



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

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

1999 Cross-sectional File
1998-1999 Longitudinal File
(issued October 18, 1999)

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. Interviews are conducted in English and Spanish. At least three annual interviews are planned; the first two years have now been completed. Preliminary findings from 1998 and 1999 are available at <http://medicine.ucsf.edu/programs/cwhs/>.

This codebook provides general guidelines for use of both the 1999 cross-sectional and the 1998-1999 longitudinal file. The cross-sectional dataset includes the responses from the 2,040 participants interviewed between April 30 and July 8, 1999. The longitudinal dataset contains interviews with the 909 participants who responded to both years' interviews.

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 Field Research

Corporation, 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, Director of the UCLA Lewis Center for Regional Policy Studies; 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 the South Bay AFL-CIO Labor Council; James Head, President of the National Economic Development and Law Center in Oakland; Chris Benner of the Department of City and Regional Planning, University of California, Berkeley; 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 2,040 respondents, 909 were re-interviews with 1998 participants, representing 57% of the 1,597 1998 respondents who had agreed to be re-contacted. Of the remaining 1,131 1999 respondents, 700 were obtained through random-digit dialing and the rest through oversampling procedures described below. For new respondents, 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. The overall cooperation rate¹ for the new baseline respondents is 56.7%.

In order to increase the accuracy of survey estimates among selected groups of interest, an additional 431 interviews were conducted among African Americans, Asian/Pacific Islanders, persons with disabilities, and persons aged 45- 70. The oversampling procedures differed for each of the four populations. To oversample African Americans, phone numbers were drawn at random from a list of phone exchanges covering 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). Persons aged 45-70 were screened into the survey through an age question at the beginning of the interviews.

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,

¹ Cooperation rate = # of completed interviews / (# completed+ # refused+# ineligible)

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: WGTTOT99—a proportional weight; using this weight does not inflate the sample size—and WGTEXT99—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 WGTEXT99. However, to compare the likelihood of multiple jobholding for men and women using a chi-square test, use WGTTOT99, so as not to inflate the sample size for the statistical test. Sample SAS code for these two examples is included in Appendix I.

For the longitudinal file, there are also two weight variables, WGT9899L and WGT9899X, representing the proportional and expansion weight for the longitudinal dataset, respectively. These weights are based on the 1998 sampling weights, with an adjustment to account for attrition. Appendix III describes in detail how we developed that adjustment.

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. The variable SAMPLE99 indicates whether the household was selected through random digit-dialing or from one of the oversampling processes. 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. Also, it is not possible to determine from the SAMPLE99 variable how the 1998 respondents were originally obtained, as all of them have been given the same code for that variable.

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 questions². In addition, there is an alphabetical listing of the variables in both datasets included in Appendix II.

The dataset can be provided in several ways: as a SAS file (in version 6.12 for Windows or as a transport dataset), as an SPSS "portable" file or as a text file. In the SAS and SPSS files, the question numbers are identified in the variable labels. Appendix I gives sample SAS code for converting the transport dataset to a working file. 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 892 to a maximum of 1,158 characters. A second text file lists the variable names in order. We may be able to create datasets in other formats as well; contact that CWHHS office with your specific request.

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 small counties (defined here as having fewer than 100,000 residents) into a single code -- "36".

² Due to the process of revising the survey skip patterns, the questions are frequently out of numeric order.

Notes on variables

Most variable names in this 1999 dataset have "99" in the last 2 positions (e.g. WORK99). The only exceptions to this are demographic variables that will not change over time, the respondent's ID, and the code for the first year in the survey. In subsequent years the variables names will be changed to 00, 01, etc, to prevent users from confusing multiple years of data.

We have 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. 112-116) ask about numbers of friends or relatives. Responses to these questions range up to 98.

