**Guidance Document for Use of Workplace Health in America Public Data Files**

**Study Design Description**: The Workplace Health in America (WHA) Survey gathered information from a cross-sectional, nationally representative sample of US worksites. The sample was drawn from the Dun & Bradstreet (D&B) database of all private and public employers in the United States with at least 10 employees. Like previous national surveys, the worksite served as the sampling unit rather than the companies or firms to which the worksites belonged. Worksites were selected using a stratified simple random sample (SRS) design, where the primary strata were ten multi-state regions defined by the Centers for Disease Control and Prevention (CDC), plus an additional stratum containing all hospital worksites. The hospital worksites were assigned to their own stratum to ensure we could obtain sufficient sample size for hospital-specific estimates. Within each CDC region, secondary strata were constructed using the cross-classification of eight worksite size categories (i.e., number of employees at the worksite) and six combined industry groups based on the North American Industry Classification System (NAICS) sectors. Table 1 includes descriptions of the three stratification characteristics. The total sample size for each CDC region was proportionally allocated across the secondary strata (size category by industry group).

Table 1. Description of Sampling Stratification Variables (file variable names are in bold all caps font)

| Key Analysis Group | Description |
|---|---|
| Multi-State Region (n=10) **CDC_REGION** | ▪ Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont<br>▪ New Jersey, New York<br>▪ Delaware, District of Columbia, Maryland, Pennsylvania, Virginia, West Virginia<br>▪ Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee<br>▪ Illinois, Indiana, Michigan, Minnesota, Ohio, Wisconsin<br>▪ Arkansas, Louisiana, New Mexico, Oklahoma, Texas<br>▪ Iowa, Kansas, Missouri, Nebraska<br>▪ Colorado, Montana, North Dakota, South Dakota, Utah, Wyoming<br>▪ Arizona, California, Hawaii, Nevada<br>▪ Alaska, Idaho, Oregon, Washington |
| Worksite Size Groups (i.e., number of employees at worksite) (n=8) **SIZE** | ▪ 10–24 employees<br>▪ 25–49 employees<br>▪ 50–99 employees<br>▪ 100–249 employees<br>▪ 250–499 employees<br>▪ 500–749 employees<br>▪ 750–999 employees<br>▪ 1,000 or more employees |
| Industry Group (n=6) **INDUSTRY** | ▪ Agriculture, Forestry, Fishing, and Hunting; Mining, Quarrying, and Oil and Gas Extraction; Utilities; Construction; Manufacturing<br>▪ Wholesale Trade; Retail Trade; Transportation and Warehousing<br>▪ Arts, Entertainment, and Recreation; Accommodation and Food Services; Other Services (except Public Administration)<br>▪ Information; Finance and Insurance; Real Estate and Rental and Leasing; Professional, Scientific, and Technical Services; Management of Companies and Enterprises; Administrative and Support and Waste Management and Remediation Services<br>▪ Educational Services; Health Care and Social Assistance [excluding hospital worksites]<br>▪ Public Administration (includes local, state, and federal government) |

| | |
|---|---|
| Hospital Worksites | ▪ Worksites classified into one of three NAICS industries. 622110: General Medical and Surgical Hospitals, 622210: Psychiatric and Substance Abuse Hospitals, or 622310: Specialty (except Psychiatric and Substance Abuse) Hospitals |

For analytic purposes, collapsed versions of the region and size stratification variables were created to produce estimates for these domains. The collapsed region variable (5 levels) is available on the public use data file and is called REGION. Definitions summarizing the states included in each category of the 5-level REGION variable are available in the corresponding codebook. The collapsed size variable (6 levels) is not available on the data file but can be constructed using the variable called SIZE. The collapsed size variable retains categories 1–5 from the SIZE variable but combines categories 6–8 into a "500+ employees" category. This collapsed definition is recommended for analysis. Some example SAS code to create the 6-level size variable is provided below.

```
if size ge 6 then size6=6;
  else size6=size;
```

note: "ge" represents "greater than or equal"

**Variance Estimation Structure**: To account for the complex sampling design of this study, analyses should be conducted using statistical software that can properly estimate variances for complex survey designs (e.g., SAS, SUDAAN, Stata). The appropriate variance estimation structure for this design is specified in the variable VARSTRATA.

**Analysis Weights**: Because this survey represents the population of US worksites with 10 or more employees, analysis weights should be used to accurately reflect the target population. The appropriate analysis weight variable is called FINALWT_WORKSITE.

