Description	CTS Baseline Questionnaire Question(s)	Variable Name	Values	Data Type
Participant Information				
CTS cohort participant identification number		CTS_Cohort_ID	1 unique value per participant	Numeric
Participant's date of birth	2. Please fill in your birthdate.	date_of_birth	MM/DD/YYYY	Date
Participant's date of death		date_of_death	MM/DD/YYYY if participant died before 12/31/2015 Missing if participant did not die before 12/31/2015	Date
Participant's date moved out of California		date_of_move	MM/DD/YYYY if participant moved out of California before 12/31/2015 Missing if participant did not move out of California before 12/31/2015	Date
Participant's study start date		baseline_date	MM/DD/YYYY	Date
The censoring date for current analyses		analysis_censoring_date	12/31/2015 if participant was alive and in CA as of 12/31/2015 Participant's date of death if participant died before 12/31/2015 Participant's date of move if participant moved out of California before 12/31/2015	Date
Participant Characteristics				
Age in years at the time of baseline questionnaire		age_at_baseline	Range: 22-104	Numeric
Participant's self-reported race/ethnicity	6. To what race/ethnic group do you and your biological parents belong?	race	Range: 0-6 0=None reported by participant 1=White 2=Black 3=Hispanic 4=Native American 5=Asian/Pacific Islander 6=Other/Mixed	Numeric
Quartiles of socioeconomic status (SES) index in the 1990 Census block that included the participant's CTS baseline address		SES_in_1990CensusBlock	1=1st quartile 2=2nd quartile 3=3rd quartile 4=4th quartile Missing if unknown	Numeric
Height in inches at baseline	65. What are your height and weight today, and what were they when you were age 18?	height_at_baseline	Range: 45-81 999=Unknown	Numeric
Weight in pounds at baseline	65. What are your height and weight today, and what were they when you were age 18?	weight_at_baseline	Range: 67-450 999=Unknown	Numeric

Description	CTS Baseline Questionnaire Question(s)	Variable Name	Values	Data Type
Calculated body mass index (BMI; kg/m^2) at baseline	65. What are your height and weight today, and what were they when you were age 18?	BMI_at_baseline	Range: 16.0-75.0 Missing if unknown	Numeric
Calculated BMI (kg/m^2) at age 18, based on self-reported height and weight at age 18 years	65. What are your height and weight today, and what were they when you were age 18?	BMI_at_age18	Range: 15.0-54.8 Missing if unknown	Numeric
Reproductive History				
Age at menarche	21. How old were you when you had your first menstrual period?	age_at_menarche	Range: 1-10 1=Never had a menstrual period 2=<10 3=10 4=11 5=12 6=13 7=14 8=15 9=16 10=17+ 999=Unknown	Numeric
Total years used oral contraceptives at baseline	24. Did you ever take birth control pills (oral contraceptives) for one month or longer? 25. How old were you when you first used birth control pills, and (if no longer taking) how old were you when you last used them?	total_yrs_ocps	Range: 0-8 0=No oral contraceptive use, or born before 1915 1=<1 2=1-2 3=3-4 4=5-9 5=10-14 6=15-19 7=20-24 8=25+ 999=Unknown	Numeric
Total number of pregnancies at baseline	28. Please indicate the outcome of each of your pregnancies next to the age when it ended (consider a multiple birth as one pregnancy).	parity_at_baseline	Range: 0-15 999=Unknown	Numeric

Description	CTS Baseline Questionnaire Question(s)	Variable Name	Values	Data Type
Menopausal status at baseline	34. Have your menstrual periods stopped permanently? 35. When did you have your last period? 36. Why did your periods stop?	meno_status_at_baseline	Range: 1-11 1=Premenopausal 2=Peri-menopausal 3=Postmenopausal: natural menopause 4=Postmenopausal: bilateral oophorectomy 5=Postmenopausal due to medication/chemo/radiation 6=Postmenopausal due to other reason 7=Postmenopausal: hysterectomy at age 56+ 8=Postmenopausal: hysterectomy at age <56 but aged 56+ at baseline 9=Postmenopausal: unknown type of menopause 999=Unknown menopausal status	Numeric
Summary variable for menopausal hormonal therapy at baseline: includes past vs. current and estrogen vs. estrogen + progestin	41. Have you ever taken estrogen ("female hormones") for symptoms of menopause (the change of life) or for other reasons? 42. Have you ever taken Premarin by mouth (as pills)? 44. At what age did you last take Premarin? 47. Have you ever used estrogen other than Premarin pills? 49. At what age did you last use estrogens other than Premarin pills? 50. Have you ever used progesterone or a progestin? 53. At what age did you last take progesterone or progestin?	hormone_therapy_use	Range: 0-4 0=No menopausal hormone therapy use 1=Past menopausal hormone therapy use 2=Current use of estrogen alone 3=Current use of estrogen + progestin 4=Progestin only 999=Unknown	Numeric
Self-reported hysterectomy status at baseline	37. Have you ever had a hysterectomy (that is, surgery to remove your uterus or womb)?	hysterectomy_at_baseline	1=Never had hysterectomy 2=Yes, had hysterectomy 999=Unknown	Numeric
Self-reported oophorectomy status at baseline	38. Have you ever had an ovary removed?	oophorectomy_at_baseline	1=Never 2=One Ovary Removed 3=Both Ovaries Removed 999=Unknown	Numeric
Self-reported tubal ligation status at baseline	40. Have you ever had your "tubes tied" (tubal ligation)?	tubal_ligation_at_baseline	1=Never had tubal ligation 2=Yes, had tubal ligation	Numeric
Personal and Family Health History				
Personal history of any type of cancer at baseline	67. Have you or any relatives had any of the following cancers or other medical conditions?	personal_history_of_cancer	0=No 1=Yes Missing if unknown	Numeric

