Surveys of Health Related Behaviors among Active Duty and Reserve Military Personnel. These techniques will be illustrated by comparing several statistics and their associated standard errors between the civilian population, active duty military personnel and the personnel in the United States guard and reserves.

2. Introduction

An integral part of many analyses in social science research is to measure any difference in some outcome measure between two or more groups of interest. When examining any potential difference between the subgroups of interest, it is often desirable to take into account other confounding factors; for example other sociodemographic characteristics such as age, race/ethnicity, gender, marital status and education. This is particularly desirable when both of the following are true:

- One or more sociodemographic or other auxiliary characteristic is highly correlated with the outcome measure of interest,

- The groups being compared have significantly different population distributions with respect to the sociodemographic/auxiliary characteristics under consideration.

Standardization is a technique commonly used to account for differences in population composition which may have an impact on estimates of an outcome measure. (Kalton, 1968; Konijn, 1973). When dealing with survey data, standardization can be thought of conceptually as creating an adjustment to the final sample weights so that the distribution of the reweighted sample in a group of interest equals some fixed distribution. This fixed distribution is often referred to as the **standardization population**. The standardization population can be obtained from some outside source, or is often estimated from the entire sample, without regard to the group(s) being considered. The standardized estimate (sometimes referred to as the adjusted mean) can be interpreted as the estimate that would have been obtained if the group exhibited the distribution of the standardizing population with respect to those characteristics being controlled for, all other things being equal (Little, 1982).

3. Direct Standardization

Consider the following, very simple hypothetical example. Suppose one were examining survey data and was interested in comparing some outcome measure between males and females as illustrated in **Table 1**. This table indicates the weighted survey distribution of males across the three age groups is 12.0%, 50.0% and 38.0% respectively. In comparison, females are distributed across the three age groups as 25.0%, 25.0% and 50.0% and the total population distribution is 10.0%, 27.0% and 63.0%. Our interest is in comparing the estimates between males and females.

Table 1. Hypothetical Example Illustrating Standardization

Age Group	Males		Females		Total
	Population Distribution	Estimate	Population Distribution	Estimate	Population Distribution
Nonstandardized (Direct) Estimates					
< 24	12.0%	12.0%	25.0%	30.0%	10.0%
25-34	50.0%	50.0%	25.0%	75.0%	27.0%
35+	38.0%	45.0%	50.0%	89.0%	63.0%
Total	100.0%		100.0%		
Estimate		**43.5%**		**70.8%**	
Standardized Estimates					
< 24	10.0%	12.0%	10.0%	30.0%	
25-34	27.0%	50.0%	27.0%	75.0%	
35+	63.0%	45.0%	63.0%	89.0%	
Total	100.0%		100.0%		
Estimate		**43.1%**		**79.3%**	

First, note from **Table 1** that the direct survey estimates, i.e. the nonstandardized estimates, can be computed as follows:

$$.435 = (.120 \cdot .120) + (.500 \cdot .500) + (.380 \cdot .450)$$
$$.708 = (.250 \cdot .300) + (.250 \cdot .750) + (.500 \cdot .890)$$

The second part of the **Table 1** illustrates the computation of a standardized estimate where the standardized population is estimated from the total population distribution. In this case, we are assuming the distribution of both the males and females across the three age groups is 10.0%, 27.0% and 63.0%. The standardized estimates are then derived as follows:

$$.431 = (.100 \cdot .120) + (.270 \cdot .500) + (.630 \cdot .450)$$
$$.793 = (.100 \cdot .300) + (.270 \cdot .750) + (.630 \cdot .890)$$

The above example illustrates a standardization method that is commonly referred to as **direct standardization** (see for example, Kalton, 1968). With direct standardization, suppose one wishes to incorporate k sociodemographic or other auxiliary variables into the standardization process. These k variables may be age group, gender, education, etc. First, cells are defined by the complete cross-classification of the k standardizing variables. Then, means calculated for each cell are weighted by the proportions in the standardizing population to calculate the standardized (or adjusted) overall mean – similar to what was illustrated with **Table 1**.

Direct standardization offers many advantages, including:

- The standardization population can come from some outside source. In other words, it does not need to be estimated from the entire sample distribution.

- Many software packages already have options that will allow one to compute direct standardized estimates and their associated standard errors. For example, the Stata 11 (StataCorp, 2009) and SUDAAN (Research Triangle Institute, 2008) software, will compute standardized estimates and their associated standard errors while also accounting for the complex design features of a study such as stratification, clustering and unequal weighting.

However, direct standardization also has many disadvantages, including:

- The number of standardizing variables one can use is limited by the sample size of the study under consideration. Specifically, the sample size in each cell of the cross-classification must be sufficiently large to adequately estimate the outcome measure under consideration within each cell.