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 pooling multiple years of data, 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:

Longitudinal and sampling variables

1. ID = Respondent identifier; will not change from year to year.
2. YEAR99 = survey year (1999 for all respondents).
3. YENTER = the year the respondent entered the survey (1998 for panel respondents, 1999 for new respondents).
4. INTRVW99: Identifies interview number for each respondent. New respondents have a 1 and continuing respondents have a 2 in this field.
5. PANEL99: 1=Continuing respondents, 0=new respondent entering in 1999
6. SAMPLE99: 1=Cross section (1999 Baseline), 2=Asian augment (excluding panel), 3=Black augment (excluding panel), 4=Disabled augment (excluding panel), 5=Aging supplement (excluding panel), 6 Panel re-interviews
7. AGESUP99= flags respondents included in the 1999 aging supplement (aged 45-70).
8. LANG99 = Language of interview: 0=English 1=Spanish

Weight variables in the cross-sectional file

1. WGTTOT99 = proportional weight for all 1999 respondents
2. WGTEXT99 = expansion weight for all 1999 respondents

Weight variables in the longitudinal file (see Appendix III for details)

1. WGT9899L= proportional weight for 1998-1999 respondents
2. WGT9899X = expansion weight for 1998-1999 respondents
3. FNLWGT98 = original 1998 sampling weight
4. ATTRADJ = adjustment to FNLWGT98 to account for loss to follow-up.

Calculated variables

1. HASJOB99 = combination of q. 1, q.2, and q. 9 -- 1=has job, 0=does not have job
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 Survey labor force status variable, although it is quite similar to the employment/population ratio (see below).
2. TOTADU99 = Total adults in household

3. TOTCHI99 = Total children in household
4. DEPRES99 = 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.

5. BMI99 = Body mass index (kg/m²)
6. RACE = a variable identifying 1 racial category for each person.
 - 1 = White
 - 2 = African American
 - 3 = Asian/Pacific Islander
 - 4 = Other

Note: In keeping with current US Census guidelines, the CWHS questionnaire allows for persons to identify themselves as a member of multiple racial groups. In cases of multiple responses, non-"other" (i.e. white, African American, or Asian/Pacific Islander) takes precedence over "other," and "non-white" (i.e. African American or Asian/Pacific Islander) takes precedence over white. There were no respondents who identified themselves as both African American and Asian/Pacific Islander. The original responses have been retained in the file, as well.

7. RURAL99 = whether the person resides in a rural county (1) or not (0); taken from the California Rural Health Policy Council definition which is based on population density (<http://www.ruralhealth.cahwnet.gov/>). Rural counties include all those coded "36" in the COUNTY99 variable, as well as a number of large, but sparsely-populated counties.
8. SOCA99 = resides in Southern (1) or Northern California (0). Southern California includes the 9 most southerly counties, and Northern California includes all other counties.

Labor force variables

Several variables 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 LFCPS99 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 Appendix I.

1. LFCPS99 = 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. LFPRT99 = labor force participation rate:
 - 1 = In labor force (LFCPS99= 1,2,3), 0 = out of labor force (LFCPS99 = 4)
3. EMPRT99 = employment/population ratio:
 - 1 = Has job (LFCPS99 = 1,2), 0 = No job (LFCPS99 = 3,4)
4. UNEMPR99 = unemployment rate:
 - 1 = Unemployed (LFCPS99= 3), 0 = Has job (LFCPS99 = 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

We welcome your comments 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 I -- Sample SAS code

Working with weights

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

```
Proc freq; tables ONEJOB99*GENDER; weight WGTEXT99; run;
```

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

```
Proc freq; tables ONEJOB99*GENDER /chisq; weight WGTTOT99; run;
```

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

Sample SAS code to copy the transport file to a working dataset:

Note: These statements are provided for users who have requested the transport dataset for using SAS under a platform other than Windows. Refer to the documentation for your SAS platform for the correct way to refer to file and library names, etc.

```
*****Set up incoming and outgoing libraries;
libname frxpt XPORT 'C:\lwork\cwhs99pb.xpt'; *name/location of transport data;
libname tosas 'C:\lwork\cawork\cwhsdata'; *location for destination dataset;

*****Run proc copy, selecting CWHS dataset;
proc copy in=frxpt out=tosas; select cwhs99pb; run;

***Confirm successful transport;
proc contents data=tosas.cwhs99pb; run;
```