**Appropriate Degrees of Freedom (DoF)**: Because this design is a stratified SRS, the appropriate degrees of freedom for testing will change depending on the analysis domain (i.e., the DOF vary for subgroup analyses by REGION, CDC_REGION, SIZE6 and INDUSTRY).

Example SAS code for properly assigning the degrees of freedom (for calculation of confidence intervals or for "between group" testing) is provided below. When conducting tests between groups, it is recommended to use the smallest DoF among the groups in the analysis as conservative approach.

```
if region=1 then dof=459;
else if region=2 then dof=457;
else if region=3 then dof=529;
else if region=4 then dof=567;
else if region=5 then dof=471;
else if cdc_region=1 then dof=189;
else if cdc_region=2 then dof=142;
else if cdc_region=3 then dof=226;
else if cdc_region=4 then dof=312;
else if cdc_region=5 then dof=294;
else if cdc_region=6 then dof=245;
else if cdc_region=7 then dof=377;
else if cdc_region=8 then dof=281;
```

```
else if cdc_region=9 then dof=191;
else if cdc_region=10 then dof=308;
else if industry=1 then dof=477;
else if industry=2 then dof=275;
else if industry=3 then dof=393;
else if industry=4 then dof=391;
else if industry=5 then dof=496;
else if industry=6 then dof=205;
else if industry=7 then dof=337;
else if size6=1 then dof=1114;
else if size6=2 then dof=594;
else if size6=3 then dof=306;
else if size6=4 then dof=207;
else if size6=5 then dof=96;
else if size6=6 then dof=194;
else dof=2574;
```

**Suppression criteria for descriptive statistics**: The following criteria are recommended for determining reliability of estimates. These criteria are intended to be used inclusively (i.e., all three criteria should be met to assess reliability).

1) The unweighted denominator for proportions is equal to or greater than 50.
2) The relative standard error (RSE) defined as the ratio of the standard error over the mean is less than or equal to 30%.
3) Caution is advised for interpreting estimate values of zero (0) or 100%. Though these represent valid estimates, these values are less likely to occur in the population of interest and may instead result from homogeneity in the observed sample.

**Comma-separated values (.CSV) and Formats file:** For users interested in data file formats other than Statistical Analysis Software (SAS), there are two options available: (1) a delimited .csv file, for which the delimiter is a "~" (tilde) and (2) format statements, which may be used to input data into another type of statistical software. The format statements provide descriptions for all possible variable values and are attached at the end of this document. When importing the .csv file, users should be aware that the first row contains the variable names.

**Example Code for Producing Estimates in SUDAAN:** The code below provides an example of how to specify the design parameters in SUDAAN statistical software to appropriately estimate a proportion and standard error that accounts for the complex sample design. This code assumes that the WHA public use data have already been imported into SAS.

```
proc descript data=<WHA data file name> design=strwr filetype=sas;
nest varstrata;
weight finalwt_worksite;
subgroup size6 region cdc_region industry;
levels     6          5          10         7;
var  <names of variable of interest: coded as 0/1 for proportions***>
tables size6 region cdc_region industry ;
SETENV DECWIDTH=6 COLWIDTH=18;
```

```
print nsum wsum mean semean total setotal;
output nsum wsum total mean semean / replace filename=out1; run;
```

***

1 = represents the outcome of interest to estimate
0 = represents all other eligible outcomes
For outcomes not of interest for analysis (e.g., Don't Know, Refusal responses) leave blank (or "." = sas missing value)

