Health Information National Trends Survey

hints

Health Information National Trends Survey 5 (HINTS 5)

HINTS-SEER Methodology Report

NOVEMBER 7, 2023

Prepared for
National Cancer Institute
9609 Medical Center Drive
Bethesda, MD 20892-9760

Prepared by
Westat
1600 Research Boulevard
Rockville, MD 20850

Revised February 17, 2025 in accordance with the January 20, 2025 Executive Order titled “Defending Women From Gender Ideology Extremism and Restoring Biological Truth to the Federal Government."

Westat

Table of Contents

Appendixes

Tables

Tables **Pages**

1. HINTS-SEER Overview

Purpose

The Health Information National Trends Survey (HINTS) is a nationally-representative survey which has been administered every few years by the National Cancer Institute (NCI) since 2003. Up to this point, HINTS has always been fielded with a probability-based sample of civilian, non-institutionalized adults living in the United States. HINTS has enabled NCI to monitor population trends in cancer communication practices, information preferences, risk behaviors, attitudes, and cancer knowledge. While the cancer communication needs of the general public continue to be a priority for NCI, the agency identified an additional priority to specifically examine the information support needs and cancer communication experiences of people who have been diagnosed with cancer at some point in their lives, commonly called "cancer survivors." Probability sampling inevitably leads to some cancer survivors being included in each round of HINTS data collection. To date, cancer survivors represent approximately nine percent of the total HINTS sample across administrations from 2003 to 2020 (N=5,762). Individual HINTS datasets include between 500 and 600 survivors per round of data collection. HINTS data users have requested a more robust survivor sample in order to assess responses by cancer type and years since diagnosis. As of 2019, 35 peer-reviewed HINTS publications had focused on the experiences of cancer survivors (Finney Rutten, et al., 2020). Those papers have been limited by the small sample sizes of survivors available in individual HINTS administrations.

To address this need, NCI developed a pilot project to oversample cancer survivors for HINTS using selected cancer registries from the Surveillance, Epidemiology, and End Results (SEER) Program (https://seer.cancer.gov) as a sampling frame of cancer survivors. SEER, also supported by NCI, collects cancer incidence and survival data from population-based cancer registries covering approximately 50 percent of the U.S. population. These registries routinely collect data on cancer survivor demographics, primary tumor site, tumor morphology and stage at diagnosis, first course of treatment, and follow-up for vital status (survival). These data are collected on every cancer case within each of the SEER cancer registry catchment areas.

The pilot project, called HINTS-SEER, was designed to provide a larger sample of cancer survivors for analysis using HINTS 5, Cycle 4 (2020) survey items and topics, in addition to other topics

relevant to cancer survivors. A unique aspect of HINTS-SEER compared to other iterations of HINTS is that in the HINTS-SEER dataset, key data elements from the cancer registry datasets are linked to the HINTS survey responses, providing a more in-depth view of each respondent's cancer diagnosis. Several noted research gaps will be possible to examine with the larger sample of cancer survivors obtained through HINTS-SEER. They include but are not limited to: barriers to cancer communication and cancer care; late effects of cancer treatment; the effect of cancer and cancer treatment on health insurance, healthcare access, employment, personal relationships, financial stability, daily activities, cognitive functioning, media use, and emotional health; and unmet care planning needs, including the communication-specific needs of long-term cancer survivors.

Methodology Overview

HINTS-SEER aimed to maintain as much of the established HINTS methodology as possible, while making a few alterations to accommodate the specialized sample. The most important alterations were to the sampling procedures. Three cancer registries were used as the basis for the HINTS-SEER sample: Iowa Cancer Registry, New Mexico Tumor Registry, and the Greater Bay Area Cancer Registry. Because each registry has its own IRB, database structure, and consenting procedures, Westat worked closely with each registry to accommodate individual registry requirements. The details of the selection of registries and the sampling procedures can be found in Chapter 2.

Data collection for HINTS-SEER adhered as closely as possible to the traditional HINTS data collection through the mail, involving three survey mailings and a $2 pre-paid incentive. Because the registries provided their HINTS-SEER samples to Westat at different times, each registry cohort was fielded individually starting in January 2021. Details about data collection are provided in Chapter 3.

Survey Instrument

The foundation for the HINTS-SEER instrument was the HINTS 5, Cycle 4 instrument which was fielded in early 2020. Edits were made to this instrument both to make individual items more appropriate for cancer survivors versus the general population, as well as to add some topics of specific concern for those who have a personal history of cancer. In addition, the timing of the data

collection warranted the inclusion of a section of questions about the COVID-19 pandemic's effect on the respondent as a cancer survivor. Specific survey item edits and additions are listed below.

- An item about whether the respondent had spoken to a mental health professional was added (item C6).
- The Cancer History section was moved to an earlier part of the instrument and questions were added about the respondent's cancer treatment as well as the physical, financial, and work impacts of their cancer diagnosis (Items E3 through E8).
- A section of items was added specifically about the COVID-19 pandemic (Section F). Questions asked about COVID's impact on cancer treatment, follow-up cancer care, cancer screening, and preventive care. Also included were questions about patient-provider discussions and trust in sources of COVID information.
- Questions were added to determine the respondent's experience with genetic testing and precision medicine related to their cancer (Items G4, G5, G11, and G12).
- Activities of Daily Living items were added in Section J to assess the respondent's current ability to care for themselves (Items J4-J11).
- A series of items to measure social isolation were added (Items J12-J15).

Although the standard HINTS procedure is to cognitively test any new survey content or any HINTS item that is altered, time did not allow for cognitive testing of the full HINTS-SEER instrument. However, many of the items added to the HINTS-SEER instrument had been cognitively tested and fielded on HINTS in earlier cycles, and others had only minor edits to their original wording. The new COVID-19 items were pre-tested in an online survey that was being conducted as part of another NCI study. The full HINTS-SEER instrument can be found in **Appendix A**.

2. Sample Selection

2.1 SEER Registry Selection

To determine which registries would participate in HINTS-SEER, the project was presented to the SEER Research Group, which facilitates research across the SEER registries, in July 2019, after which Westat sent out a short survey to SEER Registry directors in the fall of 2019. The responses identified the registries that were interested in participating and provided initial information about each registry's procedures, including their consent procedures. Following review of the registries' responses, Westat estimated the number of qualifying cancer survivors, the percent minority, the percent rural, and the region of the country for each SEER registry that responded. Based on this information and the survey data, four registries were selected and asked to participate in HINTS-SEER, with the goal of having diversity in terms of the racial/ethnic and geographic composition of the cancer survivor populations represented by the selected registries.

Table 2-1. Estimated number of qualifying cancer survivors, percent minority, percent rural, and census region by SEER registry

SEER registry	Estimated number of qualifying cancer survivors	Percent minority	Percent rural	Census region
Iowa	130,854	9	36	Midwest
Georgia	299,012	41	25	South
Greater Bay Area	270,217	57	2	West
New Mexico	70,151	55	23	West

Westat developed a plan to sample the sites in proportion to the number of cancer survivors in the records of each cancer registry. The estimated response rate to the HINTS-SEER survey (based on previous HINTS experience) was used to calculate the target sample that would be needed to get a total of at least 1,000 completed HINTS-SEER surveys. Because of the larger size of the Georgia Tumor Registry and the Greater Bay Area Cancer Registry, the plan included a larger sample from those registries. See Table 2-2 below for the initial sample of consented respondents requested from each registry. It should be noted that although these were the sample size goals, the number of cases received from each registry varied and is outlined in Sections 2.5, 2.6 and 2.7.

Table 2-2. Sample size targets of consented respondents to be invited to participate in HINTS, per registry

Registry	Target starting HINTS sample size
Georgia Cancer Registry	1,022
Iowa Cancer Registry	447
New Mexico Tumor Registry	240
Greater Bay Area Cancer Registry	923
Total	**2,632**

Once selected, Westat worked with each registry to address their specific needs in order to participate. Unfortunately, due to staffing and other issues resulting from the COVID-19 pandemic in 2020, the Georgia Tumor Registry was unable to continue involvement with HINTS-SEER. The remaining three registries went on to conduct the sampling procedures outlined below.

2.2 Eligibility Criteria

Cancer survivors that met the following criteria were included in the sampling frame for each participating SEER registry:

1. **Cancer survivors with a vital status of alive.**

2. **Survivors with Cancer Diagnosis, Specified to Invasive Cancers.** Because the SEER registries track all tumors, this criterion was designed to ensure that all cancer survivors sampled for HINTS had a history of cancer rather than a non-malignant tumor.

3. **Survivors Older Than 18 Years of Age.** The Iowa Cancer Registry and New Mexico Tumor Registry sampled people that were age 18 as of as of December 31, 2020. The Greater Bay Area Cancer Registry used the date of December 1, 2020. This criterion was included to ensure that the cancer survivor met the HINTS requirement that the survey respondent is an adult.

4. **Survivors Whose Last Contact was no Earlier than January 1, 2016.** "Last contact" is defined as the last time the SEER registry had either active contact with the cancer survivor (through a study, for example) or passive contact (by linking the cancer survivor's data with another dataset, for example). January 1, 2016 was the latest date available for the current study and was included in the eligibility criteria to reduce the likelihood of nonresponse due to incorrect mailing addresses.

5. **Survivors with Date of Diagnosis Prior to 2018 Based on Certified Data.** Certified SEER data are data that have undergone cleaning and vetting by central cancer registry staff according to the national standard for clearance and inclusion in national datasets and statistics. Cancer registry data has a two-year lag period for certification to allow for the ongoing data collection of treatment and other follow up data.

Excluded from the sampling procedures were survivors whose only diagnosis was non-melanoma skin cancer. Survivors with a diagnosis of non-melanoma skin cancer in addition to another cancer were included.

For cancer survivors that had more than one tumor, one eligible tumor was sampled using SEER cancer sequence numbers. The sequence number indicates the order in which a reportable primary tumor is discovered in relation to the total number of primaries for a given patient. A lower cancer sequence number represents an earlier or more aggressive tumor (if two primary tumors are diagnosed at the same time), with a person's first primary tumor coded as 00 (if no other primary tumors) or 01 (if first of multiple tumors). However, if a participant's first (or second) cancer did not meet the inclusion criteria for the study (for example, a non-melanoma skin cancer), then the next tumor was selected, and its corresponding sequence number is therefore higher (e.g., 02, 03). For the small percent of participants who had multiple eligible tumors, one eligible tumor was chosen for reporting. In Iowa and New Mexico, the selection of this tumor was done in a *systematic* way, with the lowest cancer sequence number selected. The Greater Bay Area Cancer Registry used a different tumor sampling method, in which, for people with multiple eligible, the tumor was selected *randomly* rather than systematically, which led to slightly more cases having a sequence number higher than 01 among respondents. However, the overall percent of cases with a first primary tumor (CSEQ 00 or 01) selected was high across the three registries for respondents (Iowa 98.5%, GBACR 90.2%, New Mexico 96.5%). Additional details on tumor sampling by cancer sequence number follow below.