Appendix I -- Sample SAS code, continued

Labor force variables (using Bureau of Labor Statistics definitions)

```
LIBNAME SSD 'C:\1WORK\CAWORK';
Data one; set ssd.cwhs99pb;
**1999 Labor force statistics -- per approx. CPS definition;
  *LFCPS99: 1=working, 2=have job, not at work, 3=unemployed, 4=out of LF;
LFCPS99 = 4; ** Out of LF, unless specified below;
** Working;
If WORK99 = 1 or anywrk99 = 1 then LFCPS99 = 1;
If anywrk99 = .R and WORK99 = .R then LFCPS99 = .R;

***Have job, but not at work (i.e., illness, vacation, bad weather, labor
dispute or other reason -- but clearly have job, based on subsequent
answers);
If 1 le REASO99 le 4 then LFCPS99 = 2;
If REASO99 eq 8 then LFCPS99 = 2;
If ABSENT99 eq 1 and REASO99 eq .R then LFCPS99 = .R;
If ABSENT99 eq 2 and LKING99 eq .R then LFCPS99 = .R;

***Unemployed or Out of LF;
If REASO99 in (5,7) then LFCPS99 = 3; *temp layoff or about to start new job;
If LKING99 eq 1 and TAKJOB99 eq 1 then LFCPS99 = 3; *looking for work & avail;
If REASO99 eq 6 and TAKJOB99 eq 1 then LFCPS99 = 3; *on indef. layoff & avail;
If lookno99 eq 1 then LFCPS99 = 4; *not actively looking;

*CPS rates;
  *LFPR;
    If 1 le LFCPS99 le 3 then LFPRT99 = 1; *in labor force;
    Else if LFCPS99 eq 4 then LFPRT99 = 0; * out of labor force;
  *Emp/Pop;
    If 1 le LFCPS99 le 2 then EMPRT99 = 1; *Employed;
    Else if 3 le LFCPS99 le 4 then EMPRT99 = 0; *Not empl.
                                     (includes unemp & out of lf);
  *Unemp;
    If LFCPS99 eq 3 then UNEMPR99 = 1; *Unemployed;
    Else if 1 le LFCPS99 le 2 then UNEMPR99 = 0; *Employed;

*Employ99 (per CWHs);
  If anywrk99 eq 1 or WORK99 eq 1 then EMPLOY99 = 1;
  Else EMPLOY99 = 0;
  If anywrk99 = .R and WORK99 = .R then EMPLOY99 = .R;

*HASJOB99 (in 1999, treated as "employed" for purposes of employment
questions);
  If EMPLOY99 = 1 or REASO99 in (1,2,3,4) then HASJOB99 = 1;
  Else HASJOB99 = 0;
```