**Format Statements**

```
Proc format;
value cr 1="CDC Region 1: CT ME MA NH RI VT"
      2="CDC Region 2: NY NJ"
      3="CDC Region 3: DE MD PA VA WV DC"
      4="CDC Region 4: AL FL GA KY MS NC SC TN"
      5="CDC Region 5: IL IN OH MI MN WI"
      6="CDC Region 6: AR LA NM OK TX"
      7="CDC Region 7: IA KS MO NE"
      8="CDC Region 8: CO MT ND SD UT WY"
      9="CDC Region 9: AZ CA HI NV"
      10="CDC Region 10: AK ID OR WA";

value rg 1="Region 1: ME VT CT RI MA NH NY PA NJ MD DC"
      2="Region 2: WV VA NC SC GA FL AL TN KY MS LA AR"
      3="Region 3: MN IA MO WI IL MI IN OH"
      4="Region 4: ID MT ND SD NE KS CO WY UT OK TX NM"
      5="Region 5: AZ WA OR CA NV AK HI";

value ind 1="Industry Category 1: NAICS Sectors: 11, 21, 22, 23, 31-33"
       2="Industry Category 2: NAICS Sectors: 42, 44-45, 48-49"
       3="Industry Category 3: NAICS Sectors: 71, 72, 81"
       4="Industry Category 4: NAICS Sectors: 51, 52, 53, 54, 55, 56"
       5="Industry Category 5: NAICS Sectors: 61, 62 (excluding hospital worksites)"
       6="Industry Category 6: NAICS Sectors: 92"
       7="Industry Category 7: Hospital worksites (NAICS6 = 622110, 622210, 622310)";

value sz 1="Size Category 1: 10-24"
      2="Size Category 2: 25-49"
      3="Size Category 3: 50-99"
      4="Size Category 4: 100-249"
      5="Size Category 5: 250-499"
      6="Size Category 6: 500-749"
      7="Size Category 7: 750-999"
      8="Size Category 8: 1,000+";

value oca 1= "Human Resources and/or Benefits"
       2= "Health Promotion/Fitness/Wellness"
       3= "Safety"
       4= "Medical"
       5= "PR/Marketing"
       6= "Finance"
       7= "Other"
       97="Don't know"
       98="Refusal"
       99="Blank";

value ocb 1= "For profit, public"
```

```
        2= "For profit, private"
        3= "Non-profit"
        4= "State or local government"
        5= "Federal government"
        6= "Other"
        97="Don't know"
        98="Refusal"
        99="Blank";

value fa 1= "Full insurance coverage offered"
       2= "Partial insurance coverage offered"
       3= "No insurance coverage offered"
       96="Legitimate skip"
       97="Don't know"
       98="Refusal"
       99="Blank";

value fb 1= "Larger"
       2= "Smaller"
       3= "About the same"
       96="Legitimate skip"
       97="Don't know"
       98="Refusal"
       99="Blank";

value fc 1= "Yes"
       2= "No"
       5= "Logically Assigned No (HP1=No)"
       9= "Logically Assigned No (All HPR2_3a-d = No)"
       95="Supplemental Questionnaire Non-participant"
       96="Legitimate skip"
       97="Don't know"
       98="Refusal"
       99="Blank";

value fcb 1= "Yes"
        2= "No"
        5= "Logically Assigned No (HP1=No)"
        9= "Logically Assigned No (HS1 Screening = No)"
        96="Legitimate skip"
        97="Don't know"
        98="Refusal"
        99="Blank";

value fcc 1= "Yes"
        2= "No"
        5= "Logically Assigned No (HP1=No)"
        9= "Logically Assigned No (WL1 = No)"
```

```
        95="Supplemental Questionnaire Non-participant"
        96="Legitimate skip"
        97="Don't know"
        98="Refusal"
        99="Blank";

value fcd 1= "Yes"
        2= "No"
        5= "Logically Assigned No (HP1=No)"
        9= "Logically Assigned No (OSH8_S = No)"
        95="Supplemental Questionnaire Non-participant"
        96="Legitimate skip"
        97="Don't know"
        98="Refusal"
        99="Blank";

Value fd            1= "The employer"
                2= "The health insurance plan"
                        3= "A different third party vendor"
                        96="Legitimate skip"
           97="Don't know"
           98="Refusal"
           99="Blank";

Value fe            1= "1-25%"
                        2= "26-50%"
                        3= "51-75%"
                        4= "76-100%"
        96="Legitimate skip"
           97="Don't know"
           98="Refusal"
           99="Blank";

Value ff 1= "Given their results"
                        2= "Given results and provided feedback and education"
                        3= "Neither of the above"
                        96="Legitimate skip"
           97="Don't know"
           98="Refusal"
           99="Blank";

Value fg            1= "Less than 1 year"
                        2= "1-2 years"
                        3= "3 to 5 years"
        4= "6 to 9 years"
        5= "10 years or more"
                        96="Legitimate skip"
           97="Don't know"
```

```
        98="Refusal"
        99="Blank";

Value fh        1= "Staff employed by our organization"
            2= "Staff employed by our insurance provider"
                3= "Staff employed by a vendor/other third party provider"
                96="Legitimate skip"
        97="Don't know"
        98="Refusal"
        99="Blank";

Value fi 1= "Yes, one committee, includes both health promotion and safety"
            2= "Yes, one committee, includes health promotion but not safety"
            3= "Yes, one committee, includes safety but not health promotion"
      4= "Yes, there are two separate committees"
      5= "No committee addressing health promotion or safety"
            96="Legitimate skip"
        97="Don't know"