2.3 Implicit Stratification

Eligible cancer survivors were sorted by two characteristics: years since diagnosis and race/ethnicity. The values of both characteristics were further grouped into categories as follows:

- Years since diagnosis
  - Less than 5 years
  - 5 Years to 9 years
  - 10 or more years
- Race/ethnicity
  - Hispanic regardless of race
  - Non-Hispanic African American alone or in combination with another race(s)
  - All other race groups

The use of implicit stratification along with the systematic sampling procedure described in the next section ensured that the sample of eligible cancer survivors selected from each registry was proportionally represented with respect to the above categories on each sampling frame.

2.4 Sampling Procedures

Westat developed a SAS program that implemented the sampling as follows:

1. Created the sampling frame by extracting records meeting the eligibility criteria
2. Sorted the variables by the stratification categories
3. Systematically selected records from the sorted frame by:
   a. Computing a sampling interval (K) by dividing the number of records on the frame (N) by mail out sample size (n): K = N/n.
   b. Computing a random starting point (RS) between 1 and the sampling interval.
   c. Selecting the record corresponding to random starting point and every K^{th} record after that where K corresponds to sampling interval until n records were selected.

This program was provided to each of the participating SEER registries to assist in drawing the sample of addresses to be used for the study. The consent and sampling procedures followed by each participating SEER registry are further described below.

2.5 Iowa Cancer Registry Consent and Sampling

Consent

The Iowa Cancer Registry's procedures include an active consent process that requires the registry to contact potential study participants in advance and get a signed release form before their contact information could be shared with Westat for the HINTS study. Based on the HINTS request for a sample of 447 Iowa participants to be included in the sampling frame, the Iowa Cancer Registry determined that they needed to send out 6,433 release forms based on a 6% anticipated agreement rate. The Iowa Cancer Registry used the Westat-provided sampling program (including the sampling of first eligible primary tumor) to identify and select the potential respondents on August 14, 2020, and sent the release form mailing on September 8, 2020.

Sampling

On December 2, 2020, the Iowa Cancer Registry delivered to Westat a file of 482 addresses of registry cancer survivors who consented to be contacted to participate in the HINTS-SEER study. Westat conducted a review of the address information to ensure completeness. Selected characteristics, including the cancer sequence number, of the final sample of 482 cases from the Iowa Cancer Registry are detailed below.

Table 2-3. Selected demographic distributions of the Iowa Cancer Registry sample

Selected demographics	Number	Percent
Sex		
Male	218	45.23
Female	264	54.77
Total	**482**	**100.00**
Race/ethnicity		
White	468	97.10
Black	<5	0.21
American Indian or Native American	<5	0.62
Asian/Pacific Islander	<5	0.62
Hispanic (any race)	<5	0.62
Unknown	<5	0.83
Total	**482**	**100.00**
Age at time of survey		
90 or more years	17	3.53
80-89 years	101	20.95
70-79 years	166	34.44
60-69 years	126	26.14
50-59 years	50	10.37
40-49 years	17	3.53
30-39 years	<5	0.83
18-29 years	<5	0.21
Total	**482**	**100.00**
Years since diagnosis		
Diagnosis between 2014-2018	160	33.19
Diagnosis between 2009-2013	113	23.44
Diagnosis from 2008 or earlier	209	43.37
Total	**482**	**100.00**
Cancer sequence number*		
00	321	66.60
01	91	18.88
02	35	7.26
03	35	7.26
Total	**482**	**100.00**

*The Iowa Cancer Registry conducted systematic sampling wherein the lowest eligible cancer sequence number was chosen for the sample.

2.6 Greater Bay Area Cancer Registry Consent and Sampling

Consent

The Greater Bay Area Cancer Registry does not pre-consent study participants to provide mailing addresses, but expects that respondents will be consented during the study procedures. The Greater Bay Area Cancer Registry's consent form, which was on the inside cover of the survey instrument, can be found in **Appendix B**. As was required by the state of California's Committee for the

Protection of Human Subjects (CPHS), data from any respondent sending back a survey without a signed consent form was discarded unless they could be re-consented (procedures described below).

Sampling

In response to the Greater Bay Area Cancer Registry's protocols, Westat modified the SAS program (described in section 2.4) in the following ways:

- Excluded tumors where the ordering of the tumors were missing, not applicable, or unknown.
- Edited the exclusion criteria for the non-melanoma skin cancer to include cases where the cancer survivor had at least one additional invasive cancer in their lifetime other than non-melanoma skin cancer.
- Edited the tumor sequence selection to select the tumor at random from the eligible tumors.

On January 21, 2021 the Greater Bay Area Cancer Registry delivered to Westat a file of 2,000 case listings selected for the study. Although HINTS had originally needed just 923 cases, the number of respondents sent by the Greater Bay Area Cancer Registry was based on the maximum number that can be requested from the registry. Westat conducted a review of the address information of the received sample and seven cases were removed from the sample because of incomplete addresses that could not be verified. Selected characteristics of the final sample of 1,993 cases from the Greater Bay Area Cancer Registry, including cancer sequence, are detailed below.

Table 2-4. Selected demographic distributions of the Greater Bay Area Cancer Registry sample

Selected demographics	Number	Percent
Sex		
Male	949	47.62
Female	1,043	52.33
Total	**1,993**	**100.00**
Race/ethnicity		
White	1,302	65.33
Black	94	4.72
American Indian or Native American	<5	
Asian/Pacific Islander	342	17.16
Hispanic (any race)	201	10.09
Other	<5	
Unknown	47	2.36
Total	**1,993**	**100.00**
Age at time of survey		
90 or more years	156	7.83
80-89 years	428	21.48
70-79 years	622	31.21
60-69 years	425	21.32
50-59 years	229	11.49
40-49 years	75	3.76
30-39 years	41	2.06
18-29 years	17	0.85
Total	**1,993**	**100.00**
Years since diagnosis		
Diagnosis between 2014-2018	533	26.75
Diagnosis between 2009-2013	493	24.73
Diagnosis from 2008 or earlier	967	48.52
Total	**1,993**	**100.00**
Cancer sequence number*		
00	1,643	82.44
01	203	10.19
02	130	6.52
03	16	0.80
04	1	0.05
Total	**1,993**	**100.00**

*The Greater Bay Area Cancer Registry conducted random sampling of eligible cancer sequence number to be chose for inclusion in the sample.

2.7 New Mexico Tumor Registry Consent and Sampling

Consent

The New Mexico Tumor Registry's procedures include a passive consent process that requires the registry to contact potential study participants and collect study refusals. Cancer survivors that did not respond to the registry to refuse to participate in the study were assumed to have consented to

be contacted to participate in the HINTS-SEER study. Based on the HINTS request for a sample of 240 New Mexico participants, the New Mexico Tumor Registry used the Westat-provided sampling program (described in section 2.4), including the instructions to select the first eligible tumor, to select 1,400 survivors for the consent mailing. The number of survivors selected for the consent mailing was determined by the New Mexico Tumor Registry based on their past experience with similar requests. The consent mailing was sent out over several days in January 2021.

Sampling

The New Mexico Tumor Registry excluded the following cases from their dataset according to the registry's internal policies and procedures:

- Carcinoid tumors;
- Native Americans;
- Cancer survivors who were flagged as "Do not contact" in their registry;
- Cancer survivors seen at Veterans Affairs only; and
- Cancer survivors with invalid addresses such as a correction center, nursing home, assisted living, hospice, or social service.

On February 25, 2021, the New Mexico Tumor Registry delivered to Westat a file of 850 registry cancer survivors who did not refuse to participate in the HINTS study. Westat conducted a review of the address information and also worked with the registry to implement additional deletions from the sample. Per the registry's request, nine cases were removed from the sample. Three of the cases were additional refusals, five cases were marked as having "bad addresses" by the registry, and one case was deceased.

The final sample of 841 cases from the New Mexico Tumor Registry is detailed in Table 2-5.

Table 2-5. Selected demographic distributions of the New Mexico Tumor Registry sample

Selected demographics	Number	Percent
Sex		
Male	367	43.64
Female	474	56.36
Total	**841**	**100.00**
Race/ethnicity		
White	513	61.00
Black	11	1.31
American Indian or Native American	0	0.00
Asian/Pacific Islander	10	1.19
Hispanic (any race)	297	35.32
Other	<5	0.24
Unknown	8	0.95
Total	**841**	**100.00**
Age at time of survey		
90 or more years	41	4.88
80-89 years	151	17.95
70-79 years	263	31.27
60-69 years	223	26.52
50-59 years	92	10.94
40-49 years	34	4.04
30-39 years	27	3.21
18-29 years	10	1.19
Total	**841**	**100.00**
Years since diagnosis		
Diagnosis between 2014-2018	270	32.10
Diagnosis between 2009-2013	214	25.45
Diagnosis from 2008 or earlier	357	42.45
Total	**841**	**100.00**
Cancer sequence number*		
00	729	86.68
01	85	10.11
02	25	2.97
03	1	.12
04	1	.12
Total	**841**	**100.00**

*The New Mexico Tumor Registry conducted systematic sampling wherein the lowest eligible cancer sequence number was chose for inclusion in the sample.

2.8 SEER Data

In addition to the contact information for sampled cancer survivors, the registries also provided information about the sampled survivors' cancers to be included in the HINTS-SEER dataset. These variables included:

- Primary cancer site: where the cancer was located in the body;
- Cancer histology: the type of tissue from which the cancer originated;
- SEER summary stage: tumor stage at diagnosis.
- Date of diagnosis: the date that the cancer survivor was diagnosed based on the selected tumor.

3. Data Collection

Data collection for HINTS-SEER started on January 11, 2021 and concluded on August 20, 2021. The survey was conducted exclusively by mail with a $2 pre-paid monetary incentive to encourage participation. The specific mailing procedures and outcomes for this data collection effort are described in detail below.

3.1 Mailing Protocol

The mailing protocol for all three HINTS-SEER cohorts (Iowa Cancer Registry, Greater Bay Area Cancer Registry, and New Mexico Tumor Registry) followed a modified Dillman approach (Dillman, et al., 2009) with a total of four mailings: an initial mailing, a reminder postcard, and two follow-up mailings. Individuals in each sample received the first mailing and reminder postcard, while only non-respondents received the subsequent survey mailings. The second survey mailing was sent via USPS Priority Mail, while all other mailings were sent First Class. The HINTS-SEER questionnaire was administered in English only.