Description	CTS Baseline Questionnaire Question(s)	Variable Name	Values	Data Type
First-degree family history of any type of cancer at baseline	67. Have you or any relatives had any of the following cancers or other medical conditions?	family_history_of_cancer	Range: 0-3 0=No family history of specific cancer 1=Yes 1+ 1st degree relative 2=Adopted 3=No info on family history Missing if unknown	Numeric
Personal history of high blood pressure at baseline	67. Have you or any relatives had any of the following cancers or other medical conditions?	high_blood_pressure	0=No 1= Yes Missing if unknown	Numeric
Personal history of diabetes at baseline	67. Have you or any relatives had any of the following cancers or other medical conditions?	diabetes_at_baseline	0=No 1= Yes Missing if unknown	Numeric
Personal history of heart attack or stroke at baseline	67. Have you or any relatives had any of the following cancers or other medical conditions?	CVD_at_baseline	0=No 1=Yes Missing if unknown	Numeric
Physical Activity				
Hours per week of strenous or moderate exercise during the 3 years before baseline	68. How often did you participate in strenuous exercise activities or sports (e.g., swimming laps, aerobic, calisthenics, running, jogging, basketball, cycling on hills, racquetball)? 69. How often did you participate in moderate exercise activites or sports (e.g., brisk walking, golf, volleyball, cycling on level streets, recreational tennis, or softball)?	Hrswk_exercise_3yrs	Range: 0-24 999=Unknown	Numeric
Hours per week of strenous or moderate exercise during the participant's lifetime to date	68. How often did you participate in strenuous exercise activities or sports (e.g., swimming laps, aerobic, calisthenics, running, jogging, basketball, cycling on hills, racquetball)? 69. How often did you participate in moderate exercise activites or sports (e.g., brisk walking, golf, volleyball, cycling on level streets, recreational tennis, or softball)?	Hrswk_exercise_life	Range: 0-24 999=Unknown	Numeric
Medication and Vitamin Use				
Total years taking any non-steroidal anti-inflammatory drug (NSAID) at baseline	66. Have you taken any of the following medications regularly (at least once a week)? If so, indicate how many total years you took it and how often you took it.	duration_NSAID_use	Range: 1-8 1=Not a regular user of NSAIDs 2=<1 3=1 4=2 5=3-4 6=5-9 7=10+ 8=Regular user but duration unknown 999=Unknown	Numeric

Description	CTS Baseline Questionnaire Question(s)	Variable Name	Values	Data Type
Frequency of NSAID use at baseline	66. Have you taken any of the following medications regularly (at least once a week)? If so, indicate how many total years you took it and how often you took it.	frequency_NSAID_use	Range: 1-5 1=1-3 2=4-6 3=Every day 4=Regular user but frequency of NSAID use unknown 5=Did not take regularly 999=Unknown	Numeric
Current multivitamin and single vitamin use at baseline	71. During the past year have you taken any vitamins or minerals regularly (at least once a week)?	vitamin_use	Range: 1-4 1=Non-user 2=Current single vitamin only 3=Current multivitamin only 4=Current multivitamin + single vitamin Missing if unknown	Numeric
Alcohol and Tobacco Use				
Grams of alcohol (beer, wine, and liquor) consumed per day in the year before baseline: none, <20 grams/day or >20 grams/day	83. Please indicate how much and how often you usually drank each beverage.	alcohol_gramsday_pstyr	0=No alcohol consumption 1=<20 grams/day 2=≥20 grams/day Missing if unknown	Numeric
Smoking status at baseline	84. Have you smoked at least 100 cigarettes in your entire life? 85. How old were you when you first smoked fairly regularly, and (if no longer smoking) how old were you when you last smoked?	smoking_status	1=Never smoked 2=Former smoker 3=Current smoker Missing if unknown	Numeric
Total years smoked at baseline	85. How old were you when you first smoked fairly regularly, and (if no longer smoking) how old were you when you last smoked?	total_years_smoked	Range: 0.5-75 Missing if unknown	Numeric
Cancer Endpoints				
Indicator variables for inclusion/exclusion criteria				
Was participant living in California at baseline?		California_resident_at_baseline	0=California resident at baseline 1=Participant lived outside California at baseline	Numeric
Did participant consent to only breast cancer research, or to all CTS research?		Consent_status	0=Participant consented to all CTS research 1=Participant consented to breast cancer research only	Numeric
The data of 2 CTS participants who asked to be removed from the study has been excluded from all analyses. Four additional participants are routinely excluded due to the incomplete or out-of-order nature of their questionnaires. This brings our eligible population for CTS analyses to 133,473.				