- Often times it is difficult to find standardized population estimates for each cell if the number of variables in the cross-classification is large. One can use the entire sample to estimate these cell percentages but only if the cell sample size is adequately large.

- Depending on the outcome measure under consideration, the main effects of some potential standardization variables may be significant predicators of the outcome variable but the interaction of the standardization variables may not be.

- In general, continuous variables cannot be used in direct standardization unless they are treated as categorical and the sample size is sufficiently large in each cell.

4. Weight Adjustment Approach

One common method of computing standardized estimates is to explicitly create a weight adjustment to the analysis weights that will force the adjusted weights to sum to the same set of control totals for each subgroup of interest. There are several methods of computing a weight adjustment including the commonly known weighting class approach and an approach that computes weight adjustment by fitting a generalized exponential model. A discussion of software available to compute weight adjustments is provided in (Witt, 2009). Computing weights adjustments by fitting a generalized exponential model is the approach used in the SUDAAN software (SUDAAN, 2008) and the approach that was considered in this comparative analysis.

There are many advantages to creating a standardized weight that would be used to create the desired standardized estimates, including:

- The standardized weights can be included on a public use file or other analysis file. Placing these adjusted weights on the data files allows other researchers to easily replicate any standardized estimates.

- Using standardized weights allows one to compare estimates from non-overlapping domains of interest. This would be very difficult to do via a predictive modeling approach because various records would, by definition, belong to more than one subgroup of interest. For example, someone may wish to compute and compare standardized estimates between personnel in the Army and total male personnel in the military.

- Depending on how the weight adjustments are computed, one can include only main effect and lower order interactions of variables in the weight adjustment process. In other words, it is not necessary to create a complete cross-classification of standardization variables as is needed with the direct standardization method. This is advantageous because it allows one to use more variables in the standardization process compared to what could be used with the direct standardization approach. The model-generated weight adjustment approaches such as the generalized exponential model technique in SUDAAN (for example) allow one to only include main effect and lower order interactions of variables in the weight adjustment process.

- As with the direct standardization approach, one can use the weight adjustment approach to standardize subgroups to a population distribution obtained from some outside source or from some distribution estimated from the sample itself. The predictive marginal approach would not allow one to create standardized estimates that have been adjusted to a population distribution from an outside source. This will be discussed in **Section 5**.

There is one very important disadvantage to creating standardized estimates by simply using an adjusted weight, and that relates to variance estimates of the standardized statistics. Creating standardized estimates using an adjusted weight with standard survey software packages ignores the fact that the estimates have been standardized. Variance estimates will account for any change in the effects of unequal weighting induced by the adjusted weights but the variance estimates will not properly account for the fact that weights have been adjusted to a specified population distribution. This primary disadvantage does not exist with the direct standardization technique or the predictive marginal approach, because standard software packages can provide appropriate variance estimates with these standardization approaches (see for example, Stata or SUDAAN).

5. Predictive Marginals

A third approach to computing standardized estimates is to generate model-based standardized estimates, often called **predictive marginals**. In what many consider to be a landmark paper, Graubard and Korn (1999) discussed the application and computation of the predictive marginals and their associated standard errors with survey data.

The predictive marginal approach is applicable when the standardization population can be estimated from the entire sample and therefore would not be applicable if one were interested in computing standardized estimates using a standardization population estimated from an outside source. However, the advantages of predictive marginals are numerous:

- Predictive marginals do not require one to cross a set of standardization variables. It takes a model-based approach to computing standardized estimates. However, if one were to include all the interaction terms associated with a set of standardization variables into an appropriate model, then the predictive marginal approach would reproduce the exact same results as direct standardization. For this reason, **predictive marginals is a generalization of the direct standardization approach and is equivalent when an appropriate model is considered and the appropriate**

interaction of a set of variables are used as explanatory variables. This is illustrated in an example below.

- Since predictive marginals do not require one to interact a set of categorical, standardization variables, more main effect and lower order interaction terms can be included in the modeling process. This is very similar to the weight adjustment approach.

- Continuous variables can be included in the predictive marginals. Continuous variables could also be used with the weight adjustment approach depending on the weight adjustment methodology used. But in general, continuous variables could not be used with direct standardization because of the small sample size in cells generated by the continuous variable(s).

- Since this is a model-based approach, one could identify the set of predictors and associated interactions that are statistically significant and only include those in the predictive marginal.

- Predictive marginals can be computed from virtually any type of model including linear regression models, logistic regression models, multinomial logistic models and even proportional hazards models[1].

- Both STATA and the SUDAAN statistical software packages currently compute predictive marginals, the standard error of predictive marginals and the contrasts between marginals in all or most of its modeling procedures, while simultaneously accounting for the complex design features of a study including unequal weighting, stratification and clustering.