The survey and contact materials for the Greater Bay Area cohort differed slightly from the materials that were developed for the Iowa and New Mexico cohorts. The state of California's Committee for the Protection of Human Subjects (CPHS) required respondents from the Greater Bay Area to sign and return a consent form with their completed questionnaire. Thus, a separate version of the questionnaire was developed for respondents from the Greater Bay Area Cancer Registry sample. This version of the instrument included a consent form inside the front cover of the survey that respondents from the Greater Bay Area could sign if they agreed to participate in HINTS. In addition, respondents from the Greater Bay Area received a patient notification brochure (developed by the California Cancer Registry—the parent organization of the Greater Bay Area Cancer Registry) with their first mailing. The California Cancer Registry (CCR) provided 2,000 copies of this brochure to Westat to include with the Greater Bay Area's first mailing. Furthermore, after obtaining a second approval from the state of California, a fourth mailing to the Greater Bay Area was conducted to obtain consent from respondents who completed a survey but did not sign the consent form included with their survey. This fourth mailing included a cover letter, the Greater Bay Area Cancer Registry consent form, and one postage-paid return envelope.

The contents of the mailings are further described in Table 3-1. The cover letters for Iowa and New Mexico and the reminder postcard for all three cohorts can be found in **Appendix C**. The cover letters for the Bay Area can be found in **Appendix D**. Each cover letter included a list of Frequently Asked Questions (FAQs) on the back. The FAQs for Iowa and New Mexico are in **Appendix E**. The FAQs prepared for the Bay Area as well as the patient notification brochure provided by the CCR are in **Appendix F**.

Table 3-1. Mailing protocol for HINTS-SEER

Mailing	Date(s) mailed	Mailing method	Cycle 3 materials
Mailing 1	• Iowa: January 11, 2021 • Greater Bay area: February 16, 2021 • New Mexico: March 22, 2021	1st Class Mail	• Cover letter with FAQs • Questionnaire • Postage-paid return envelope • $2 bill • Patient Notification Brochure (Bay Area, only)
Postcard	• Iowa: January 19, 2021 • Greater Bay Area: February 23, 2021 • New Mexico: March 29, 2021	1st Class Mail	• Reminder/thank you postcard
Mailing 2	• Iowa: February 10, 2021 • Greater Bay Area: March 17, 2021 • New Mexico: April 21, 2021	USPS Priority Mail	• Cover letter with FAQs • Questionnaire • Postage-paid return envelope
Mailing 3	• Iowa: March 3, 2021 • Greater Bay Area: April 7, 2021 • New Mexico: May 12, 2021	1st Class Mail	• Cover letter with FAQs • Questionnaire • Postage-paid return envelope
Mailing 4	• Greater Bay Area: July 9, 2021	1st Class Mail	• Cover letter with FAQs • Consent form • Postage-paid return envelope

The number of packets sent per mailing is outlined in Table 3-2. Individuals who sent in completed questionnaires were removed from further mailings. In addition, individuals with packets that were returned by the Postal Service as undeliverable were removed from any further mailings.

Table 3-2. Number of packets per mailing by cohort

Mailing	Iowa	Greater bay area	New Mexico	Total
Mailing 1	482	1,993	**841**	3,316
Mailing 2	161	1,488	**562**	2,211
Mailing 3	87	1,188	**484**	1,759
Mailing 4	N/A	158	N/A	158
Total	**730**	**4,827**	**1,887**	**7,444**

3.2 In-bound Telephone Calls

A toll-free telephone number was provided to all respondents. This number was provided in each mailing. Respondents were told that they could call the number if they had any questions or concerns about HINTS. This number had a HINTS-specific voicemail message that instructed callers to leave their contact information and the reason for the call and that a study staff member would return their call. When voicemails were received, they were logged into the Study Management System (SMS) and the request was either processed (such as recording their desire for an additional copy of the questionnaire) or the respondent was called back to ascertain the respondent's need if it was not clear from the message. Callers stating that they did not want to participate in the study were coded as "refusal" and removed from any subsequent mailings.

The toll-free line received 26 calls throughout the HINTS-SEER field period (see Table 3-3 below). The majority of in-bound calls were refusals or callers who wanted to let the study team know that that the recipient of the survey had passed away or was incapacitated. The rest were respondents who wanted to let the study team know that they had completed their survey or respondents calling in with some form of comment or question. One caller wanted the study team to know that the survey was mailed to the wrong address. Three calls could not be resolved because they were either hang-ups or non-informative messages and study staff were not able to reach the respondents.

Table 3-3. Telephone calls received

Reason for call	Number of calls received
Refusal	7
Respondent let the study team know that the survey had been completed	5
Deceased/Sick	6
Non-locatable/Undeliverable	1
Respondent asked a question or made a comment. Topics included: • Whether participation was required • The recipient of the survey may not be able to complete the survey due to their age • They wanted to provide a new mailing address • They wanted to know how we obtained their mailing address	4
Calls that were never resolved due to hang ups or non-informative messages	3
Total	**26**

3.3 Incoming Questionnaires

Field room staff receipted all returned questionnaires into the SMS using each questionnaire's unique barcode. The SMS tracked each received questionnaire as well as the status of each sampled

participant. Once an individual was recorded as complete, they no longer received additional mailings. Packages that came back as undeliverable were marked as such in the SMS and those addresses did not receive further mailings.

In addition to refusing by calling the toll-free line, some respondents also refused by sending a letter stating that they did not wish to participate or asking to be removed from the mailing list.

These individuals were marked in the system as refusals and were removed from subsequent mailings. Respondents who sent back a blank questionnaire were also marked as refusals and removed from subsequent mailings.

The status of each HINTS-SEER cohort at the end of data collection (but before data cleaning and editing) can be found in Table 3-4.

Table 3-4. Status of HINTS-SEER cohorts at close of data collection

Respondent status	Iowa		Greater Bay Area		New Mexico	
	N	%	N	%	N	%
Complete	412	85.48	494	24.79	352	41.85
Missing Consent	N/A	N/A	102	5.12	N/A	N/A
Refusal	3	0.62	37	1.86	12	1.43
Deceased/Sick	2	0.41	14	0.70	0	0.00
Undeliverable	21	4.36	294	14.75	56	6.66
Nonresponse	44	9.13	1,052	52.78	421	50.06
Total	**482**	**100.0**	**1,993**	**100.0**	**841**	**100.0**

The number of questionnaires returned by date during the field periods for the Iowa, Bay Area, and New Mexico cohorts can be found in Tables 3-5, 3-6, and 3-7. The majority of returns for the Iowa cohort were early in the field period, with 85 percent of returns coming in after the first mailing of the survey and the mailing of the reminder postcard. The second mailing resulted in an additional 10 percent and the remaining five percent were in response to the final mailing.

Table 3-5. Iowa cohort response by date

Date of mailing	Period of returns	Number of returns
Mailing 1: January 11	January 12 – January 21	0
Postcard: January 19	January 22 – February 11	351
Mailing 2: February 10	February 12 – March 5	41
Mailing 3: March 3	March 6 – April 14	20
	Total	**412**

The majority of returns including consent for the Greater Bay Area cohort were early in the field period with 55 percent of returns including consent coming in after the first mailing of the survey and the mailing of the reminder postcard. The second and third mailings resulted in an additional 34 percent and the remaining 11 percent were in response to the fourth and final mailing.

Table 3-6. Greater Bay Area cohort response by date

Date of mailing	Period of returns	Returns including consent	Returns missing consent	Total number of returns
Mailing 1: February 16	February 1 – February 25	1	1	2
Postcard: February 23	February 26 – March 18	270	46	316
Mailing 2: March 17	March 19 – April 9	126	35	161
Mailing 3: April 7	April 10 – May 19	44	20	64
Mailing 4: July 9	July 10 – August 20	53*	0	53*
	Total	**494**	**102**	**596**

*Includes respondents who only returned a signed consent form in response to Mailing 4. The fourth mailing included only the consent form not the full instrument.

The majority of returns for the New Mexico cohort were early in the field period, with 74 percent of returns coming in after the first mailing of the survey and the mailing of the reminder postcard. The second mailing resulted in an additional 16 percent and the remaining 10 percent were in response to the final mailing.

Table 3-7. New Mexico cohort response by date

Date of mailing	Period of returns	Number of returns
Mailing 1: March 22	March 23 – March 31	0
Postcard: March 29	April 1 – April 22	260
Mailing 2: April 21	April 23 – May 14	57
Mailing 3: May 12	May 15 – June 23	35
	Total	**352**

4. Data Management

After being processed and receipted into the SMS, each returned paper questionnaire was scanned, and verified, cleaned, and edited. Imputation procedures were also conducted. These procedures are described below.

4.1 Scanning

All completed paper questionnaires were scanned using a data capture software (TeleForm) to capture the survey data and images were stored in SharePoint. Staff reviewed each form as it was prepared for scanning. The review included:

- Determining if the form was not scannable for any reason, such as being damaged in the mail. Some questionnaires or individual responses needed to be overwritten with a pen that was readable by the data capture software. Numeric response boxes were pre-edited to interpret and clarify non-numeric responses and responses written outside the capture area.
- Reviewing potential problem questionnaires or pertinent comments made by respondents.
- The reviewed paper surveys were then sent through the high-speed scanner to capture the responses. TeleForm read the form image files and extracted data according to HINTS-SEER rules established prior to the field period. Scanned data were then subject to validation according to HINTS specifications. If a data value violated validation rules (such as marking more than one choice box in a mark-only-one question) the data item was flagged for review by verifiers who looked at the images and the corresponding extracted data and resolved any discrepancies.

Decisions made about data issues as a result of scanning were recorded in a data decision log. The decision log contains the respondent ID, the value triggering the edit, the updated value, and the reason for the update. A total of 21 entries were made into the data decision log during the course of data scanning and processing. These were attributed to decisions made about numeric entries outside variable parameters (i.e., 2-digit numbers written on single digit question).

A 10 percent quality control check was then conducted on the scanned data and the electronic images of the survey. Quality Assurance (QA) staff compared the hard copy questionnaire to the data captured in the database item-for-item and the images stored in the repository page-for-page to

ensure that all items were correctly captured. If needed, updates were made. In addition, QA staff closely reviewed frequencies and cross tabulations of the HINTS-SEER raw data to identify outliers and open ended items to be verified. ID reconciliation across the database, images, and the SMS, was completed to confirm data integrity.

4.2 Data Cleaning and Editing

Once the paper questionnaires had been scanned, all survey data were cleaned and edited. General cleaning and editing activities are described briefly below, with more detailed information found in **Appendix G** (Variable Values and Data Editing Procedures).