Description	CTS Baseline Questionnaire Question(s)	Variable Name	Values	Data Type
Follow up questionnaires				
Questionnaire 2 completion date		date_completed_questionnaire2	MM/DD/YYYY if participant completed questionnaire 2 Missing if participant did not complete questionnaire 2	Date
Questionnaire 3 completion date		date_completed_questionnaire3	MM/DD/YYYY if participant completed questionnaire 3 Missing if participant did not complete questionnaire 3	Date
Questionnaire 4 completion date		date_completed_questionnaire4	MM/DD/YYYY if participant completed questionnaire 4 Missing if participant did not complete questionnaire 4	Date
Questionnaire 5 completion date		date_completed_questionnaire5	MM/DD/YYYY if participant completed questionnaire 5 Missing if participant did not complete questionnaire 5	Date
Cancer				
Cancer status at baseline		cancer_survivor_at_baseline	0=No Personal History of Cancer at Baseline 1=CCR-Based Cancer Before Baseline 2=Self-Reported Cancer on Baseline Questionnaire	Numeric
Number of times participant has been diagnosed with cancer		Ccr_count	Range: 1-8	Numeric
SEER codes, diagnosis dates, cancer behavior, cancer histology, ICD-O codes, and cancer stage have been included for up to participants' 8th cancer diagnosis, because 8 was found to be the greatest number of cancers for any CTS participant since the baseline questionnaire.				
SEER site recode for participant's cancer diagnosis (1st through 8th)		cancer_SEER_site1 cancer_SEER_site2 cancer_SEER_site3 cancer_SEER_site4 cancer_SEER_site5 cancer_SEER_site6 cancer_SEER_site7 cancer_SEER_site8	5-digit SEER site code e.g. 37000 Missing if participant was not diagnosed with this cancer	Numeric

Description	CTS Baseline Questionnaire Question(s)	Variable Name	Values	Data Type
Diagnosis date for participant's cancer diagnosis (1st through 8th)		cancer_diagnosis_date1 cancer_diagnosis_date2 cancer_diagnosis_date3 cancer_diagnosis_date4 cancer_diagnosis_date5 cancer_diagnosis_date6 cancer_diagnosis_date7 cancer_diagnosis_date8	MM/DD/YYYY if participant was diagnosed with this (1st, 2nd, 3rd, etc.) cancer Missing if participant was not diagnosed with this cancer	Date
Cancer behavior SEER code for participant's cancer diagnosis (1st through 8th)		cancer_behavior1 cancer_behavior2 cancer_behavior3 cancer_behavior4 cancer_behavior5 cancer_behavior6 cancer_behavior7 cancer_behavior8	0=Benign in both ICD-O-2 and ICD O-3 1=Borderline malignancy in both ICD-O-2 and ICD-O-3 2=In situ in both ICD-O-2 and ICD-O-3 3=Malignant (invasive) in both ICD-O-2 and ICD-O-3 4=Only malignant in ICD-O-3 5=No longer reportable in ICD-O-3 6=Only malignant for cases diagnosed 2010+ Missing if participant was not diagnosed with this cancer	Char
Histology code for participant's cancer diagnosis (1st through 8th)		cancer_histology1 cancer_histology2 cancer_histology3 cancer_histology4 cancer_histology5 cancer_histology6 cancer_histology7 cancer_histology8	4-digit histology code e.g. 8046 Missing if participant was not diagnosed with this cancer	Numeric
Cancer ICD_0 code for participant's cancer diagnosis (1st through 8th)		cancer_ICDO_site1 cancer_ICDO_site2 cancer_ICDO_site3 cancer_ICDO_site4 cancer_ICDO_site5 cancer_ICDO_site6 cancer_ICDO_site7 cancer_ICDO_site8	4-digit ICD-0 code beginning with C e.g. C209 Missing if participant was not diagnosed with this cancer	Numeric
Summary stage code for participant's cancer at the time of diagnosis (1st through 8th)		cancer_stage1 cancer_stage2 cancer_stage3 cancer_stage4 cancer_stage5 cancer_stage6 cancer_stage7 cancer_stage8	0=In situ 1=Localized 2=Regional by direct extension 3=Regional by lymph nodes 4=Regional by direct extension and lymph nodes 5=Regional, NOS 7=Remote 8=Not abstracted 9=Unknown or not specified Missing if participant was not diagnosed with this cancer	Numeric