- Depending on how well the model under consideration fits the data, in practice one will often find that the predictive marginal approach to deriving standardized estimates will yield more precise estimates than direct standardization.

To illustrate predictive marginals in the context of this discussion, suppose y_i is some outcome measure under consideration for person *i*, d_i is a 0/1 indicator that will equal 1 if person i belongs to a group under consideration and zero otherwise and suppose X_i is some vector of explanatory variables. The vector X_i can have continuous variables and any number of interaction terms. Suppose we use some regression technique to estimate the model parameters in some model defined by:

$$y = f(d, X, \hat{\beta})$$

Where $\hat{\beta}$ are the estimated model parameters. The function *f* can include interaction terms between the d and *X*. Then the weighted predictive marginal for the group under consideration is:

---

[1]See (Aragon-Logan, Brown, Shah and Barnwell, 2004) for a discussion of predictive marginals for Cox's proportional hazards model.

$$\bar{y} = \frac{1}{\sum_i w_i} \sum_i w_i f(d_i = 1, X_i, \hat{\beta})$$

In other words, the predictive marginal for a group is found by computing a prediction from the model for every person, assuming every person belongs to the group and assuming the person retains their values for the other explanatory variables (i.e. the X_i).

It's worth pointing out that predictive marginals can be computed for a specific value of d_i in the case where d_i is a continuous variable.

6. Summary of Data Sources

Examples illustrating the predictive marginal approach, the weight adjustment approach and the direct standardization approach are presented in the next two sections. These examples use data from three sources:

- Data from the **2005 and 2006 National Survey on Drug Use and Health** (NSDUH). This is a national household study conducted by the Substance Abuse and Mental Health Services Administration (SAMHSA) on a yearly basis. The target population for this study is the civilian, noninstitutionalized population aged 12 years old or older that reside in the United States. The study excludes active duty military personnel. The final respondent sample size was 68,308 for 2005 and 67,802 for 2006. Data from the public use file were used for these analyses.

- Data from the **2005 Department of Defense Survey of Health Related Behaviors Among Active Duty Military Personnel** (AD Study). This is study of all active-duty military personnel stationed through-out the world. Data were collected from personnel in the Army, Navy, Marine Corps and Air Force. The final respondent sample size was 16,146. Data from the public use file were used for these analyses.

- Data from the **2006 Departments of Defense Survey of Health Related Behaviors in the Reserve Component** (RC Study). This study was very similar to the AD study. In this study, data were collected from personnel in the U.S. National Guard and Reserves. The total respondent sample size was 18,342.

Data on substance use, physical health and mental health were collected in all three of these studies (as well as numerous other data). The examples presented here use data that were collected from similarly worded questions in the three surveys. Ignoring differences related to varying reference periods (the RC study collected data for 2006 while the AD study collected data from 2005 and the NSDUH collected data for both 2005 and 2006), the three target populations from the three studies nearly represent non-overlapping segments of the U.S. population. The only segment of the population that overlaps is personnel in the U.S. Guard and Reserve. These individuals are included in both the NSDUH and RC study target populations. In order to partially address this, those individuals that indicate they are currently in a "reserve component" were omitted from the NSDUH sample. Data were not collected in NSDUH that would indicate whether a respondent was currently in the National Guard or not.

Data for individuals age 12-17 were also omitted from NSDUH so that the age group representation of the three studies was equivalent to the extent possible. Data for the full-time and/or activated Guard and Reservists were also omitted from the RC study file so that estimates presented in this discussion match those presented elsewhere. After excluding these domains, the final respondent sample size used in these analyses were 37,065 from the 2005 NSDUH, 36,778 from the 2006 NSDUH, 16,146 from the AD Study and 15,212 from the RC Study.

7. Illustrative Examples Using Data from the RC Study

Using data from the 2006 RC Study, consider the outcome measure of whether a person reported an alcohol binge episode in the past 30 days. An alcohol binge episode is defined as the consumption of five or more drinks (four for females) on the same occasion at least once in the past 30 days.

Table 2. Percent of Military Personnel in Reserves Who Reported an Alcohol Binge Episode[a] in the Past 30 Days

	Some College or Less		College Graduate or Higher		Total	
Characteristic	**Population Distribution**	**Estimate**	**Population Distribution**	**Estimate**	**Population Distribution**	**Estimate**
Total	**100.0%**	**44.5%**	**100.0%**	**29.1%**	**100.0%**	**40.4%**
Age Group						
24 and younger	41.7%	52.8%	6.1%	67.5%	32.3%	53.6%
25-34	28.7%	45.3%	26.3%	37.9%	28.0%	43.5%
35 and Older	29.7%	31.9%	67.6%	22.2%	39.6%	27.6%
Race/Ethnicity						
White, Non-Hispanic	67.6%	45.4%	75.6%	32.6%	69.7%	41.8%
African American, Non-Hispanic	14.2%	35.4%	13.1%	12.4%	13.9%	29.7%
Hispanic	12.7%	51.1%	5.9%	26.2%	10.9%	47.5%
Other	5.5%	40.6%	5.4%	22.9%	5.5%	36.1%
Gender						
Male	83.6%	46.3%	79.4%	32.1%	82.5%	42.7%
Female	16.4%	34.9%	20.6%	17.3%	17.5%	29.5%

Note: Estimates exclude full-time and/or activated guard/reservists (Membership Category, Q2; Current Work Status, Q13).