- Customized range and logical inconsistency edits, following predetermined processing rules to ensure data integrity, were developed and applied against the data.
- Edit rules were created to identify and recode nonresponse or indeterminate responses.
- Missing values were recoded for some responses to questions that featured a forced-choice response format and for filter questions where responses to later questions suggested a particular response was appropriate.
- Derived variables were created to reflect each response recorded for certain "mark-one" questions in order to facilitate the imputation process implemented when respondents did not follow the instruction to mark only one response. For these variables (listed below), imputation, as described in Section 4.3, was carried out. For other "mark-one" questions where respondents marked multiple responses, editing rules were used to determine which response was retained.

Table 4-1. Derived variables for imputation

Item number	Variable name
A5	SEERStrongNeedCancerInfo_IMP
G5	WhoOrderedCaTest_IMP
H3	FirstInfoClinTrials2_IMP
H4	TrustInfoClinTrials2_IMP
J22	MostImportantValues_IMP
P11	SexualOrientation

- Categorical variables were created to summarize the responses for the "mark all that apply" questions in the instrument. These variables (listed below) indicated each response selected for respondents selecting only one response, and a code was created to indicate "multiple categories selected" for all of the respondents who answered multiple responses.

Table 4-2. Categorical variables from select all questions

Item number	Variable name
B6	HaveDevice_Cat
E1	SEER_Cancer_Cat
F1	COVIDCa_Cat
F2	COVIDRoutine_Cat
G1	HeardGenTest_Cat
G2	TestSource_Cat
G3	HadTest2_Cat
G4	CaTest_Cat
G6	UndGenTest2_Cat
G7	SharedRes3_Cat
P5	Occupation_Cat
P8	Hisp_Cat
P9	Race_Cat2

- Data cleaning was carried out for the two height variables: Height_Feet and Height_Inches. The rules (detailed in Appendix G) that were applied minimized the number of out-of-range values by accounting for response measurements in incorrect boxes, responses using metric measures, responses using only one unit of measurement and other response errors.

- "Other, specify" responses were examined, cleaned for spelling errors, categorized, and upcoded into preexisting response codes when applicable.

4.3 Imputation

In the HINTS-SEER data, there are questions for which respondents incorrectly selected more than one response and therefore were recoded to -5 (respondent selected more response options than appropriate for the question) and subject to imputation. A single answer was imputed by selecting one response among those selected by the respondent. The imputed response was based on the distribution of answers among the single-answer responses on each question. If a respondent selected two responses, for example, where the first response comprised 40 percent of the single-answer responses and the second response comprised 10 percent, the first response was likely to be the imputed response 4 out of 5 times (40% / (40% + 10%)), and the second response was likely to be the imputed response 1 out of 5 times (10% / (40% + 10%)). The items imputed and the number of imputations conducted are shown in Table 4-3 below. An imputation flag is included on the dataset to indicate imputed values.

Table 4-3. Items with imputation

Item number	Variable name	Number of imputations
A5	SEERStrongNeedCancerInfo_IMP	75
G5	WhoOrderedCaTest_IMP	12
H3	FirstInfoClinTrials2_IMP	67
H4	TrustInfoClinTrials2_IMP	33
J22	MostImportantValues_IMP	13
	Total	200

4.4 Survey Eligibility

Returned surveys were reviewed for completion and duplication (more than one questionnaire returned from the same individual) to ensure they were eligible for inclusion in the final dataset. Of the 1,305 questionnaires received, 20 were returned blank, 5 were determined to be incompletely filled out, and 27 surveys were identified as duplicates (i.e., the same individual returned multiple surveys).

Nineteen individuals returned a survey and reported that they were never diagnosed as having cancer. These cases were brought to the registries' attention and registry staff determined that although these individuals were diagnosed with cancer, they were often a less aggressive or early stage of cancer and the patient may not have been aware of the diagnosis. The registries also reported that certain ethnic or cultural groups tend to try keep this kind of diagnosis from patients so these individuals perhaps were told by family that they did not have cancer. These 19 responses that stated they did not have cancer were treated as ineligible. The remaining 1,234 surveys were determined to be eligible. The processes for these reviews are detailed below.

Definition of a Complete and Partial Complete Questionnaire

Consistent with prior HINTS administrations, a complete questionnaire was defined as any questionnaire with at least 80 percent of the required questions answered in Sections A and B. For HINTS-SEER, only questions required of every respondent were factored into the completion rate calculation. Questions that followed skip patterns were excluded from the analysis. A partial-complete was defined as when between 50 percent and 79 percent of the questions were answered in Sections A and B. There were 45 partially-completed questionnaires. Both partially-completed and completely-answered questionnaires were retained. Five questionnaires with fewer than 50 percent of the required questions answered in Sections A and B were coded as incompletely filled out and

discarded. The 5 incomplete questionnaires represented 0.4 percent of all eligible surveys, which was consistent with all prior cycles of HINTS 5. Data for the 3 registries breaks down as follows:

Table 4-4. Summary of questionnaire completeness by SEER registry

Registry	Partial complete	Complete	Incomplete*	Respondent reported no cancer diagnosis*	Total questionnaires retained
Iowa	15	393	2	2	408
Greater Bay Area	22	459	2	11	481
New Mexico	8	337	1	6	345
Total	45	1,189	5	19	1,234

*Ineligible and therefore discarded

4.5 Additional Analytic Variables

Included in the datasets are four sets of analytical variables: (1) National Center for Health Statistics (NCHS) urban-rural classification scheme for counties; (2) 2013 Urban Influence Codes; (3) USDA Rural-Urban Commuting Area (RUCA) codes that classify census tracts using measures of population density, urbanization, and daily commuting; and 4) USDA Rural-Urban Continuum Codes (RUCC).

The **NCHS Urban–Rural Classification Scheme for Counties (NCHSURCODE2013)** was developed in 2013 for use in studying associations between urbanization level of residence and health and for monitoring the health of urban and rural residents. The scheme groups counties and county-equivalent entities into six urbanization levels (four metropolitan and two nonmetropolitan), on a continuum ranging from most urban to most rural.

The **2013 Urban Influence Codes (UIC2013)**, developed by the United States Department of Agriculture, form a classification scheme that distinguishes metropolitan counties by population size of their metro area, and nonmetropolitan counties by size of the largest city or town and proximity to metro and micropolitan areas. The standard Office of Management and Budget (OMB) metro and non-metro categories have been subdivided into two metro and 10 non-metro categories, resulting in a 12-part county classification.

The **two RUCA codes (primary and secondary)** provide a detailed and flexible way for delineating sub-county components of rural and urban areas. They are based on the 2006-10 American Community Survey (ACS) and have been updated using data from the 2010 decennial

census. The primary codes (PR_RUCA2010) delineate metropolitan and nonmetropolitan areas based on the size and direction of primary commuting flows. The secondary codes (SEC_RUCA2010) further subdivide the primary codes to identify areas where classifications overlap based on the size and direction of the secondary, or second largest, commuting flow.

The **2013 Rural-Urban Continuum Codes (RUC2013)** form a classification scheme that distinguishes metropolitan counties by the population size of their metro area, and nonmetropolitan counties by degree of urbanization and adjacency to a metro area. The Office of Management and Budget (OMB) metro and non-metro categories have been subdivided into three metro and six non-metro categories. Each county in the country is assigned one of the nine codes.

4.6 SEER Registry Variables

As noted in Chapter 2, selected tumor and diagnosis data were provided by each of the registries for each sampled participant. These variables (year of diagnosis, cancer site, histology, and SEER summary stage) are detailed below. The cancer sequence variable (CSEQ) that was used as part of the registries' sampling procedures is not provided on the HINTS-SEER dataset. As noted in Chapter 2, for people with multiple eligible tumors, the Iowa and New Mexico registries selected tumors systematically, selecting the first eligible tumor while the Greater Bay Area Cancer Registry used random selection of eligible tumors. A descriptive table of the cancer sequence numbers for respondents in the final HINTS sample are listed below in Table 4-5.

Table 4-5. Unweighted Frequencies of Cancer Sequence Number for HINTS Respondents

Cancer Sequence Number	Iowa		New Mexico		Greater Bay Area		Total Sample	
	N	%	N	%	N	%	N	%
00	327	80.15	295	85.51	378	78.59	1,000	81.04
01	75	18.38	38	11.01	56	11.64	169	13.70
02	5	1.23	11	3.19	39	8.11	55	4.46
03	1	.25	1	.30	8	1.66	10	.81
Total	408	100.00	345	100.00	481	100.00	1,234	100.00

The below variables are included in the HINTS-SEER dataset for each respondent to the survey. They were renamed with the prefix "Registry_" to differentiate them from self-reported HINTS survey variables with similar names. A description of these variables and any modifications made are described below.

Year of Cancer Diagnosis

Registry_Year_Of_Diagnosis indicates the year of diagnosis of the cancer used by the registry to include the respondent in the sample based on the sampling strata outlined in section 2.3 . If the respondent had been diagnosed with more than one cancer in their lifetime, diagnosis years for cancers other than the one used for sampling are unknown. In other words, the cancer used for sampling may not have been the respondent's first cancer. For consistency and to help with respondent confidentiality, each of the three SEER datasets was restricted to year of diagnosis rather than the full date.

Cancer Site

Registry_Cancer_Site is a raw, uncategorized SEER variable which indicates the anatomical location of the cancer that was used during sampling. In addition to the individual cancer site codes, the dataset includes some recodes that combine the individual site codes and histology codes in order to facilitate analysis. These are:

- Standard re-codes (Registry_Cancer_Site_StdRecode, see **Appendix J**) as provided by the ICD-O-3 SEER standard recode chart[1];
- Standard re-codes that were reviewed by certified tumor registrars and other experts and slightly edited (Registry_Cancer_Site_Group, see **Appendix H**); and
- A composite variable (Registry_Cancer_Site_OrganSys, see **Appendix I**) that combines the site codes into just 22 categories.

Histology

Registry_Histology is a raw, uncategorized SEER variable which provides a code for the type of tissue from which the cancer originated.

[1] Site Recode ICD-0-3/WHO 2008 Definition, published by the National Cancer Institute, Surveillance, Epidemiology, and End Results Program, available on seer.cancer.gov (https://seer.cancer.gov/siterecode/icdo3_dwhoheme/index.html).

From a histological standpoint there are hundreds of different cancers, so to facilitate analysis, Westat grouped the Registry_Histology variable into nine major categories and had those categories reviewed by certified tumor registrars and other cancer registry experts. The Registry_Histology_Recode categories are outlined below.