        98="Refusal"
        99="Blank";

Value fj 1= "Not representative at all"
            2= "Somewhat representative"
      3= "Mostly representative"
      4= "Entirely representative"
            96="Legitimate skip"
        97="Don't know"
        98="Refusal"
        99="Blank";

Value fk        1= "No annual budget"
                2= "Less than $1000"
                3= "$1001-5000"
                4= "$5001-10,000"
      5= "$10,001-15,000"
      6= "$15,001-20,000"
      7= "$20,001-50,000"
      8= "$50,001-100,000"
      9= "$100,000-500,000"
      10= "Over $500,000"
      96="Legitimate skip"
        97="Don't know"
        98="Refusal"
        99="Blank";

Value fl 1= "We will spend more"
            2= "We will spend less"
      3= "We will spend about the same amount"
```

```
            96="Legitimate skip"
          97="Don't know"
          98="Refusal"
          99="Blank";

Value fm          1= "Awareness or informational"
            2= "Skill-building"
       3= "Both informational and skill-building"
            96="Legitimate skip"
          97="Don't know"
          98="Refusal"
          99="Blank";

Value fn          1= "Offered mostly by employer"
                    2= "Offered mostly by health plan"
                      3= "Offered mostly by outsourced vendor"
       4= "Offered by combined efforts of employer, health plan and/or vendor"
                      96="Legitimate skip"
          97="Don't know"
          98="Refusal"
          99="Blank";

Value fo          1= "1-25%"
                      2= "26-50%"
                      3= "51-75%"
                      4= "More than 75%"
       96="Legitimate skip"
          97="Don't know"
          98="Refusal"
          99="Blank";

Value fp          1= "Not allowed in ANY public areas"
                    2= "Allowed in some public areas"
                      3= "Allowed in ALL public areas"
       4= "Not applicable"
                      96="Legitimate skip"
          97="Don't know"
          98="Refusal"
          99="Blank";

Value fq          1= "Not allowed in ANY work areas"
                    2= "Allowed in some work areas"
                      3= "Allowed in ALL work areas"
       4= "Not applicable"
                      96="Legitimate skip"
          97="Don't know"
          98="Refusal"
          99="Blank";
```

```
Value fr 1= "Onsite"
                2= "Offsite"
                3= "Both"
                96="Legitimate skip"
        97="Don't know"
        98="Refusal"
        99="Blank";

Value fs 1= "Yes, onsite"
                2= "Yes, offsite"
                3= "Yes, onsite and offsite"
                4= "No"
                5= "Logically Assigned No (HP1=No)"
                96="Legitimate skip"
        97="Don't know"
        98="Refusal"
        99="Blank";

Value ft 1= "Onsite in-person"
                2= "By phone counseling"
                3= "Online program"
                4= "Multiple ways"
                96="Legitimate skip"
        97="Don't know"
        98="Refusal"
        99="Blank";

Value fu        1= "Yes, we have offered incentives"
                2= "No, but plan to offer in the next 12 months"
                3= "No, and have no plans to offer them in the next 12 months"
                5= "Logically Assigned No Plans Offered (HP1=No), Next 12 Month plans unknown"
                96="Legitimate skip"
        97="Don't know"
        98="Refusal"
        99="Blank";

Value fv        1= "Not at all effective"
                2= "Somewhat effective"
                3= "Effective"
                4= "Extremely effective"
                96="Legitimate skip"
        97="Don't know"
        98="Refusal"
        99="Blank";

Value fw        1= "Yes, for employees"
                2= "Yes, for employees and their families"
```

```
                    3= "No"
                    96="Legitimate skip"
          97="Don't know"
          98="Refusal"
          99="Blank";

Value fx1= "Not at all challenging"
                    2= "Slightly challenging"
       3= "Somewhat challenging"
                    4= "Challenging"
                    5= "Extremely challenging"
                    96="Legitimate skip"
          97="Don't know"
          98="Refusal"
          99="Blank";

Value fy          1= "Selected"
                    2= "Not selected"
                    5= "Logically Not Selected (HP1=No)"
                    96="Legitimate skip"
          97="Don't know"
          98="Refusal"
          99="Blank";

Value fz96="Legitimate skip"
          97="Don't know"
          98="Refusal"
          99="Blank";

Value fzb         996="Legitimate skip"
          997="Don't know"
          998="Refusal"
          999="Blank\Invalid";
run;

format    cdc_region cr.
        region rg.
        industry ind.
        size sz.
        oc1 oca. oc3 ocb.
        HI1 fa.
        HI2 fb.
        HI3-HI4 HRA1 CP1-CP5 HP1 HP4 HP6 HP7A HP7B HP7C HP7D HP7D1-HP7D3 HP7E HP7E1 HP7F
HP7F1-HP7F11
        HPR1_1 HPR1_2A HPR1_2B HPR1_2C HPR1_2D HPR1_2E HPR1_2F HPR1_2G HPR1_2H HPR1_2I
HPR1_2J HPR1_2K
        HPR2_1 HPR2_2A HPR2_2B HPR2_2C HPR2_2D HPR2_3A HPR2_3B HPR2_3C HPR2_3D HPR2_4A
HPR2_4B HPR2_4C HPR2_4D
```