[a]Defined as consumption of five or more drinks (four for females) on the same occasion at least once in the past 30 days.

Source: 2006 Department of Defense Reserve Component Survey (Binge Episode Q29).

Table 2 provides estimates of total Reserve personnel who reported an alcohol binge episode in the past 30 days by education and demographic variables. This table also presents estimates of the population distribution. For example, 41.7% of those that reported some college or less were 24 years old and younger. In comparison, 6.1% of those that reported being a college graduate or higher were 24 years old and younger. Within these two groups, the estimate of Reserve personnel who reported an alcohol binge episode were 52.8% for those with some college or less and who were 24 years old and younger. The estimate of Reserve personnel who reported an alcohol binge episode was 67.5% for those who were college graduates or higher and 24 years old and younger.

Although many of the differences may not be statistically significant, the table suggests the alcohol binge rate is less for those who were a college graduate or higher, across all levels of the demographic variables exhibited in **Table 2** except for the 24 and younger age group. The table also suggests that (not surprisingly) the population distribution is fairly different across the age group levels between those with some college or less and those who are a college graduate or higher. The population distribution by race and gender is not extremely different between these education groups.

Table 3. Percent of Military Personnel in Reserves Who Reported an Alcohol Binge Episode[a] in the Past 30 Days, Direct Estimates, Standardized Estimates and Predictive Marginals

Estimates	Some College or Less	College Graduate or Higher
Direct Estimates		
Estimate	44.5%	29.1%
Standard Error	1.7%	2.1%
Direct Standardization Estimate		
Estimate	42.1%	41.0%
Standard Error	1.4%	2.1%
Weight Adjustment Estimate		
Estimate	42.1%	40.7%
Standard Error	1.6%	3.6%
Predictive Marginals (All Interactions)		
Estimate	42.1%	41.0%
Standard Error	1.6%	2.7%
Predictive Marginals (Main Effects Only)		
Estimate	41.9%	35.8%
Standard Error	1.6%	2.2%

Note: Estimates exclude full-time and/or activated guard/reservists (Membership Category, Q2; Current Work Status, Q13). Adjusted estimates have been standardized by age group, race/ethnicity, and gender to the total Reserve military personnel distribution.

[a]Defined as consumption of five or more drinks (four for females) on the same occasion at least once in the past 30 days.

Source: 2006 Department of Defense Reserve Component Survey (Binge Episode Q29).

Table 3 compares the direct survey estimate of an alcohol binge episode with an estimate obtained via direct standardization, obtained using an adjusted standardized weight, and one obtained via the predictive marginal approach. Specifically, in **Table 3**:

- The **Direct Estimates** are the unadjusted, weighted survey estimates.

- The **Direct Standardization Estimates** are the estimates obtained when the two education groups are standardized by the cross-classification of age group, race/ethnicity and gender.

- The **Weight Adjustment Estimates** are standardized estimates obtained by adjusting the sample weight within the two education groups by the cross-classification of age group, race/ethnicity and gender.

- The **Predictive Marginals (All interactions)** are the predictive marginals associated with the two education groups that resulted from a logistic model with all main effects and interactions terms of education, age group, race/ethnicity and gender included in the model. As can be seen from this table, and as noted in **Section 5**, when all the interaction terms are included in the model then the predictive marginals are equivalent to the direct standardization estimates. Notice in this case though that the standard errors of the predictive marginals are slightly higher, which is likely due to a model that does not fit the data very well.

- The **Predictive Marginals (Main Effects Only)** are the predictive marginals with only the main effects of age group, race/ethnicity and gender included in the model.

In this example, the direct standardization estimates for the college graduate or higher group (41.0%) is quite a bit larger than the direct estimate for the same group (29.1%). This indicates that a good portion of the difference in the direct estimates between the some college or less group (44.5%) and the college graduate or higher group (29.1%) is due to differences in the distribution of these populations across the demographic variables considered in this example. The predictive marginal (main effects only) lie in between the direct estimate and the direct standardization estimate and has a slightly larger standard error. The estimates created using the adjusted weights are similar to the direct standardized estimate but have slightly larger estimated standard errors. The larger standard error estimate is likely because creating standardize estimates using an adjusted weight and obtaining a typical variance estimate of a mean (or percent) ignores the fact that the estimates have been standardized and the variance estimate is therefore only accounting for a change in the unequal weighting effect.