Table 4-6. Histology Recode Values

Data value	Value Label for Registry_Histology_Recode	Registry_Histology Values
1	Carcinoma	8010, 8013, 8046, 8050-8052, 8070-8072, 8074, 8076, 8120, 8122, 8130-8131, 8140, 8145, 8160, 8170, 8200, 8210-8211, 8230, 8240, 8244, 8246, 8249, 8252, 8255, 8260-8261, 8263, 8310, 8312, 8323, 8330-8331, 8335, 8340-8341, 8345, 8380, 8401, 8441, 8460-8461, 8470-8471, 8480-8482, 8490, 8500-8501, 8503, 8507, 8510, 8520, 8522-8524, 8530, 8542, 8550, 8560, 8570, 8575, 8585, 8950, 8980
2	Mesenchymal Tumor/Sarcoma	8801, 8811, 8830, 8858, 8890, 8936, 9020, 9140, 9251
3	Neural/Glial Neoplasm	9400, 9450
4	Germ Cell Tumor	9061, 9065, 9070, 9080-9081, 9085
5	Lymphoma/Lymphocytic Leukemia	9590-9591, 9650, 9652, 9663, 9670, 9675, 9679-9680, 9689-9691, 9695, 9698-9699, 9761, 9823, 9833, 9835, 9837, 9940
6	Plasma Cell Neoplasm	9732, 9734
7	Myelogenous Leukemia	9861, 9863, 9872-9873, 9875
8	Myelodysplastic Syndrome/Other Myeloproliferative Neoplasm	9950, 9961, 9975, 9982-9983, 9985, 9989
9	Melanocytic Tumor	8720-8722, 8730, 8742-8745, 8761, 8770

SEER Summary Stage

Registry_Summary_Stage categorizes how far a cancer has spread from its point of origin. Summary Stage uses all information available in the medical record. In other words, it is a combination of the most precise clinical and pathological documentation of the extent of disease.[2] The range of stages provided by the registries and available in the dataset are outlined in Table 4-7 below.

[2] Summary Stage 2018, released September 9, 2021 (Version 2.1), published by the National Cancer Institute, Surveillance, Epidemiology, and End Results Program, available on seer.cancer.gov (https://seer.cancer.gov/tools/ssm/).

Table 4-7. SEER Summary Stage Variable Values

Data value	Value label for Registry_Summary_Stage
1	Localized only
2	Regional by direct extension only
3	Regional lymph nodes only
4	Regional by BOTH direct extension AND lymph node involvement
5	Regional, NOS
7	Distant site(s)/node(s) involved
9	Unknown if extension or metastasis (unstaged, unknown, or unspecified) Death Certificate only case

4.7 Codebook Development

Following cleaning, editing, and weighting (described below), a detailed codebook including frequencies was created for both the weighted and the unweighted data for the 3 combined HINTS-SEER registries. The codebooks define all variables in the questionnaires, provide the question text, list the allowable codes, and explain the inclusion criteria for each item. The survey instrument was annotated with variable names and allowable codes to support the usability of the delivery data.

5. Weighting and Variance Estimation

Weighting was conducted for all HINTS-SEER participants using control totals of the eligible population within each of the three registries. Because weighting for HINTS-SEER was conducted separately for each registry, the weights reflect the eligible registry population of each specific registry rather than the overall population of the state. Comparing HINTS-SEER estimates to other HINTS data collections should be done with caution since other HINTS data collections are weighted to represent the full US population. See Chapter 7 for more information about comparing HINTS-SEER to HINTS.

Every sampled cancer survivor who completed a questionnaire for HINTS-SEER from a SEER registry received a full-sample weight and a set of 50 replicate weights. The full-sample weight is used to calculate population and subpopulation estimates. Replicate weights are used to compute standard errors for these estimates. The use of sampling weights is done to ensure valid inferences from the responding sample to their respective population, correcting for nonresponse and noncoverage biases to the extent possible. Population in this context is defined as all cancer survivors from each registry that met the eligibility criteria defined in Section 2.2.

The computation of the full-sample weights consisted of the following steps:

- Calculating base weights;
- Adjusting for nonresponse; and
- Calibrating cancer survivor weights to counts of eligible cancer survivors from the corresponding registry (referred to below as control totals).

Replicate weights were calculated using the 'delete one' jackknife (JK1) replication method.

5.1 Base Weights

The initial step in the weighting process is calculating the base weight for each cancer survivor in the sample. The base weight is the reciprocal of the probability of selecting the survivor for the survey from the list of eligible survivors from their respective SEER registry. Since the sample was selected

using a single-stage equal probability sample design, every sampled survivor from a registry had the same base weight. Table 5-1 shows the base weight for each SEER registry.

Table 5-1. Eligible survivors, sample size, and survivor base weights by SEER registry

SEER registry	Eligible survivors	Sample size	Survivor base weights
Iowa	127,881	482	265.3133
New Mexico	60,597	850	71.2906
Greater Bay Area	239,221	2,000	119.6105

5.2 Nonresponse Adjustments

Nonresponse is generally encountered, to some degree, in every survey. The first and most obvious effect of nonresponse is the reduction in the effective sample size, which in turn increases the sampling variance. In addition, if there are systematic differences between the respondents and the nonrespondents, there will be a bias of unknown size and direction. This bias is generally adjusted for in the case of unit nonrespondents (nonrespondents who refuse to participate in the survey at all) with the use of a weighting adjustment term multiplied to the base weights of sample respondents. Item nonresponse (nonresponse to specific questions only) is generally adjusted for through the use of imputation. This section discusses weighting adjustments for unit nonresponse.

The most widely accepted paradigm for unit nonresponse weighting adjustment is the quasi-randomization approach (Oh and Scheuren, 1983). In this approach, nonresponse cells are defined based on those measured characteristics of the sample members that are known to be related to response propensity. For example, if it is known that males respond at a lower rate than females, then sex should be one characteristic used in generating nonresponse cells. Under this approach, sample units are assigned to a response cell based on a set of defined characteristics. The weighting adjustment for the sample unit is the reciprocal of the estimated response rate for the cell. Any set of response cells must be based on characteristics that are known for all sample units, responding and nonresponding. Thus, questionnaire items on the survey cannot be used in the development of response cells because these characteristics are only known for the responding sample units.

Under the quasi-randomization paradigm, Westat models nonresponse as a "sample" from the population of adults in that cell. If this model is in fact valid, then the use of the quasi-randomization weighting adjustment eliminates any nonresponse bias (see, for example, Chapter 4 of Little and Rubin, 1987). The weighting procedure for HINTS-SEER used a survivor-level

nonresponse adjustment procedure based on this approach. The base weights of the survivors that did return the questionnaire were adjusted to reflect nonresponse by the remaining eligible survivors. The software package called SI-CHAID[3] was used to identify variables highly correlated with survivor-level response, and these variables were used to create the nonresponse adjustment cells. These cells were formed separately by SEER registry sample, and the variables used to define nonresponse cells were:

- Race/ethnicity
- Year of birth (categorized by decade)
- Metropolitan Status (county in Metro areas; county in Non-Metro areas)
- Year of diagnosis (categorized by decade)

Nonresponse adjustment factors were computed for each nonresponse cell *b* using the formula below. This formula is consistent with the RR4 formula of the American Association of Public Opinion Research (AAPOR) for calculating response rates. This is the same formula that was used to compute nonresponse adjustment factors for HINTS 5, Cycle 4.

$$HH\_NRAF(b) = \frac{RESPONSE + NONRESPONSE + UNKNOWN \times e}{RESPONSE},$$

where

- *RESPONSE* is the sum of survivor base weights for all responding survivors in nonresponse cell *b*,
- *NONRESPONSE* is the sum of the survivor base weights for all known nonresponding survivors in nonresponse cell *b*,
- *UNKNOWN* is the sum of the survivor base weights for all survivors who were not mailed a survey because of an undeliverable mailing address and whose eligibility is unknown in nonresponse cell *b*, and
- *e* is the estimated percentage of eligible cancer survivors among the cancer survivors with known response status for a specific SEER registry sample.

[3] SI-CHAID 4.0 User's Guide by J. Magidson, published by Statistical Innovations Inc., available on StatisticalInnovations.com (https://www.statisticalinnovations.com/wp-content/uploads/SICHAIDusersguide.pdf).

Table 5-2 summarizes the nonresponse adjustments for each SEER registry sample. It also includes the percentage of eligible survivors among the survivors with known response status (e).

Table 5-2. Nonresponse adjustments summary by SEER registry

SEER registry	Percentage of eligible among known response status (e)	Average nonresponse factor	Smallest nonresponse factor	Largest nonresponse factor
Iowa	99.6	1.12	1.05	1.20
New Mexico	99.2	2.27	1.78	2.94
Greater Bay Area	98.7	1.18	1.15	1.28

5.3 Calibration Adjustments

In this step, sampling weights after nonresponse adjustments were calibrated to population counts of eligible cancer survivors from each HINTS-SEER registry. The purpose of calibration is to reduce the sampling variance of estimators using reliable auxiliary information (see, for example, Deville and Sarndal, 1992) or information obtained directly from the sampling frame. In the ideal case, this auxiliary or frame information usually takes the form of known population totals for particular characteristics (called *control totals*). However, calibration also reduces the sampling variance of estimators if the auxiliary information has sampling errors, if these sampling errors are significantly smaller than those of the survey itself.

Calibration reduces sampling errors particularly for estimators of characteristics that are highly correlated to the calibration variables in the population. The extreme case of this would be the calibration variables themselves. The survey estimates of the control totals would have considerably higher sampling errors than the "calibrated" estimates of the control totals, which would be the control totals themselves. The estimator of any characteristic that is correlated to any calibration variable will share partially in this reduction of sampling variance, though not fully. Only estimators of characteristics that are completely uncorrelated to the calibration variables will show no improvement in sampling error. Deville and Sarndal (1992) provide a rigorous discussion of these results.

Control Totals

For each of the three HINTS-SEER samples, the control totals reflecting the distributions of demographic characteristics of the eligible population of survivors were provided by the administrators of the individual SEER registries. The controls totals were based on cancer survivors that met the eligibility requirements for the study and came directly from each registry's corresponding sampling frame. For the HINTS-SEER survey, all the registries provided estimates of age, race/ethnicity, and sex. Iowa additionally provided year of diagnosis, cancer stage (localized, regional, distant, unstaged), and cancer site. Table 5-3 summarizes the characteristics used as control totals for the separate SEER registry samples.

Table 5-3. Characteristic used as control totals by SEER registry sample

SEER registry	Characteristics used for raking						
		Age	Race	Ethnicity	Year of diagnosis	Cancer stage	Cancer site
Iowa		✓			✓	✓	✓
New Mexico		✓		✓			
Greater Bay Area		✓	✓				

In some instances, specific characteristics were not used for raking because the number of sample cases for that characteristic was too small. Specifically, the responding sample for Iowa had too few non-White survivors and Hispanic survivors to effectively use in the raking process. As a consequence, Westat used year of diagnosis, cancer stage, and cancer site, instead.

