



UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

BERKELEY · DAVIS · IRVINE · LOS ANGELES · RIVERSIDE · SAN DIEGO · SAN FRANCISCO · SANTA BARBARA · SANTA CRUZ

The California Work and Health Survey
Institute for Health Policy Studies
Box 0920
San Francisco, CA 94143-0920

(415) 476-1001
email: trupin@itsa.ucsf.edu

A User's Guide to the 1998 California Work and Health Survey

Overview

The California Work and Health Survey (CWHS) is a telephone-based, longitudinal survey of California adults, designed by faculty and staff of the Work and Health Program at the University of California, San Francisco, with input from researchers and practitioners in the fields of health and economics. It is conducted by the Field Institute and funded by the California Wellness Foundation. The survey includes extensive coverage of employment status, recent job loss, working conditions and environment, and of physical and mental health status. At least three annual interviews are planned. This dataset includes respondents from the initial interview. The survey was administered between June 10 and July 21, 1998 to 1,771 adults (age 18 or older) living in California. Interviews were conducted in both English and Spanish.

Public Use Agreement

The California Work and Health Survey data are available for use by the general public, with the understanding that they will be used solely for research purposes. Users of the CWHS must agree to make no attempt to identify individual respondents on the datafile. Additionally, we ask that you provide us with a copy of any publication or report that you produce using the CWHS. Send such publications to us at:

The California Work and Health Survey
University of California, San Francisco
Box 0920
San Francisco, CA 94143-0920

Acknowledgments

The principal and co-principal investigators for the CWHS are Edward Yelin, Ph.D. and Laura Trupin, M.P.H. of the Institute for Health Policy Studies and Department of Medicine at the University of California, San Francisco. Irene Yen, Ph.D., of the School of Public Health at the University of California, Berkeley, assisted in the design of the survey. The project is funded by the Work and Health Initiative of The California Wellness Foundation. Ruth Brousseau, Ph.D. and Lucia Corral, J.D., are the Senior Program Officer and Program Officer, respectively, for the Work and Health Initiative at the Foundation. The survey was conducted by The Field Institute, an independent public policy research organization in San Francisco, under the direction of Mark DiCamillo. Dean Sonneborn prepared this public release file; Jim Calvert assisted with the preparation of the codebook.

The following individuals provided additional expertise in the design of the survey: Steve Levy of the Center for the Continuing Study of the California Economy; David Dooley of the School of Social Ecology at the University of California, Irvine; E. Richard Brown and Thomas Rice of the Center for Health Policy Research at UCLA, Paul Ong, Chair of Public Policy and Social Research at UCLA; David Hayes-Bautista, Director of the Center for the Study of Latino Health at UCLA; Nancy Adler, Director of the Health Psychology Program at the University of

California, San Francisco; Amy Dean, President of South Bay AFL-CIO; James Head, President of the National Economic Development and Law Center in Oakland; Chris Benner of Working Partnerships in San Jose; Dan Friedland, George Kaplan, and Rick Price of the Michigan Prevention Research Center at the University of Michigan, Ann Arbor, and Harriet Presser of the University of Maryland.

Sample

Of the 1,771 respondents, 1,500 were obtained through random-digit dialing. Up to six attempts were made to contact an adult member of the household. If more than one adult was available, the interviewer requested the youngest adult male; if there were no men at home, the oldest female was asked to participate. This strategy is one that has been developed by the major public opinion research firms, in order to balance participation in phone surveys.

In order to increase the accuracy of survey estimates among selected groups of interest, an additional 271 interviews were conducted among African Americans, Asian/Pacific Islanders, and persons with disabilities. The oversampling procedures differed for each of the three populations. To oversample African Americans, phone numbers were drawn at random from a list of phone exchanges from areas known to have higher proportions of African Americans than the state average, based on Census block information. For Asian/Pacific Islanders, phone numbers were drawn at random from a list of Asian surnames and phone numbers compiled by an independent survey sampling firm. A screening question determined if the respondent was a member of the particular group of interest. Persons with disabilities were identified through a screening question that asked if the respondent had any limitations in activities as a result of long-term physical or mental impairments (see item 104 in the questionnaire).