As noted in **Section 3**, one benefit of the predictive marginal approach is that more significant main effect and lower order interaction terms can be included in the standardization process. To illustrate this see **Table 4** below. In this illustration, the percent of military personnel who reported an alcohol binge episode in the past 30 days is displayed by Reserve Component.

The categorical variables considered in the direct standardization estimates and the predictive marginals were gender, age group, enlisted/officer indicator, married/other, education and race/ethnicity. These six categorical variables crossed with Reserve component yields a total of 2,304 cells. Because of this large number of cells, in this study we found that 78% of the cells had a sample size of 5 or less and 51% of the cells had a sample size of zero. A small sample size can adversely affect the direct standardization estimates and can seriously bias the resulting variance estimate.

Despite the large number of empty cells in the cross classification, **Table 4** suggests the direct standardization estimates are relatively close to the direct estimates and to the predictive marginals, at least for some Reserve components. In this example, the predictive marginals were derived from a fitted logistic model that included the main effects of the categorical variables only. All effects were statistically significant predictors except for the enlisted/office indicator and education.

Table 4 also shows that the for some Reserve components, the predicated marginal estimates are significantly different than the direct estimate and the direct standardization estimate. For example, for the Marine Corps Reserve Component, the predictive marginal is 19.8% less than the direct estimate and 13.6% less than the direct standardization estimate. Higher item and total person nonresponse were exhibited among the Marine Corp Reserves in this study, therefore the adverse effects of a large number of zero or near zero cells in the cross classification will be more pronounced for this Reserve group. From **Table 4**, note that for the Marine Corps, 78.6% of the cells in the cross-classification used in the direct standardization estimates had a sample size of zero and 93.0% had a sample size less than 5. Because of the large number of cells with a small or zero sample size, the predictive marginal is likely a more stable and less biased estimate for the Marines.

Table 4. Percent of Military Personnel in Reserves Who Reported an Alcohol Binge Episode[a] in the Past 30 Days, Direct Estimates, Standardized Estimates and Predictive Marginals, by Reserve Component

Estimates	Army National Guard	Army Reserve	Naval Reserve	Air National Guard	Air Force Reserve	Marine Corps Reserve
Direct Estimates						
Estimate	47.4%	37.7%	26.6%	29.5%	31.0%	59.3%
Standard Error	2.8%	2.9%	0.9%	3.6%	1.2%	3.0%
Direct Standardization Estimates						
Estimate	46.0%	39.9%	33.4%	33.8%	37.9%	55.0%
Standard Error	2.4%	1.8%	1.7%	2.2%	1.4%	1.6%
Percent of Cells With Sample Size = 0	46.9%	48.2%	43.5%	49.7%	39.6%	78.6%
Percent of Cells With Sample Size $\leq$ 5	77.3%	81.0%	71.6%	80.2%	65.6%	93.0%
Predictive Marginals (Main Effects Only)						
Estimate	45.0%	38.2%	32.2%	32.2%	37.0%	47.5%
Standard Error	2.7%	2.6%	1.2%	3.8%	1.5%	2.6%
Percent Difference Between Predictive Marginal and:						
Direct Estimate	-5.19%	1.39%	21.28%	9.27%	19.26%	-19.83%
Direct Standardization Estimate	-2.25%	-4.21%	-3.53%	-4.54%	-2.59%	-13.62%

Note: Estimates exclude full-time and/or activated guard/reservists (Membership Category, Q2; Current Work Status, Q13). Adjusted estimates have been standardized by age group, race/ethnicity, enlisted/officer indicator, married/other, education and gender to the total Reserve military personnel distribution.

[a]Defined as consumption of five or more drinks (four for females) on the same occasion at least once in the past 30 days.

Source: 2006 Department of Defense Reserve Component Survey (Binge Episode Q29).

8. Comparing Standardized Military and Civilian Estimates

Table 5 and **Table 6** compares the direct estimates, direct standardized estimates, weight adjustment estimates (i.e. standardized estimates done via a weight adjustment), and predictive marginals for several outcome measures. Estimates in these tables were computed using data from the 2005 AD Study, the 2006 RC Study, and the 2005 and

2006 NSDUH studies. The direct standardized, weight adjustment and predictive marginal estimates in these tables were computed from a pooled Civilian/Active Duty/Reserve dataset.