HPR2_4E HPR3_1 HPR3_2 HPR4_1 HPR4_2A HPR4_2B HPR4_2C HPR4_2D HPR4_2E HPR4_2F HPR4_2G HPR4_2G1 HPR4_2G2 HPR4_2G3 HPR4_2G4 HPR4_2G5 HPR5_1 HPR6_1 HPR7_1 HPR8_1 HPR9_1

HS11-HS19

DM2B DM3 KP3_1 KP3_2 KP3_3 KP5A KP5C KP5E KP5F KP5G KP5H KP5J WL2 WL5 WL6 WL7 WL8 WL9 WL11 WL12 WL14 WL15

OSH1-OSH6 OSH7_1 OSH7_2 OSH7_3 HPR5_2A_S HPR5_2B_S HPR5_2C_S HPR5_2D_S HPR5_2E_S HPR9_2A_S HPR9_2B_S HPR9_2C_S HPR9_2D_S HPR9_2E_S HPR9_2F_S

HPR8_2A_S HPR8_2B_S HPR8_2C_S HPR8_2D_S HPR8_2E_S HPR7_2A_S HPR7_2C_S HPR7_2D_S HPR7_2E_S HPR7_2F_S

HPR6_2A_S HPR6_2B_S HPR6_2C_S HPR6_2D_S HPR6_2E_S OSH7_1_S OSH7_2_S OSH7_4_S OSH7_6_S OSH8_S

KP1A_S KP1B_S KP1C_S KP1D_S KP1E_S KP1F_S KP1G_S fc.

HS11A HS11_2 HS12_2 HS13_2 HS14_2 HS15_2 HS16_2 HS17_2 HS18_2 HS19_2 fcb.

WL1_1_S WL1_2_S WL1_3_S WL1_4_S WL1_5_S fcc.

OSH81A_S OSH81B_S OSH81C_S fcd.

HRA1A fd.

HRA1B fe.

HRA1E ff.

HP2 fg.

HP3 fh.

HP5 fi.

HP5A fj.

HP7G fk.

HP8 fl.

HPR1_1A HPR2_1A HPR3_1A HPR4_1A HPR5_1A HPR6_1A HPR7_1A HPR8_1A HPR9_1A fm.

HPR1_1B HPR2_1B HPR3_1B HPR4_1B HPR5_1B HPR6_1B HPR7_1B HPR8_1B HPR9_1B fn.

HPR1_1C HPR2_1C HPR3_1C HPR4_1C HPR6_1C HPR7_1C HPR8_1C HPR9_1C HS2B fo.

HPR4_2G6 fp.

HPR4_2G7 fq.

HS2A fr.

HS3 fs.

DM2A ft.

KP2 fu.

KP4 fv.

WL1 fw.

B1_1 B1_2 B1_3 B1_4 B1_5 B1_6 B1_7 B1_8 B1_9 B1_10 B1_11 B1_12 fx.

DM11M1 DM11M2 DM11M3 DM12M1 DM12M2 DM12M3 DM13M1 DM13M2 DM13M3 DM14M1 DM14M2 DM14M3

DM15M1 DM15M2 DM15M3 DM16M1 DM16M2 DM16M3 DM17M1 DM17M2 DM17M3 DM18M1 DM18M2 DM18M3

DM19M1 DM19M2 DM19M3 DM20M1 DM20M2 DM20M3 WL3M1 WL3M2 WL3M3 WL3M4 WL3M5

E1M1 E1M2 E1M3 E1M4 E1M5 E1M6 E1M7 E1M8 E1M9 fy.

E2 fz.
WD1_1 WD1_1 WD2-WD7 fzb.
;