Appendix II -- Alphabetical variable lists

1999 cross-sectional file

ABSENT99	DIABET99	HRSWK99	MBAR299	PEN299	SSI99
ACCESS299	DISAGE99	ICDEDC99	MDRINK99	PENADJ99	SSI299
ACCESS99	DISETH99	ICW299	MEDOFC99	PENPLN99	SSOC99
ADC99	DISHAN99	ID	MEDREG99	PENTY299	SSOC299
AFDC99	DISRAC99	INDCN299	MEDVIS99	PENTYP99	STAIR99
AG51799	DISSEX99	INDINC99	MGAP99	PERJOB99	STOOP99
AGE99	DROPAC99	INDUL199	MIGRAI99	PERSON99	SUFINC99
AGE1899	EDUC99	INDUL299	MOHRS199	PREFER99	SUPER99
AGECAT99	ELICON99	INDUST99	MOHRS299	PREMI99	TAKJOB99
AGERET99	ELIPLN99	INS12M99	MOST99	PRESOC99	TALL
AGESOC99	EMPLAN99	INSCHG99	MOST299	PROBLE99	TANF99
AGESUP99	EMPLOY99	INSTYC99	MREST99	PROJEC99	TEMPAY99
AGLES599	EMPRT99	JOB99	MREST299	PROMO99	TIME99
ALCDNK99	ENERGY99	JOBINC99	MWORK99	RACE	TMPHLT99
ALCFIV99	EXERCI99	KIDNEY99	MWORK299	RACE_DK	TOOLS99
ALCMTH99	EXMTH99	LAGEN99	NCHILD99	RACE_OTH	TOTADU99
ALCOHO99	EXTIME99	LAI DOF99	NEW99	RAGE99	TOTCHI99
ALCWEK99	EXTRBW99	LANGHOME	NOINS99	REASO99	TRAS299
ANYPLN99	EXTYPT99	LANGWK99	NOIS299	REASO199	TRASH99
ANYWRK99	FAM99	LASMED99	NOISE99	REDUC99	TYPEMP99
ARTHR I99	FAM299	LASTYR99	NONEG99	RELCLS99	ULCER99
ASIAN	FAMHLY99	LEAVJB99	NOSKIL99	RELPHC99	UNEMPR99
ASSET99	FAST99	LFCPS99	NOTHIR99	RELVIS99	UNICOV99
ASSET299	FINBET99	LFPRT99	NOTPRO99	RENT99	UNIMEM99
ASSIST99	FIRE99	LIFEMP99	OCC199	RESTDK99	USAGE
ASTHMA99	FLXHRS99	LIFSAT99	OCC299	RETAGE99	WALK99
ATHOME99	FODSTP99	LIFT99	OCCL199	RETIRE99	WEIGHT99
BACK99	FREDOM99	LIGHT299	OCCL299	RURAL99	WEKWRK99
BAD99	FRIEJW99	LIGHT99	ONE99	SAMPLE99	WELFAR99
BARDNK99	FRIENW99	LIMACT99	ONEJOB99	SAV99	WGTEXN99
BEDDAY99	FULPRT99	LIMHO299	ORIFIR99	SAV299	WGTEXT99
BEND99	GAPPAY99	LIMHOU99	ORIG199	SELFEM99	WGTNEW99
BETTER99	GENDER	LIMWR299	ORIG299	SHIFT99	WGTTOT99
BLACK	GRPDOW99	LIMWRK99	ORIG399	SHOP299	WHITE
BMI99	HAND99	LKAD99	ORIG499	SHOPS99	WHOME99
BORED99	HAPPY99	LKEMP99	ORIG599	SIT99	WIC99
BORN	HBP99	LKFRND99	ORIG699	SLEEP99	WKDNK99
CALBAK99	HEALTH99	LKING99	ORIG799	SMAGE99	WKSLOK99
CANCER99	HEART99	LKOTHE99	ORIG899	SMBAR99	WONDER99
CARE99	HINC399	LONGEM99	ORIGIN	SMDAY99	WORK99
CARHRS99	HISPANIC	LONGER99	ORIHIR99	SMMORE99	WORK1899
CARPAL99	HLPLSS99	LOOK299	ORINO99	SMOKE99	WORKD99
CARRY99	HLTH99	LOOKNO99	ORIPRO99	SMPAST99	WORKE99
CHANGE99	HLTINS99	LOSEJO99	OTHACT99	SMREST99	WORKED99
CHEKUP99	HMO99	LOSJOB99	OVERCO99	SMWORK99	WORKER99
CIGAV199	HOPLES99	LOSTJC99	OWN99	SMYRS99	WORKN99
CIGAV299	HOSINC99	LSTWK299	PANEL99	SOCA99	WORKS99
CONSLT99	HOSP99	LSTYR99	PARENT99	SOCSE299	WORKTI99
CONTRC99	HOURSC99	LTERM99	PART99	SOCSEC99	WORTH99
CONTRL99	HOUSEH99	LUNG99	PARTED99	SPIRIT99	YEAR99
COUNTY99	HOWLG299	MAIRET99	PAYEND99	SPORIT99	YENTER
CRIM299	HOWLN199	MANYJ99	PAYPAR99	SPOSOC99	YOURWA99
CRIME99	HOWLN299	MARIED99	PDAYCA99	SPOUS299	
DEPRES99	HRDHP99	MBAR99	PEN99	SPOUSE99	