Overall sample results are subject to a sampling error of +/- 2.4 percentage points. Users are cautioned against attempting to make estimates for small subsets of the data.

Weights

In order to have the sample of respondents align as closely as possible to the known characteristics of the California population (in terms of household size, age, gender, race/ethnicity, region) and to account for the oversampling of the groups described above, the Field Institute calculated weights for each respondent. Two types of weights have been included in the file: FNLWGT98—a proportional weight; using this weight does not inflate the sample size—and EXPWGT98—using this variable inflates the sample to the entire California adult population.

The proportional weight should be used for any statistical calculations. The sum of the weighted frequencies will be equal to the sample size. Use the expansion weight to make prevalence estimates for the California adult population.

For example, to estimate the number of California workers with more than 1 job, use EXPWGT98. However, to compare the likelihood of multiple jobholding for men and women using a chi-square test, use FNLWGT98, so as not to inflate the sample size for the statistical test. Sample SAS code for these two examples is included in the appendix.

We strongly recommend using the weights. If you choose not to include the weight variables in your analysis, you should select only those respondents obtained through random digit-dialing. Note, however, that this group cannot be considered a true random sample of California adults, because the individual respondent in a given household was not selected at random. The variable SAMPLE98 indicates whether the household was selected through random digit-dialing or from one of the oversampling processes.

Data

Accompanying this document is a questionnaire that has been annotated with the variable names corresponding to each question. The variables in the dataset are in the same order as the questions¹.

The dataset can be provided in one of two ways: as a SAS file (version 6.11, for Windows), or as a text file. In the SAS dataset, the question numbers are identified in the variable labels.

The text file contains one row per respondent, with a space between each variable. Depending on the individual responses, the record length varies from a minimum of 1050 to a maximum of 1280 characters. A second text file lists the variable names in order. Sample SAS input statements are included in the appendix.

Masking for confidentiality

To ensure that no individual can be identified from these data, we do not include geographic information below the county level. Further, we have combined 23 rural counties (defined here as having fewer than 100,000 residents) into a single code -- "36".

Notes on variables

Most variable names in this 1998 dataset have "98" in the last 2 positions (eg. WORK98). The only exceptions to this are demographic variables that will not change over time, the respondent's ID, and the code for the survey year. In subsequent years the variables names will be changed to 99, 00, 01, etc, to prevent users from confusing multiple years of data.

Whenever a question allows for multiple responses, a separate variable was created for each possible response (see, for example, q. 7). Occasionally, a potential response category was not selected by any respondents (e.g. q.25 "vacation"). The SAS label indicates this with NR. We have retained the variable for consistency with future years of data.

In the case of respondent's occupation, the categories were developed based on responses to two open-ended questions about work tasks (q. 41 and q. 42).

All variables referring to time of day have been converted to a 24 hour clock. We have similarly recoded responses from hours and minutes to minutes, and from months and years to years, in order to have all information from a single question contained in a single variable.

The social network questions (q. 112-126) ask about numbers of friends or relatives. Any responses of 97 or greater are coded as 97.

Not all variables are listed on the questionnaire. These include: the weight and sampling variables described above, two longitudinal variables that will be useful when multiple years of data are available, a handful of calculated variables for the convenience of the user, and labor force status variables based on U.S. Bureau of Labor Statistics definitions. All of these additional variables are defined below:

¹ Due to the process of revising the survey skip patterns, the questions are occasionally out of numeric order.

Longitudinal variables

1. ID = Respondent identifier; will not change from year to year.
2. YEAR = survey year (1998 for all respondents).
3. INTRVW98 = 1. Identifies interview number for each respondent. For the first year, everyone has a 1 in this field. In 1999, new respondents will have a 1 and continuing respondents will have a 2 in this field.