Table 5. Percents and Standard Errors for Selected Dependent Measures Comparing Civilian, Total Active Duty and Total Reserves Estimates, Aged 18 or Older

Estimates	Civilian	Active Duty	Reserves
Cigarette Use in Past Month			
Direct Estimate	26.7 (0.3)	32.2 (1.1)	23.7 (1.3)
Direct Standardized Estimate	26.8 (0.3)	29.4 (1.4)	24.5 (1.6)
Weight Adjustment Estimate	26.8 (0.3)	27.2 (1.4)	25.5 (1.7)
Predictive Marginal	26.8 (0.3)	25.5 (0.8)	19.7 (1.2)
Marijuana Use in Past Month			
Direct Estimate	6.0 (0.1)	1.3 (0.2)	3.0 (0.5)
Direct Standardized Estimate	6.0 (0.1)	1.1 (0.4)	2.3 (0.4)
Weight Adjustment Estimate	6.0 (0.1)	0.5 (0.2)	1.9 (0.5)
Predictive Marginal	6.0 (0.1)	0.6 (0.1)	1.5 (0.2)
Illicit Drug Use in Past Month[a]			
Direct Estimate	8.0 (0.1)	5.0 (0.4)	6.6 (0.8)
Direct Standardized Estimate	8.0 (0.1)	3.9 (0.5)	6.8 (0.8)
Weight Adjustment Estimate	8.0 (0.1)	3.7 (0.4)	6.3 (0.8)
Predictive Marginal	8.0 (0.1)	2.6 (0.2)	3.7 (0.4)
Heavy Drinking in the Past Month[b]			
Direct Estimate	7.2 (0.1)	18.5 (1.0)	16.7 (0.9)
Direct Standardized Estimate	7.3 (0.1)	10.8 (0.5)	13.4 (1.5)
Weight Adjustment Estimate	7.2 (0.1)	10.2 (0.8)	11.5 (0.8)
Predictive Marginal	7.3 (0.1)	9.8 (0.5)	9.1 (0.5)
Alcohol Dependence in Past Year[c]			
Direct Estimate	3.4 (0.1)	2.9 (0.3)	3.1 (0.3)
Direct Standardized Estimate	3.4 (0.1)	2.6 (0.3)	2.7 (0.3)
Weight Adjustment Estimate	3.4 (0.1)	1.6 (0.4)	2.5 (0.5)
Predictive Marginal	3.4 (0.1)	1.7 (0.2)	1.9 (0.2)
Driving Under the Influence of Alcohol in Past Year			
Direct Estimate	13.7 (0.2)	13.3 (0.6)	15.4 (0.9)
Direct Standardized Estimate	13.8 (0.2)	8.9 (0.8)	11.0 (0.7)
Weight Adjustment Estimate	13.8 (0.2)	7.8 (0.6)	11.1 (0.8)
Predictive Marginal	13.8 (0.2)	7.4 (0.4)	9.0 (0.6)

Note: Percentages displayed above based on 2005 AD, 2006 RC, and 2005-2006 NSDUHs. Estimates in parentheses are estimated standard errors. Reserve estimates exclude full-time and/or activated guard/reservists. Weight Adjustment Estimate refers to standardized estimates computed via a weight adjustment. Estimates standardized over gender, age, marital status, race/ethnicity, and education level. The Weight Adjustment Estimate and the Predictive Marginal accounted for the main effect of the standardization variables only (i.e. no interactions terms were included).

[a]Defined as the nonmedical use of marijuana, cocaine, hallucinogens, amphetamines/stimulants, tranquilizers, sedatives, heroin, analgesics, or inhalants.

[b]In the RC and AD studies, this is defined as consuming five or more drinks on the same occasion at least once a week in the past 30 days. In the NSDUH studies, this is defined as drinking five or more drinks at least once a week in past 30 days.

[c]In the RC and AD studies, this is defined as having experience four or more alcohol dependence symptoms at any time during the year. In the NSDUH studies, alcohol dependence is based on definition found in the 4th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV).

The standardized distribution considered in this analysis was derived from the pooled Civilian/Active Duty/Reserve population distribution. The categorical variables considered in the direct standardization estimates, predictive marginals, and the weight adjustment estimates were gender, age group, married/other, education, and race/ethnicity. Only the main effect of these variables was considered in the standardization process for the weight adjustment estimates and the predictive marginals.

Table 5 shows that, in general, the standardized estimates for the Active Duty and Reserves tend to be noticeably less than the direct survey estimates for these same groups – at least for most of the measures depicted in this table. Consider, for example, the outcome measure of whether a person drove under the influence of alcohol in the past year. Looking at **Table 5**, the direct estimates show that the rates are similar for Civilians, Active Duty military personnel, and military personnel in the Reserve components at 13.7%, 13.3%, and 15.4% respectively. However, after adjusting for some of the sociodemographic differences between the three groups, the predictive marginal estimates show that all three groups are significantly different from each other (Civilians at 13.8%, Active Duty at 7.4%, and Reserves at 9.0%). Since the estimates for both military groups decreased, this indicates that these groups have a higher percentage of people in the sociodemographic groups that exhibit higher rates for driving under the influence – compared to the total population.