Appendix II -- Alphabetical variable lists, continued

1998-1999 longitudinal file

ABSENT98	ATTRADJ	DIABET98	FREDOM98	INDCN299	LIFSAT98
ABSENT99	BACK98	DIABET99	FREDOM99	INDINC98	LIFSAT99
ACCES299	BACK99	DISAGE99	FRIEJB98	INDINC99	LIFT98
ACCESS98	BAD98	DISETH99	FRIEJW99	INDUL199	LIFT99
ACCESS99	BAD99	DISHAN99	FRIEND98	INDUL299	LIGH299
ADC98	BARDNK99	DISRAC99	FRIENW99	INDUST98	LIGHT98
ADC99	BEDDAY98	DISSEX99	FULPRT98	INDUST99	LIGHT99
AFDC98	BEDDAY99	DONE198	FULPRT99	INS12M99	LIMACT98
AFDC99	BEND98	DONE298	FULTI198	INSHG99	LIMACT99
AG51298	BEND99	DONE398	GAPPAY99	INSTYC99	LIMHO298
AG51799	BETTER98	DONE498	GENDER	INSTYP98	LIMHO299
AG131798	BETTER99	DONE598	GRPDOG98	INTRVW98	LIMHOU98
AGE98	BLACK	DONE698	GRPDOW99	JBTRN198	LIMHOU99
AGE99	BMI98	DONE798	HAND98	JBTRN298	LIMWR298
AGE1898	BMI99	DONE898	HAND99	JOB99	LIMWR299
AGE1899	BORED98	DONE998	HAPPY98	JOBINC99	LIMWRK98
AGECAT98	BORED99	DONE1198	HAPPY99	KEPWRK98	LIMWRK99
AGECAT99	BORN	DONE1298	HBP98	KIDNEY98	LKAD98
AGERET99	BREAK98	DROPAC98	HBP99	KIDNEY99	LKAD99
AGESOC99	CALBAK98	DROPAC99	HEALTH98	LAGEN98	LKEMP98
AGESUP99	CALBAK99	EDUC98	HEALTH99	LAGEN99	LKEMP99
AGLES598	CANCER98	EDUC99	HEART98	LAIDOF98	LKFRND98
AGLES599	CANCER99	EFFECT98	HEART99	LAIDOF99	LKFRND99
ALCDNK98	CARE99	ELICON99	HELP98	LANG98	LKING98
ALCDNK99	CARHRS98	ELIPLN99	HINC398	LANG99	LKING99
ALCFIV98	CARHRS99	EMPLAN99	HINC399	LANGHOME	LKOTHE98
ALCFIV99	CARPAL98	EMPLOY98	HISPANIC	LANGWK98	LKOTHE99
ALCMTH98	CARPAL99	EMPLOY99	HLPLSS98	LANGWK99	LONGEM99
ALCMTH99	CARRY98	EMPRT98	HLPLSS99	LASMED98	LONGER99
ALCOHO98	CARRY99	EMPRT99	HLTH99	LASMED99	LOOK298
ALCOHO99	CHANGE98	END2498	HLTINS98	LASTYR98	LOOK299
ALCWEK98	CHANGE99	ENERGY98	HLTINS99	LASTYR99	LOOKNO98
ALCWEK99	CHEKUP99	ENERGY99	HMO98	LEAVJB98	LOOKNO99
ANYPLN98	CHGJOB98	EXERCI98	HMO99	LEAVJB99	LOSEJO98
ANYPLN99	CHGTIM98	EXERCI99	HOPLES98	LEWK198	LOSEJO99
ANYWRK98	CIGAV198	EXMTH98	HOPLES99	LEWK298	LOSJOB99
ANYWRK99	CIGAV199	EXMTH99	HOSINC98	LEWK398	LOSTJC99
ARDK98	CIGAV298	EXPWGT98	HOSINC99	LEWK498	LOSTJO98
ARJNT98	CIGAV299	EXTIME98	HOSP98	LEWK598	LSTWK298
ARLIM98	CLASS98	EXTIME99	HOSP99	LEWK698	LSTWK299
ARLYME98	COMHOM98	EXTRBE98	HOURS98	LEWK798	LSTWRK98
AROSTE98	COMPNO98	EXTRBW99	HOURSC99	LEWK898	LSTYR99
AROTH98	COMPUT98	EXTYPT98	HOUSEH98	LEWK998	LTERM99
ARRHEU98	COMSCH98	EXTYPT99	HOUSEH99	LEWK1198	LUNG98
ARRHSM98	COMWRK98	FAM99	HOWLG298	LEWK1298	LUNG99
ARSYM98	CONSLT99	FAM299	HOWLG299	LEWK1398	MAIRET99
ARTHRI98	CONTRC99	FAMDIN98	HOWLN199	LEWK1498	MANYJ98
ARTHRI99	CONTRL98	FAMHLY99	HOWLN299	LEWK1598	MANYJ99
ARTX98	CONTRL99	FAST98	HRDSHP98	LEWK1698	MARIED98
ASIAN	COUNTY98	FAST99	HRDSHP99	LEWK1798	MARIED99
ASSET99	COUNTY99	FINBET99	HRSWK98	LFCPS98	MBAR99
ASSET299	CRIM299	FIRE99	HRSWK99	LFCPS99	MBAR299
ASSIST99	CRIME98	FLXHRS98	ICDEDC99	LFPRT98	MDRINK99
ASTHMA98	CRIME99	FLXHRS99	ICW299	LFPRT99	MEDOF98
ASTHMA99	DAYCAR98	FODSTP98	ID	LIFEMP98	MEDOF99
ATHOME98	DEPRES98	FODSTP99		LIFEMP99	
ATHOME99	DEPRES99				