Weight and sampling variables

4. FNLWGT98 = proportional weight
5. EXPWGT98 = expansion weight
6. PREWGT98 = an interim variable (should not be used)
7. SAMPLE98: 1=Random sample, 2=African American sample, 3=Asian/Pacific Islander sample, 4=Persons with disabilities sample

Calculated variables

8. EMPLOY98 = combination of q. 1 and q. 2 -- 1=working, 0=not working
Respondents with a 1 in this variable are asked the set of questions about current job(s).
Note that this is **not** a Current Population labor force status variable (see below).
9. TOTADU98 = Total adults in household
10. TOTCHI98 = Total children in household
11. DEPRES98 = The Short Geriatric Depression Scale (S-GDS) score ^{1,2}. This is a count of all the depressive symptoms identified in questions 127a - 127o. The score is calculated by converting the responses so that 1 signifies a depressive symptom for all 15 items. A cut-point of 7 (out of 15 responses) has been suggested as an indicator of a high degree of depressive symptomology³.

Note that all respondents who answered *any* of the items in question 127 have a GDS score. Users may wish to omit respondents who skipped items, or to make the cut-point for depression be based on the total number of items answered.

References

1. Sheikh JI & Yesavage JA. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clin Gerontol* 5:165-173.
2. Rule BG, Harvey HZ & Dobbs AR. (1989). Reliability of the Geriatric Depression Scale for younger adults. *Clin Gerontol* 9:37-43.
3. Cwikel J & Ritchie K. (1989). Screening for depression among the elderly in Israel: An assessment of the short Geriatric Depression Scale (S-GDS). *Isr J Med Sci* 25:131-137.

12. BMI98 = Body mass index (kg/m²)
13. RACE = a variable identifying 1 racial category for each person.
1 = White
2 = African American
3 = Asian/Pacific Islander

4 = Other

We followed these decision rules to assign individuals with multiple responses to a single category:

Non-“other” (i.e. White, African American, or Asian) takes precedence over “other”

Non-white (i.e. African American or Asian) takes precedence over White

There were no respondents who identified themselves as both African American and Asian.

Labor force variables

Several variables were created to help users identify labor force status in keeping with the Bureau of Labor Statistics (BLS), which has a strict definition of unemployment. The CWHS employment status questions were patterned after the Current Population Survey (CPS), and the LFCPS98 variable was coded following the CPS guidelines. To be unemployed, a person must be on temporary lay-off and awaiting a callback, actively looking for work and available for work, or waiting to start a new job in the next 30 days. All other persons without jobs are considered to be out of the labor force. The SAS code used to identify the BLS labor force status is included in the appendix to this codebook.

1. LFCPS98 = Labor force status, modeled after the BLS' Current Population Survey

1 = Working

2 = With a job, but not at work

3 = Unemployed

4 = Out of the labor force

2. LFPRT98 = labor force participation rate:

1 = In labor force (LFCPS98 = 1,2,3), 0 = out of labor force (LFCPS98 = 4)

3. EMPRT98 = employment/population ratio:

1 = Has job (LFCPS98 = 1,2), 0 = No job (LFCPS98 = 3,4)

4. UNEMPR98 = unemployment rate:

1 = Unemployed (LFCPS98 = 3), 0 = Has job (LFCPS98 = 1,2)

Missing values

There are two types of missing values on the file. If a question was irrelevant to the respondent and appropriately skipped, the variable has a value of “.” in the SAS file, and -99 in the text file. Responses of don't know, refused, or not answered are coded with “.R” in the SAS file and -98 in the text file. Users of the text file are advised to convert the missing values before running numerical procedures.

Feedback

This is the first release of data from the first year of the California Work and Health Survey; we view it as a work-in-progress. We welcome your feedback about the survey topics, the specific questions, and the documentation. If you think that you have identified a coding error, please contact us promptly. Thank you.

Appendix -- Sample SAS code

Working with weights

- a. Estimating prevalence of male and female workers with more than 1 job:

```
Proc freq; tables ONEJOB98*GENDER; weight EXPWGT98; run;
```

- b. Testing gender difference in proportion of population with more than 1 job:

```
Proc freq; tables ONEJOB98*GENDER /chisq; weight FNLWGT98; run;
```

(Note that the percentages in both examples will be identical).