Raking to the control totals for these variables listed in Table 5-3 was then performed. As a result of raking HINTS-SEER weights to the control totals, the weights for each of the sites sum to the total eligible population for each of the sites and reflect the above control totals for each site. For instance, the sum of the final HINTS-SEER weights by age will match up to the distribution of age of the population of eligible SEER cancer patients provided by each SEER registry.

5.4 Replicate Variance Estimation

In addition to the full-sample weight, a set of 50 replicate weights were provided for each respondent in each HINTS-SEER sample. These replicate weights are used to calculate standard error of estimates obtained from the HINTS-SEER data, using the delete one jackknife (JK1) replication method. Replicate weights were calculated for each registry sample separately.

The JK1 jackknife technique is compatible with the sample design and weighting procedures for HINTS. This jackknife variance estimation technique takes carefully selected subsets of the data for each "replicate," and for each respondent in the replicate subset and determines a sampling weight, as if the replicate subset were in fact the responding sample. (This replicate subset is usually almost the entire sample, except for a group of respondents that are "deleted" for that replicate.) The resulting weights are called replicate weights.

The jackknife variance estimator requires the use of replicate weights. A set of 50 replicate weights was assigned to each responding cancer survivor. To illustrate how the replicate variance estimates are computed, suppose P is a percentage of survivors in a SEER registry having a particular characteristic (e.g., answering one of the HINTS questions in a particular way). A representative estimator p can be computed by aggregating the sampling weights of all responding survivors with this characteristic (e.g., all responding survivors in the survey answering the survey question in a particular way). A JK1 jackknife variance estimator of the sampling variance of p can be computed in two steps:

Step 1. Recompute estimators $p(r)$, $r = 1,...,50$, by aggregating the replicate sampling weights corresponding to replicate r for all responding cancer survivors with the characteristic.

Step 2. Compute the JK1 jackknife variance estimator

$$v(p) = \frac{R-1}{R}\sum_{r=1}^{50}(p(r)-p)^2$$

The replicate weights are computed by systematically deleting a portion of the original sample, and recomputing the sampling weights as if the remaining sample (without the deleted portion) were the actual sample. The remainder of the sample with the deleted portion removed is called the replicate subset, and it should mirror the full sample design, as if it were a reduced version of the original sample.

For the purposes of JK1 jackknife variance estimation, each survivor was assigned to one of 50 replicate "deletion" groups $D(r)$, $r = 1,..., 50$. Each replicate sample is the full sample minus the deletion group (i.e., it is roughly 49/50 of the original sample).

The replicate sampling weights were generated in a series of steps that parallel the steps computing the full-sample sampling weights. The replicate base weight for each sampled survivor and each

replicate is either equal to *R/(R-1)* times the full sample base weight (if the survivor is contained in the replicate subset) or equal to 0 (if the survivor is not contained in the replicate subset, but instead is contained in the "deleted" set for that replicate).

Nonresponse and calibration adjustments were then computed for each set of replicate weights, using the replicate weights in the computation of nonresponse and calibration adjustments in place of the original weights. These calculations generated a set of replicate nonresponse and post-stratification adjustments for each responding survivor. The final replicate weights were products of the replicate weights, nonresponse adjustments, and calibration adjustments.

5.5 Taylor Series Variance Estimation

Even though replication is the recommended method for variance estimation for HINTS, not all software packages have a replication option to produce variance estimates. For example, SPSS has built-in options for estimating variance using Taylor Series methods but not replication methods. To accommodate SPSS users or any end user who wants to produce variances using Taylor Series methods, Westat provided the appropriate variables on the HINTS data files to do so as described below.

The full-sample weight (as described in the introduction of Chapter 5) is used as the weight to compute Taylor Series variance estimates. The variable VarStratum indicates the variance-estimation stratum, which codes for the three cancer registries, and the variable VarCluster indicates the primary sampling unit (PSU) or cluster within the variance-estimation stratum. These variables allow the analyst to produce variance estimates using Taylor Series.

6. Response Rates

6.1 Response rates by SEER site

For HINTS-SEER, response rates were calculated differently than typical HINTS data collection cycles. First, the HINTS-SEER response rates are not weighted to correct for differential selection probabilities because there was no oversampling in HINTS-SEER[4]. Second, because the participating SEER registries required active or passive consent from potential respondents to be included in HINTS-SEER, the response rate is calculated in two stages.

1. The first stage accounts for the consent rates and is calculated as the proportion of sampled registrants who consented to participate. Stage 1 consent rates are reported in Table 6-1.

2. The second stage accounts for the survey completion rate and is computed as the proportion of those who consented who returned a complete survey. Stage 2 completion rates are reported in Table 6-2.

The overall response rate is the product of the consent and completion rates (Stage 1 * Stage 2) and is presented by registry in Table 6-2.

Table 6-1. Consent rates by HINTS-SEER registry and overall

SEER registry (consenting process)	A – Total sampled	B – Total consented	Consent rate (Stage 1) (B/A)
Iowa (active consent prior to receiving survey)	6,433	482	7.5%
Greater Bay Area (active consent with survey mailing)	1,993	483	24.2%
New Mexico (passive consent)	1,400	841	60.1%
Total	**9,826**	**1,806**	**18.4%**

Table 6-2. Survey completion rates and final response rates by HINTS-SEER registry and overall

SEER registry	C – Total completed surveys	Stage 2 completion rate (C/B)	Overall response rate (Stage 1 * Stage 2)
Iowa	408	84.6%	6.3%
Greater Bay Area	481	99.6%	24.1%
New Mexico	345	40.9%	24.6%
Total	**1,234**	**68.3%**	**12.6%**

[4] While each population of registrants was stratified by two factors (years since diagnosis and race/ethnicity), all registrants were systematically sampled with the same selection probability. For samples with equal selection probability, regardless if they are stratified or not, each sample unit will have the same base weight. Thus, response rates calculated using weights will be the same as response rates calculated without weights.

The New Mexico Tumor Registry, which required passive consent, achieved the highest consent rate and lowest completion rate. The Iowa Cancer Registry, which required active consent prior to receiving a survey, achieved the lowest consent rate and a relatively high survey completion rate.

The Greater Bay Area Cancer Registry, which required active consent but included the consent form with the survey mailing, achieved a consent rate in between New Mexico and Iowa. The Greater Bay Area Cancer Registry achieved the highest completion rate because providing consent required returning a completed survey with a signed consent form.

The Greater Bay Area Cancer Registry and New Mexico Tumor Registry achieved similar overall response rates of 24.1 and 24.6 percent, respectively. Iowa achieved a substantially lower response rate (6.3%) which was attributable to the low consent rate from the very large sample that was asked to consent to have their addresses made available as part of the study. Active consent procedures are known to yield lower second-stage response rates relative to passive consent procedures in research studies (Range et al., 2001).

6.2 Nonresponse bias analysis

In this section we compare the demographic composition of the HINTS-SEER respondents to the pool of sampled registrants in each registry. Each registry provided aggregated demographic data for the pool of registrants to be included in HINTS-SEER. In Table 6-3 we compare these distributions to those of the final respondents based on their survey responses. Overall the respondents were not substantially different from the overall sample. For most of the demographic comparisons, the difference in proportions was five percentage points or less. In Iowa, there were a larger proportion of individuals age 60 or older in the respondents than in the overall sample and there was a smaller proportion who were diagnosed with cancer prior to 2010. In the Greater Bay Area, the respondents included a larger proportion of individuals over the age of 60 and Non-Hispanic Whites than the overall sample. In New Mexico there was a larger proportion of Hispanics among the respondents than in the overall sample.

Table 6-3. Comparison of demographic distributions between sampling frame and survey respondents by HINTS-SEER site (unweighted)

Iowa	Sampled (6,433)	Surveyed (408)
Age >= 60	78%	84%
Male	45%	46%
White, Non-Hispanic	96%	97%
Hispanic	1%	2%
Diagnosis prior to 2010	54%	47%
Greater Bay Area	**Sampled (1,993)**	**Surveyed (481)**
Age > 60	82%	88%
Male	48%	47%
White, Non-Hispanic	65%	74%
Hispanic	10%	7%
Diagnosis prior to 2010	53%	54%
New Mexico	**Sampled (1,400)**	**Surveyed (345)**
Age > 62	78%	79%
Male	45%	46%
White, Non-Hispanic	65%	65%
Hispanic	26%	31%
Diagnosis > 9 years ago	57%	56%

7. Analyzing HINTS-SEER Data with Other HINTS Cycles

The primary goal of sampling from cancer registries was to administer surveys to a larger sample of cancer survivors than would respond in a probability-based population survey such as HINTS. As discussed in Chapter 1, each cycle of HINTS randomly includes 500 to 600 survivors. This is adequate for some purposes, but it is difficult to do detailed analyses by important subgroups or types of cancers. For this pilot study, the sample of approximately 1,200 survivors doubles the numbers available on a traditional cycle of HINTS.[5] There are several different methods that analysts might consider when using the HINTS-SEER data. The sections below discuss analyzing the HINTS-SEER data as a single data-set, comparing the HINTS-SEER estimates to HINTS, and combining the HINTS-SEER and HINTS into a single estimate. For reasons discussed in the last section below, combining the data is not recommended at this point. More information about analyzing the data can be found in the Overview Of The HINTS-SEER (2021) Survey And Data Analysis Recommendations document.

Separate Analysis by Registry or Combining Across the 3 HINTS-SEER Samples

The most straightforward analysis is to generate estimates that represent each of the individual SEER registries included in the study. As noted above (see section on 'control totals' in Section 5.3), the weights for each registry scale up the respondents who participated in the survey to the registry populations from which they were drawn. Accounting for the respondent's probability of selection into the study, the nonresponse and calibration adjustments account for differences between the sample selected and the sample frame (i.e., the registry list from which the sample was drawn). The frame represents the members of the registry who met the eligibility requirements as described in Chapter 2. The HINTS-SEER dataset includes all three registries in a single file. This gives the analyst the flexibility to conduct analyses that combine across all three registries, or analyze the registries separately. The three registries are also identified by the SEERREGISTRY_FLAG

[5] With the new biennial cycle, the numbers for HINTS should double with respect to the number who are cancer survivors.

variable. The weights on the file were developed separately for each registry and are combined into one file. This means that when subsetting the data by registry, the weights can be used to generate estimates for that registry using the standard replication procedures or when using Taylor Series linearization. Alternatively, the weights can also be used if analyses combine across all registries into a single set of estimates. As noted in the weighting chapter, there are 50 replicates that will be used for estimating standard errors and two variables, VarStratum and VarCluster for Taylor Series . No special procedures, such as those needed when combining across HINTS data collection years,[6] are needed when combining across registries.