Table 6 is similar to **Table 5** and compares civilians with the four service components of the military (Army, Navy, Air Force, and Marine Corps). The military estimates are a combination of the Active Duty and Reserve Study estimates. Consider, for example, the outcome measure of alcohol dependence in the past year and the rates for these five groups. Comparing the direct estimates for the civilians to the four military groups, only Air Force (1.1%) had a significantly different rate than the Civilians (3.4%). Estimates for the other three military service branches were similar to the Civilian rate (Army – 3.9%, Navy – 2.5 %, and Marine – 4.6%). However, when looking at the predictive marginals, all the components are significantly different than the Civilian population (Civilian – 3.4%, Army – 2.2 %, Navy – 1.6%, Marine – 2.2%, and Air Force – 0.8%). Since the adjusted estimates are all less than the direct estimates for the four components of the military, this again suggests the military has a higher percentage of people in the sociodemograhic groups that exhibit higher rates of alcohol dependence compared to the civilian population.

In examining the estimates that were created using the adjusted weight, the estimates for the Marine Corps for several of the outcomes are quite different than the other three estimates and the standard error is larger (this can be seen, for example, with the illicit drug use and marijuana use in past month). For example, the direct estimate for past month marijuana use for the Marine Corps is 2.1%. The direct standardized estimate is 1.7% and the predictive marginal is .7%. All the standard errors for these three estimates are less than 1%. The estimate obtained using the adjusted weight is 3.5% with a standard error of 2.1%. This suggests that the current Weight Adjustment estimate may not be too good for this domain. Using a different set of variables in the weight adjustment would likely address this.

**Table 6. Selected Dependent Measures
Comparing Civilian, Army, Navy, Air Force and Marine Corps
Estimates, Aged 18 or Older**

Estimates	Civilian	Army	Navy	Air Force	Marine Corps
Cigarette Use in Past Month					
Direct Estimate	26.7 (0.3)	31.5 (1.6)	28.8 (1.9)	21.5 (1.4)	34.5 (1.9)
Direct Standardized Estimate	26.8 (0.3)	27.6 (1.5)	26.9 (1.9)	21.5 (1.5)	22.9 (1.4)
Weight Adjustment Estimate	26.8 (0.3)	29.0 (2.0)	24.9 (1.8)	21.2 (1.4)	23.1 (2.1)
Predictive Marginal	26.8 (0.3)	24.8 (1.4)	23.7 (1.2)	19.4 (0.9)	23.6 (1.4)
Marijuana Use in Past Month					
Direct Estimate	6.0 (0.1)	3.0 (0.4)	1.4 (0.5)	0.6 (0.1)	2.1 (0.6)
Direct Standardized Estimate	6.0 (0.1)	1.5 (0.2)	0.6 (0.2)	0.7 (0.3)	1.7 (0.2)
Weight Adjustment Estimate	6.0 (0.1)	1.4 (0.5)	0.4 (0.1)	0.5 (0.2)	3.5 (2.1)
Predictive Marginal	6.0 (0.1)	1.3 (0.2)	0.7 (0.3)	0.3 (0.1)	0.7 (0.2)
Illicit Drug Use in Past Month[a]					
Direct Estimate	8.0 (0.1)	7.8 (0.6)	4.4 (0.9)	2.7 (0.3)	6.2 (0.9)
Direct Standardized Estimate	8.0 (0.1)	5.9 (0.8)	4.8 (0.8)	2.3 (0.3)	5.7 (0.6)
Weight Adjustment Estimate	8.0 (0.1)	6.5 (0.9)	4.0 (0.3)	2.7 (0.4)	7.2 (2.1)
Predictive Marginal	8.0 (0.1)	3.9 (0.3)	2.5 (0.5)	1.7 (0.1)	2.5 (0.4)
Heavy Drinking in the Past Month[b]					
Direct Estimate	7.2 (0.1)	21.4 (1.2)	15.1 (1.2)	10.0 (0.9)	26.5 (1.2)
Direct Standardized Estimate	7.3 (0.1)	16.0 (0.9)	8.3 (0.7)	9.1 (1.0)	15.0 (1.2)
Weight Adjustment Estimate	7.2 (0.1)	11.9 (0.9)	8.0 (0.5)	6.7 (0.9)	17.9 (3.5)
Predictive Marginal	7.3 (0.1)	11.2 (0.6)	8.7 (0.7)	6.0 (0.5)	11.3 (0.5)
Alcohol Dependence in Past Year[c]					
Direct Estimate	3.4 (0.1)	3.9 (0.4)	2.5 (0.5)	1.1 (0.2)	4.6 (0.7)
Direct Standardized Estimate	3.4 (0.1)	2.2 (0.4)	1.4 (0.4)	1.9 (0.2)	2.5 (0.3)
Weight Adjustment Estimate	3.4 (0.1)	2.3 (0.6)	1.1 (0.3)	1.3 (0.5)	1.2 (0.3)
Predictive Marginal	3.4 (0.1)	2.2 (0.2)	1.6 (0.3)	0.8 (0.1)	2.2 (0.3)
Driving Under the Influence of Alcohol in Past Year					
Direct Estimate	13.7 (0.2)	14.6 (0.9)	14.2 (0.8)	10.4 (0.9)	21.5 (1.2)
Direct Standardized Estimate	13.8 (0.2)	9.5 (0.8)	9.4 (0.7)	7.4 (0.4)	15.8 (1.6)
Weight Adjustment Estimate	13.8 (0.2)	9.7 (1.0)	8.9 (0.6)	7.4 (0.6)	14.0 (2.1)
Predictive Marginal	13.8 (0.2)	8.1 (0.6)	8.8 (0.5)	6.0 (0.5)	10.6 (0.7)