Appendix II -- Alphabetical variable lists, continued

1998-1999 longitudinal file

MEDREG98	NWRK1298	PENTY299	RETAGE99	STAIR99	WHOME99
MEDREG99	NWRK1398	PENTYP99	RETIRE99	STOOP98	WHYNOT98
MEDVIS98	NWRK1498	PERJOB98	RURAL99	STOOP99	WIC98
MEDVIS99	NWRK1598	PERJOB99	SAMPLE98	SUFINC98	WIC99
MGAP99	NWRK1798	PERSON98	SAMPLE99	SUFINC99	WKDNK99
MIGRAI98	NWRK1898	PERSON99	SATREL98	SUPER98	WKSLOK98
MIGRAI99	NWRK1998	PREFER99	SAV99	SUPER99	WKSLOK99
MOHRS198	NWRK2098	PREMI98	SAV299	SUPNUM98	WONDER98
MOHRS199	OCC198	PREMI99	SELFEM98	TAKJOB98	WONDER99
MOHRS298	OCC199	PRESOC98	SELFEM99	TAKJOB99	WORK98
MOHRS299	OCC298	PRESOC99	SHIFT99	TALL	WORK99
MOST99	OCC299	PROBLE98	SHOP299	TANF98	WORK1898
MOST299	OCCL199	PROBLE99	SHOPS98	TANF99	WORK1899
MREST99	OCCL299	PROJEC98	SHOPS99	TEMPAY98	WORKD99
MREST299	OEND2498	PROJEC99	SHORT98	TEMPAY99	WORKD198
MWORK99	ONE99	PROMO98	SIT98	TIME98	WORKD298
MWORK299	ONEJOB98	PROMO99	SIT99	TIME99	WORKD398
NAME198	ONEJOB99	RACE	SLEEP98	TMPHLT98	WORKD498
NAME298	ORIFIR99	RACE_DK	SLEEP99	TMPHLT99	WORKD598
NAME398	ORIG199	RACE_OTH	SLEHRS98	TOOLS98	WORKE99
NAME498	ORIG299	RAGE99	SLEMED98	TOOLS99	WORKED98
NAME598	ORIG399	REASO98	SLEOK98	TOTADU98	WORKED99
NCHILD99	ORIG499	REASO99	SMAGE98	TOTADU99	WORKER98
NEGAGE98	ORIG599	REASO199	SMAGE99	TOTCHI98	WORKER99
NEGETH98	ORIG699	REDUC98	SMBAR99	TOTCHI99	WORKN99
NEGHAN98	ORIG799	REDUC99	SMDAY98	TRANS98	WORKRE98
NEGORE98	ORIG899	REL198	SMDAY99	TRAS299	WORKS99
NEGOTH98	ORIGIN	REL298	SMMORE99	TRASH98	WORKTI98
NEGRAC98	ORIHIR99	REL398	SMOKE98	TRASH99	WORKTI99
NEGSEX98	ORINO99	REL498	SMOKE99	TYPEMP98	WORTH98
NEGTRE98	ORIPRO99	REL598	SMPAST98	TYPEMP99	WORTH99
NEW98	OSTA2498	RELCAR98	SMPAST99	ULCER98	WRKL198
NEW99	OTHACT99	RELCLS98	SMREST99	ULCER99	WRKL298
NEWSKI98	OVERCO98	RELCLS99	SMWORK99	UNEMPR98	WRKL398
NOINS99	OVERCO99	RELEN198	SMYRS98	UNEMPR99	WRKL498
NOIS299	OWN98	RELEN298	SMYRS99	UNICOV98	WRKL598
NOISE98	OWN99	RELEN398	SOCA99	UNICOV99	WRKL698
NOISE99	PANEL99	RELEN498	SOCSE299	UNIMEM98	WRKL798
NONEG99	PARCAR98	RELEN598	SOCSEC99	UNIMEM99	WRKL898
NOSKIL99	PARENT99	RELHEL98	SPECEQ98	USAGE	WRKL1098
NOTHIR99	PART99	RELPHC99	SPIRIT98	WAGES98	WRKL1198
NOTPRO99	PARTED98	RELPHO98	SPIRIT99	WALK98	WRKL1298
NWRK198	PARTED99	RELSAT98	SPORIT99	WALK99	WRKL1398
NWRK298	PAYEND98	RELST198	SPOSOC99	WEIGHT98	WRKL1498
NWRK398	PAYEND99	RELST298	SPOUS299	WEIGHT99	WRKL1598
NWRK498	PAYPAR98	RELST398	SPOUSE99	WEKWRK98	YEAR98
NWRK598	PAYPAR99	RELST498	SSI99	WEKWRK99	YEAR99
NWRK698	PDAYCA99	RELST598	SSI299	WELFAR98	YENTER
NWRK798	PEN99	RELVIS98	SSOC99	WELFAR99	YOURWA98
NWRK898	PEN299	RELVIS99	SSOC299	WGT9899L	YOURWA99
NWRK998	PENADJ99	RENT99	STA2498	WGT9899X	
NWRK1198	PENPLN99	RESTDK99	STAIR98	WHITE	