An analysis that ignores the registries will be representative of the three registries that participated in the study. Some caution should be taken when combining across the registries given the different procedures used to gain consent in each place. One concern is the very low response rate in the Iowa registry relative to the other two (see table 6.2). While all of the data were weighted to account for nonresponse, this adjustment may have limited value for this particular registry sample. Before conducting analyses across all sites, analysts should test whether there are differences in the outcome of interest between registries, with special attention to Iowa. Testing for differences can be completed using simple bivariate tests (e.g., t-tests) that compare the outcomes across the sites, essentially treating site as a covariate. As noted above, the weights can be used without any special adjustments.

If differences are found between the sites, they might represent differences in the methods used to recruit respondents. As noted above, if Iowa stands out, this is evidence that the low response rate might have affected results and it may not be appropriate to combine with the other sites. Or if possible, the analysis could include site as a covariate when analyzing the file with two or three registries. If there are differences across the other sites, then the investigator should examine whether other covariates may account for the differences. For example, if the registries differ by type of cancer and this is correlated with the outcome of interest, then including the type of cancer as a covariate would control for differences between sites.

[6] Rizzo, L., Moser, R.P., Waldron, W., Wang, Z. and W.W. Davis (undated) Analytic Methods to Examine Changes Across Years Using HINTS 2003 & 2005 Data.
Analytic Methods to Examine Changes Across Years Using HINTS 2003 & 2005 Data (cancer.gov)

Comparing HINTS-SEER with HINTS

While it is not nationally representative, the HINTS-SEER samples **are** probability samples from the frames from which they were drawn. It is possible to compare and contrast the HINTS-SEER results with cancer survivors that are captured in HINTS. Comparing the demographic distributions, the types of cancers, and their responses to the health and communication items on the survey can provide users of the HINTS-SEER data a way to assess how the two datasets differ along key outcomes of interest to analysts. For example, this might involve comparing the percent of survivors who are satisfied with the care they are getting from their primary care physician for the HINTS-SEER sample to the same sample of cancer survivors from the national HINTS. One can statistically compare HINTS and HINTS-SEER (i.e., conduct significance tests, run models) using the same methods analysts currently use when comparing data across HINTS cycles (Rizzo, et al., n.d.).[7]

Combining HINTS-SEER with HINTS

Combining HINTS-SEER with HINTS to generate a single estimate is potentially a powerful method to increase the precision of the estimates. When considering combining the HINTS-SEER and HINTS data, it is important to consider that the two samples (HINTS-SEER and HINTS) differ in important ways. First, the sample frame for the HINTS-SEER datasets is restricted to the three geographic regions (Iowa, New Mexico and the Greater Bay Area of California). It is useful that the registries were selected to represent different Census regions of the country (West and Midwest).[8] Nonetheless, the HINTS-SEER sites are restricted to relatively small areas within those regions. The HINTS data are sampled to cover the entire country and are therefore representative of the U.S. adult population. Second, the sample frames (registries for HINTS-SEER vs. housing units for HINTS) and methods of recruitment are very different. And third, the types of cancer survivors found in HINTS-SEER and HINTS are also likely to be different. The HINTS data contain anyone who was diagnosed with cancer up to the time the survey was filled out, and therefore may contain recently diagnosed survivors, while the HINTS-SEER sample was limited to individuals who had been diagnosed two or more years prior to sample selection. The HINTS-SEER sample also

[7] For an example, see https://hints.cancer.gov/docs/HINTS_IDA_Report.pdf.

[8] The site intended to represent the southeast (Georgia) had to drop out because of the COVID 19 pandemic.

excluded those diagnosed with non-melanoma skin cancer. These two definitional differences can be accounted for in analysis by identifying these groups on the HINTS datasets and either excluding them from any analysis or comparing the results for these groups separately.

For these reasons, combining the datasets to generate estimates is not recommended. Further exploration is needed of how the two types of samples differ geographically, demographically, and from a health perspective, as was suggested in the previous section ("Comparing HINTS-SEER and HINTS").

References

Deville, J.C., and Sarndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association, 87*, 376-382.

Dillman, D.A., Smyth, J.D., and Christian, L.M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method.* Hoboken, NJ: John Wiley and Sons.

Finney Rutten, L.J., Blake, K.D., Skolnick, V.G., Davis, T., Moser, R.P., and Hesse, B.W. (2020). Data Resource Profile: The National Cancer Institute's Health Information National Trends Survey (HINTS). *International journal of epidemiology*, *49*(1), 17–17j.

Little, R., and Rubin, D.B. (1987). *Statistical analysis with missing data.* New York: John Wiley and Sons.

Oh, H., and Scheuren, F. (1983). Weighting adjustments for unit response. In W.G. Madow, I. Olkin, and D. B. Rubin (Eds.), *Incomplete data in sampling surveys, Vol. II: Theory and annotated bibliography*. New York: Academic Press.

Range, L., Embry, T., and MacLeod, T. (2001). Active and passive consent: a comparison of actual research with children. *Ethical human sciences and services.* 3 (1), 23-31.

Rizzo, L., Moser, R.P., Waldron, W., Wang, Z., and Davis, W.W. (n.d.) *Analytic Methods to Examine Changes Across Years Using HINTS 2003 and 2005 Data.* Retrieved from https://hints.cancer.gov/publications-reports/users_data_handbook_accessible.aspx

Appendix A

HINTS-SEER Instrument

Appendix A
HINTS-SEER Instrument

Health Information

National Trends Survey

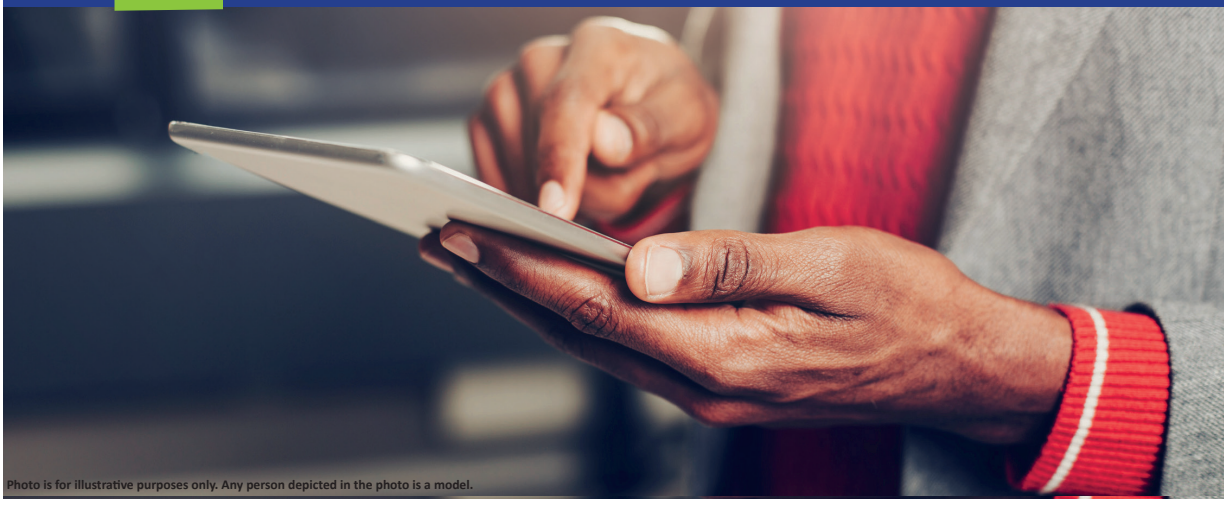


Photo is for illustrative purposes only. Any person depicted in the photo is a model.

H5-SEER

hints

NIH NATIONAL CANCER INSTITUTE

Instructions:
Please use a black or blue pen to complete this form.
Mark ☒ to indicate your answer. To change an answer, darken the box ▩ and mark the correct answer.

A: Looking For Health Information

A1. Have you ever looked for information about cancer from any source?

- ☐ Yes
- ☐ No → **GO TO A3 below**

A2. Based on the results of your most recent search for information about cancer, how much do you agree or disagree with each of the following statements?

	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
a. It took a lot of effort to get the information you needed.....	☐	☐	☐	☐
b. You felt frustrated during your search for the information.....	☐	☐	☐	☐

A3. Overall, how confident are you that you could get advice or information about cancer if you needed it?

- ☐ Completely confident
- ☐ Very confident
- ☐ Somewhat confident
- ☐ A little confident
- ☐ Not confident at all

A4. In general, how much would you trust information about cancer from each of the following?

	Not at all	A little	Some	A lot
a. A doctor.....	☐	☐	☐	☐
b. Family or friends.....	☐	☐	☐	☐
c. Government health agencies...	☐	☐	☐	☐
d. Charitable organizations.....	☐	☐	☐	☐
e. Religious organizations and leaders.....	☐	☐	☐	☐

A5. If you had a strong need to get information about cancer. Where would you go first?

Mark only one.

- ☐ Books
- ☐ Brochures, pamphlets, etc.
- ☐ Cancer organization
- ☐ Family
- ☐ Friend/Co-worker
- ☐ Doctor or health care provider
- ☐ Internet
- ☐ Library
- ☐ Magazines
- ☐ Newspapers
- ☐ Telephone information number
- ☐ Complementary, alternative, or unconventional practitioner
- ☐ Other - Specify → _____

10488

B: Using the Internet to Find Information

B1. Do you ever go on-line to access the Internet or World Wide Web, or to send and receive e-mail?

- ☐ Yes
- ☐ No → **GO TO B5 in the next column**

B2. When you use the Internet, do you access it through...

	Yes	No
a. A regular dial-up telephone line	☐	☐
b. Broadband such as DSL, cable, or FiOS	☐	☐
c. A cellular network (i.e., phone, 3G/4G)	☐	☐
d. A wireless network (Wi-Fi)	☐	☐

B3. How often do you access the Internet through each of the following?

	Daily	Sometimes	Never	Not applicable
a. Computer at home	☐	☐	☐	☐
b. Computer at work	☐	☐	☐	☐
c. Computer in a public place (library, community center, other)	☐	☐	☐	☐
d. On a mobile device (cell phone/smart phone/tablet)	☐	☐	☐	☐

B4. To what extent are you satisfied with your Internet speed?

- ☐ Not at all satisfied
- ☐ Not very satisfied
- ☐ Somewhat satisfied
- ☐ Very satisfied
- ☐ Extremely satisfied

B5. In the past 12 months, have you used a computer, smartphone, or other electronic means to do any of the following?

	Yes	No
a. Looked for health or medical information for yourself	☐	☐
b. Used e-mail or the Internet to communicate with a doctor or a doctor's office	☐	☐
c. Look up medical test results	☐	☐
d. Made appointments with a health care provider	☐	☐

B6. Please indicate if you have each of the following.

Mark <u>all that apply</u>.