Labor force variables (using Bureau of Labor Statistics definitions)

```
LIBNAME SSD 'C:\1WORK\CAWORK';
Data one; set ssd.whealth;
LFCPS98 = 4; ** Out of LF, unless specified below;
** Working;
If WORK98 = 1 or ANYWRK98 = 1 then LFCPS98 = 1;
If ANYWRK98 = .R and WORK98 = .R then LFCPS98 = .R;

***Have job, but not at work (ie. illness, vacation, bad weather, labor
dispute or other reason -- but clearly have job, based on other answers);
If 1 le REASO298 le 4 then LFCPS98 = 2;
If REASO298 = 8 and (WAGES98 ne .R or MOHRS298 ne .R) then LFCPS98 = 2;
If ABSENT98 = 1 and REASO298 = .R then LFCPS98 = .R;
If ABSENT98 = 2 and LOOKING = .R then LFCPS98 = .R;

***Unemployed or Out of LF;
If REASO298 in (5,7) then LFCPS98 = 3; *temp layoff or waiting to start new
job;
If LKING98 = 1 and TAKJOB98 = 1 then LFCPS98 = 3; *looking for work & avail;
If REASO298 = 6 and TAKJOB98 = 1 then LFCPS98 = 3; *on indef. layoff & avail;
If LKNO98 = 1 then LFCPS98 = 4; *not actively looking;

array done done198--done1298;
do over done;
if done=. then done=0;
end;
DONESUM= SUM(DONE198--DONE1298);
If DONE298 = 1 and DONE298=DONESUM then LFCPS98 = 4; *checked want ads;
If DONE1198 = 1 and DONE1198=DONESUM then LFCPS98 = 4; *not actively looking;

*CPS rates;
*LFPR;
If 1 le LFCPS98 le 3 then LFPRT98 = 1;
Else if LFCPS98 eq 4 then LFPRT98 = 0;
*Emp/Pop;
If 1 le LFCPS98 le 2 then EMPRT98 = 1;
Else if 3 le LFCPS98 le 4 then EMPRT98 = 0;
*Unemp;
If LFCPS98 eq 3 then UNEMPR98 = 1;
Else if 1 le LFCPS98 le 2 then UNEMPR98 = 0;
run;
```

Appendix -- Sample SAS code, continued

Sample SAS input statements, for use with ASCII file:

Note: The ASCII file *does not* contain the raw interview responses. Rather, the text file was created from the SAS file, and includes all calculated variables described in this Users' Guide. These statements are provided for users who will be using SAS under a platform other than Windows.

```
data one;
  infile 'FILENAME' missover lrecl=1280 ;

input ID YEAR WORK98 ANYWRK98 EMPLOY98 LFCPS98 LFPRT98 EMPRT98 UNEMPR98
ONEJOB98 MANYJ98 HOURS98 MOHRS198 WRKL198 WRKL298 WRKL398 WRKL498 WRKL598
WRKL698 WRKL798 WRKL898 WRKL1098 WRKL1198 WRKL1298 WRKL1398 WRKL1498 WRKL1598
ABSENT98 REASO98 WAGES98 MOHRS298 LKING98 LOOKNO98 LAGEN98 LKEMP98 LKAD98
LKFRND98 LKOTHE98 DONE198 DONE298 DONE398 DONE498 DONE598 DONE698 DONE798
DONE898 DONE998 DONE1198 DONE1298 LOSTJO98 LAIDOF98 WKSLOK98 FULTI198
TAKJOB98 WHYNOT98 LSTWRK98 LSTWK298 LEWK198 LEWK298 LEWK398 LEWK498 LEWK598
LEWK698 LEWK798 LEWK898 LEWK998 LEWK1198 LEWK1298 LEWK1398 LEWK1498 LEWK1598
LEWK1698 LEWK1798 FULPRT98 NWRK198 NWRK298 NWRK398 NWRK498 NWRK598 NWRK698
NWRK798 NWRK898 NWRK998 NWRK1198 NWRK1298 NWRK1398 NWRK1498 NWRK1598 NWRK1798
NWRK1898 NWRK1998 NWRK2098 LOOK298 WEKWRK98 HRSWK98 SELFEM98 INDUST98
WORKED98 PROMO98 CHANGE98 OCC198 PRESOC98 WORKER98 STA2498 END2498 FLXHRS98
NEW98 FREEDOM98 OWN98 TIME98 FAST98 SUPER98 SUPNUM98 CRIME98 NOISE98 TRASH98
LIGHT98 ACCESS98 SHOPS98 WALK98 STAIR98 SIT98 STOOP98 LIFT98 CARRY98 HAND98
BEND98 TOOLS98 PERJOB98 PROJEC98 TEMPAY98 UNIMEM98 UNICOV98 HLTINS98 INSTYP98
PREMI98 HMO98 GRPDOC98 EXTRBE98 PAYPAR98 TMPHLT98 ANYPLN98 LOSEJO98 LEAVJB98
LASTYR98 HOWLG298 TYPEMP98 OCC298 PAYEND98 JBTRN198 CLASS98 JBTRN298 EFFECT98
HEALTH98 SMOKE98 SMAGE98 SMDAY98 CIGAV198 SMPAST98 CIGAV298 SMYRS98 ALCOHO98
WEKMTH98 ALCWEK98 ALCMTH98 ALCDNK98 ALCFIV98 EXERCI98 EXTYPT98 EXMTH98
EXTIME98 HBP98 HEART98 DIABET98 CANCER98 ASTHMA98 MIGRAI98 LUNG98 ULCER98
KIDNEY98 BACK98 CARPAL98 ARJNT98 ARSYM98 ARLIM98 ARTHRI98 AROSTE98 ARRHEU98
ARRHSM98 ARLYME98 AROTH98 ARDK98 ARTX98 HOSP98 MEDVIS98 LASMED98 MEDREG98
MEDOFC98 BEDDAY98 LIMACT98 LIMWRK98 LIMWR298 LIMHOU98 LIMHO298 KEPWRK98
HELP98 SHORT98 CHGTIM98 BREAK98 TRANS98 CHGJOB98 NEWSKI98 SPECEQ98 SLEEP98
SLEMED98 SLEHRS98 SLEOK98 FRIEND98 FRIEJB98 RELCLS98 RELVIS98 RELPHO98
RELHEL98 RELSAT98 SATREL98 DEPRES98 LIFSAT98 DROPAC98 LIFEMP98 BORED98
SPIRIT98 BAD98 HAPPY98 HLPLSS98 ATHOME98 PROBLE98 WONDER98 WORTH98 ENERGY98
HOPLES98 BETTER98 CONTRL98 PERSON98 YOURWA98 OVERCO98 NEGTR98 NEGAGE98
NEGSEX98 NEGRAC98 NEGETH98 NEGHAN98 NEGOTH98 NEGORE98 GENDER AGE AGECAT98
EDUC98 COMPUT98 COMHOM98 COMWRK98 COMSCH98 COMPNO98 MARIED98 PARTED98
HISPANIC RACE ORIGIN WHITE BLACK ASIAN RACE_OTH RACE_DK TALL98 WEIGHT98 BMI98
BORN USAGE98 LANGHO98 LANGWK98 HOUSEH98 AGE1898 TOTADU98 AG131798 AG51298
AGLES598 TOTCHI98 CARHRS98 PARCAR98 RELCAR98 DAYCAR98 FAMDIN98 HOSINC98
HINC298 HINC398 WELFAR98 AFDC98 FODSTP98 ADC98 TANF98 WIC98 INDINC98 SUFINC98
HRDSHP98 REDUC98 WORK1898 WORKTI98 OSTA2498 OEND2498 WORKRE98 NAME198 NAME298
NAME398 NAME498 NAME598 NAME698 REL198 REL298 REL398 REL498 REL598 REL698
WORKD198 WORKD298 WORKD398 WORKD498 WORKD598 WORKD698 RELST198 RELST298
RELST398 RELST498 RELST598 RELST698 RELEN198 RELEN298 RELEN398 RELEN498
RELEN598 RELEN698 COUNTY98 CALBAK98 SAMPLE98 PREWGT98 EXPWGT98 FNLWGT98;

run;
```