Appendix III – Notes on the 1998 – 1999 longitudinal file and development of the longitudinal weights

The 1998-1999 Longitudinal CWSHS file contains all responses from both years' interviews from those participants who were interviewed both years. There are 909 such respondents. The records from the two years have been merged so that there is only one record per person, containing all responses from the two years. In all, 51% of the 1,771 1998 participants responded to the 1999 survey. Most non-response was due to the inability to re-contact the participant, rather than an active refusal to be re-interviewed. Of the 862 non-respondents, 155 (18%) were refusals at the time of re-contact in 1999; an additional 174 (20%) indicated in 1998 that they did not wish to be interviewed again. The remaining 62% were lost to follow-up.

Continuing participants differed from non-continuing participants in many ways. They were older, more likely to be white, more likely to report a chronic condition, more likely to speak English at home, more likely to have a college degree, less likely to report high levels of stress (see question 129 in questionnaire), and less likely to have household incomes under \$20,000 than those who were not interviewed in 1999.

Because of these differences, an adjustment to the 1998 sampling weights is needed that accounts for the differing probability of being followed among different sub-groups of the population. This adjustment for attrition would increase the weight of a person who, based on his/her known characteristics, was unlikely to have been followed-up in 1999, and reduce the weight of someone who is likely to have been followed. For example, white, college-educated respondents in their 50's were fairly likely to be re-interviewed in 1999; in fact, 72% of this group remained in the survey. Therefore, their sampling weights would be adjusted downward, because they make up a relatively larger proportion of the longitudinal dataset than they do of the initial 1998 dataset. The reduction in their sampling weight by use of the adjustment variable described below prevents their characteristics from being overestimated in the longitudinal dataset. Without such an attrition adjustment, one would lose the ability to generalize the findings to the California adult population despite the use of the sampling weights.

To calculate this adjustment for attrition, we developed a logistic regression model that predicted survey participation in 1999. The model contained the following independent variables: age, age², age³ (to capture the curvilinear relationship between age and loss-to-follow-up), race/ethnicity (African American, Latino, Asian/Pacific Islander, vs. non-Latino White), self-reported health status (poor, fair, good, very good, vs. excellent), perceived stress (Q.129a, entered as a linear variable), household income (below 125% of Federal Poverty Level, unknown household income, vs. above 125% FPL), employment status (has job vs. not working).

We took steps to determine that the model adequately fit the data; specific results of the logistic regression model and influence diagnostics are available upon request. Next, we output the predicted probability of being interviewed in 1999, and took its inverse as the attrition adjustment. Thus, persons in the longitudinal file who were more likely to have been followed based on their known characteristics have small values of the attrition adjustment, which reduces the size of their longitudinal weight relative to others in the longitudinal sample. The size of this adjustment ranges from 1.4 to 4.4.

Appendix III – Notes on the 1998 – 1999 longitudinal file and development of the longitudinal weights, continued

To create the longitudinal weight, we multiplied the attrition adjustment (ATTRADJ) by the original 1998 sampling weight (FNLWGT98). Finally, we normalized the longitudinal weights so that the sum of these values is equal to the total sample size of 909. The longitudinal file contains two 1998-1999 longitudinal weight variables, a proportional weight (WGT9899L) and an expansion weight (WGT9899X). In addition, the attrition adjustment variable (ATTRADJ) and the original 1998 sampling weight variable (FNLWGT98) are retained in the longitudinal file. In most cases, however, users will not need either the attrition adjustment nor the original 1998 sampling weight variables. They are included on the file for users experienced in longitudinal sampling issues who may wish to investigate the construction of the longitudinal weights.