Note: Percentages displayed above based on 2005 AD, 2006 RC, and 2005-2006 NSDUHs. Estimates in parentheses are estimated standard errors. Estimates for active duty personnel and reserve/guard have been combined into the four services. Predictive Marginal estimates were derived from models with main effect terms only.

Reserve estimates exclude full-time and/or activated guard/reservists (Membership Category, Q2; Current Work Status, Q13

Estimates standardized over gender, age, marital status, race/ethnicity, and education level.

[a]Defined as the nonmedical use of marijuana, cocaine, hallucinogens, amphetamines/stimulants, tranquilizers, sedatives, heroin, analgesics, or inhalants.

[b]In the RC and AD studies, this is defined as consuming five or more drinks on the same occasion at least once a week in the past 30 days. In the NSDUH studies, this is defined as drinking five or more drinks at least once a week in past 30 days.

[c]In the RC and AD studies, this is defined as having experience four or more alcohol dependence symptoms at any time during the year. In the NSDUH studies, alcohol dependence is based on definition found in the 4th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV).

One interesting thing to note about the estimates displayed in **Table 5** and **Table 6** - throughout these tables the direct estimate, direct standardized estimate, weight adjustment estimate and the predictive marginal estimate for the Civilian population do not change much. This is due to the fact that the Civilians make up about 99.2% of the pooled population so the standardization weight distribution of the Civilian population is obviously extremely close the pooled Civilian/Active Duty/Reserve population distribution.

9. Concluding Remarks

This paper presented a discussion of the use of predictive marginals with survey data and noted that it was an appealing alternative to both the direct standardization approach and the weight adjustment approach when one is interested in comparing standardized estimates.

The predictive marginal approach is a generalization of the direct standardization approach and will yield equivalent results when all interactions associated with a set of categorical variables are used as independent variables in the model. Compared to the direct standardization approach, the predictive marginal approach allows one to use a larger number of variables in the standardization process (by only including main effects and lower-order interactions), one can use continuous variables, one can easily test the significance of variables in the standardization process and several software packages have routines built in them that allow one to obtain predictive marginal estimates and their associated standard errors (e.g. Stata and SUDAAN).

The weight adjustment approach shares many of the same advantages as the predictive marginal approach to deriving standardized estimates, particularly when one computes weight adjustments from a model such as the generalized exponential model. The weight adjustment approach has one distinct disadvantage though. Most software packages do not properly account for the standardization weight adjustment when computing variance estimates for means or percents.

References

Bray, R. et el (2006). *2005 Department of Defense Survey of Health Related Behaviors Among Active Duty Military Personnel*. Report prepared for the Assistant Secretary of Defense (Health Affairs) by RTI International.

Kalton, G. (1968). Standardization: A technique to control for extraneous variables. *Applied Statistics*, *23*, 118-136.

Konijn, H.S. (1973). *Statistical theory of sample survey design and analysis*. London: North-Holland.

Korn, E. L., and B. I. Graubard (1999). Predictive Margins for Survey Data. *Biometrics* 55, 652-659.

Little, R. J. A. (1982). Direct standardization as a tool for teaching linear models for unbalanced data. *American Statistician, 36*(1), 38-43.

Research Triangle Institute (2008) *SUDAAN Language Manual, Release 10.0*. Research Triangle Park, NC: Research Triangle Institute. http://www.rti.org/sudaan/

StataCorp (2009), Producers and Distributor of *Stata 11, Statistical Software for Professionals*, 2009. http://www.stata.com/

Witt, M. B. (2009) Overview of Software that Will Produce Sample Weight Adjustments, *Proceedings of the Section on Survey Research Methods of the American Statistical Association*.