- ☐ Tablet computer (for example, an iPad, Samsung Galaxy, Motorola Xoom, or Kindle Fire)
- ☐ Smartphone (for example, an iPhone, Android, Blackberry, or Windows phone)
- ☐ Basic cell phone only
- ☐ I do not have any of the above

GO TO B10 on the next page

B7. On your <u>tablet or smartphone</u>, do you have any "apps" related to health and wellness?

- ☐ Yes
- ☐ No → **GO TO B9 on the next page**
- ☐ Don't know → **GO TO B9 on the next page**

B8. In the past 12 months, have you used any of these health or wellness apps?

- ☐ Yes
- ☐ No
- ☐ Don't know

10488

B9. Has your tablet or smartphone…

		Yes	No
a.	Helped you track progress on a health-related goal such as quitting smoking, losing weight, or increasing physical activity?	☐	☐
b.	Helped you make a decision about how to treat an illness or condition?	☐	☐
c.	Helped you in discussions with your health care provider?	☐	☐

B10. In the past 12 months, have you used an electronic wearable device to monitor or track your health or activity? For example, a Fitbit, Apple Watch, or Garmin Vivofit.

- ☐ Yes
- ☐ No → **GO TO B13 in the next column**

B11. In the past month, how often did you use a wearable device to track your health?

- ☐ Every day
- ☐ Almost every day
- ☐ 1-2 times per week
- ☐ Less than once per week
- ☐ I did not use a wearable device in the past month

B12. Would you be willing to share health data from your wearable device with...

		Yes	No
a.	Your health care provider?	☐	☐
b.	Your family?	☐	☐
c.	Your friends?	☐	☐

B13. Have you shared health information from either an electronic monitoring device or smartphone with a health professional within the last 12 months?

- ☐ Yes
- ☐ No
- ☐ Not Applicable

B14. Sometimes people use the Internet to connect with other people online through social networks like Facebook or Twitter. This is often called “social media”.

In the last 12 months, have you used the Internet for any of the following reasons?

		Yes	No
a.	To visit a social networking site, such as Facebook or LinkedIn	☐	☐
b.	To share health information on social networking sites, such as Facebook or Twitter	☐	☐
c.	To participate in an online forum or support group for people with a similar health or medical issue	☐	☐
d.	To watch a health-related video on YouTube	☐	☐

C: Your Health Care

C1. Not including psychiatrists and other mental health professionals, is there a particular doctor, nurse, or other health professional that you see most often?

- ☐ Yes
- ☐ No

C2. In the past 12 months, not counting times you went to an emergency room, how many times did you go to a doctor, nurse, or other health professional to get care for yourself?

- ☐ None → **GO TO C6 in the next column**
- ☐ 1 time
- ☐ 2 times
- ☐ 3 times
- ☐ 4 times
- ☐ 5-9 times
- ☐ 10 or more times

C3. Overall, how would you rate the quality of health care you received in the past 12 months?

- ☐ Excellent
- ☐ Very good
- ☐ Good
- ☐ Fair
- ☐ Poor

C4. The following questions are about your communication with all doctors, nurses, or other health professionals you saw during the past 12 months.

How often did they do each of the following?	Always	Usually	Sometimes	Never
a. Give you the chance to ask all the health-related questions you had	☐	☐	☐	☐
b. Give the attention you needed to your feelings and emotions	☐	☐	☐	☐
c. Involve you in decisions about your health care as much as you wanted	☐	☐	☐	☐
d. Make sure you understood the things you needed to do to take care of your health	☐	☐	☐	☐
e. Explain things in a way you could understand	☐	☐	☐	☐
f. Spend enough time with you	☐	☐	☐	☐
g. Help you deal with feelings of uncertainty about your health or health care	☐	☐	☐	☐

C5. In the past 12 months, when getting cancer care or care for other medical problems, was there a time when you...

	Yes	No
a. Had to bring an X-ray, MRI, or other type of test result with you to the appointment?	☐	☐
b. Had to wait for test results longer than you thought reasonable?	☐	☐
c. Had to redo a test or procedure because the earlier test results were not available?	☐	☐
d. Had to provide your medical history again because your chart could not be found?	☐	☐

C6. In the past 12 months, have you seen or talked to a mental health professional such as a psychologist, psychiatrist, psychiatric nurse or clinical social worker about your health?

- ☐ Yes
- ☐ No

C7. Are you currently covered by any of the following types of health insurance or health coverage plans?

	Yes	No
a. Insurance through a current or former employer or union	☐	☐
b. Insurance purchased directly from an insurance company	☐	☐
c. Medicare, for people 65 and older, or people with certain disabilities	☐	☐
d. Medicaid, Medical Assistance, or any kind of government-assistance plan for those with low incomes or a disability	☐	☐
e. TRICARE or other military health care	☐	☐
f. VA (including those who have ever used or enrolled for VA health care)	☐	☐
g. Indian Health Service	☐	☐
h. Any other type of health insurance or health coverage plan (Specify)	☐	☐

10488

D: Medical Records

Next, we are going to ask you some questions about your medical records. Medical records are defined as medical history, such as laboratory test results, clinical notes, and current list of medications.

D1. Do any of your doctors or other health care providers maintain your medical records in a computerized system?

- ☐ Yes
- ☐ No
- ☐ Don't know

D2. Have you ever been offered online access to your medical records by your…

	Yes	No	Don't know
a. health care provider?	☐	☐	☐
b. health insurer?	☐	☐	☐

D3. Have any of your health care providers, including doctors, nurses, or office staff ever encouraged you to use an online medical record?

- ☐ Yes
- ☐ No

D4. How many times did you access your online medical record in the last 12 months?

- ☐ 0
- ☐ 1 to 2 times
- ☐ 3 to 5 times
- ☐ 6 to 9 times
- ☐ 10 or more times

GO TO D6 on the next page

D5. Why have you not accessed your medical records online? Is it because…

	Yes	No
a. You prefer to speak to your health care provider directly?	☐	☐
b. You do not have a way to access the website?	☐	☐
c. You did not have a need to use your online medical record?	☐	☐
d. You were concerned about the privacy or security of the website that had your medical records?	☐	☐
e. You don't have an online medical record	☐	☐
f. You found it difficult to login (for example, you had trouble remembering your password)?	☐	☐
g. You are not comfortable or experienced with computers?	☐	☐
h. You have more than one online medical record?	☐	☐

If you have not accessed any medical records in the last 12 months, go to E1 on the next page

Otherwise, go to D6 on the next page

10488

D6. In the past 12 months, have you used your online medical record to…

	Yes	No
a. Look up test results?	☐	☐
b. Securely message health care provider and staff (for example, e-mail)?	☐	☐
c. Download your health information to your computer or mobile device, such as a cell phone or tablet?	☐	☐

D7. How did you access your online medical record?

- ☐ App
- ☐ Website
- ☐ Both app and website
- ☐ Don't know

D8. Do any of your online medical records include clinical notes (health provider's notes that describe a visit)?

- ☐ Yes
- ☐ No
- ☐ Don't know

D9. Have you electronically sent your medical information to….

	Yes	No
a. Another health care provider?	☐	☐
b. A family member or another person involved with your care?	☐	☐
c. A service or app that can help manage and store your health information?	☐	☐

D10. How easy or difficult was it to understand the health information in your online medical record?

- ☐ Very easy
- ☐ Somewhat easy
- ☐ Somewhat difficult
- ☐ Very difficult

E: Cancer History

E1. What type(s) of cancer have you been diagnosed with?

Mark all that apply.

- ☐ I have never been diagnosed as having cancer → **GO TO E9 on the next page**
- ☐ Bladder cancer
- ☐ Bone cancer
- ☐ Breast cancer
- ☐ Cervical cancer (cancer of the cervix)
- ☐ Colon cancer
- ☐ Endometrial cancer (cancer of the uterus)
- ☐ Head and neck cancer
- ☐ Leukemia/Blood cancer
- ☐ Liver cancer
- ☐ Lung cancer
- ☐ Lymphoma (Hodgkin's)
- ☐ Lymphoma (Non-Hodgkin's)
- ☐ Melanoma
- ☐ Non-melanoma skin cancer (basal cell or squamous cell carcinoma)
- ☐ Oral cancer
- ☐ Ovarian cancer
- ☐ Pancreatic cancer
- ☐ Pharyngeal (throat) cancer
- ☐ Prostate cancer
- ☐ Rectal cancer
- ☐ Renal (kidney) cancer
- ☐ Stomach cancer
- ☐ Other - Specify → ☐

E2. At what age were you first told that you had cancer?

☐☐☐ Years old

E3. Did you ever receive any treatment for your cancer?

- ☐ Yes
- ☐ No → **GO TO E6 below**

E4. About how long ago did you receive your last cancer treatment?

- ☐ Still receiving treatment
- ☐ Less than 1 year ago
- ☐ 1 year ago to less than 5 years ago
- ☐ 5 years ago to less than 10 years ago
- ☐ 10 or more years ago

E5. Overall, how would you rate the quality of the cancer care you received when you were treated for cancer?

- ☐ Excellent
- ☐ Very good
- ☐ Good
- ☐ Fair
- ☐ Poor

E6. Have you ever experienced any of the following conditions as a result of your cancer diagnosis or cancer treatment?

	Yes	No
a. Cognitive impairment (for example, having difficulty remembering things, or 'chemobrain')	☐	☐
b. Neuropathy (numbness or tingling feelings)	☐	☐
c. Severe fatigue (always tired or sleepy)	☐	☐
d. Nausea	☐	☐
e. Something else. Specify	☐	☐

E7. Looking back, since the time you were first diagnosed with cancer, how much, if at all, has cancer and its treatment hurt your financial situation?

- ☐ Not at all
- ☐ A little
- ☐ Some
- ☐ A lot

E8. At any time since you were first diagnosed with cancer, did any doctor or other health care provider ever discuss with you the impact of cancer or its treatment on your ability to work?

- ☐ Discussed it with me in detail
- ☐ Briefly discussed it with me
- ☐ Did not discuss it at all
- ☐ I don't remember
- ☐ I was not working at the time of my diagnosis

E9. The following questions ask about your knowledge about cancer in your family. By family we mean your first- and second-degree biological relatives: your parents, brothers and sisters, children, grandparents, aunts and uncles, nieces and nephews.

How well do you know your family's cancer history, including if you have no history of cancers in your family?

- ☐ Not at all
- ☐ A little
- ☐ Somewhat
- ☐ Well
- ☐ Very well

E10. Have any of your first- or second-degree biological relatives (parents, brothers and sisters, children, grandparents, aunts and uncles, nieces and nephews) ever had cancer?

- ☐ Yes
- ☐ No
- ☐ Not sure

10488

F: Impact of COVID-19

F1. The following questions are related to the coronavirus/COVID-19 pandemic that impacted the United States in 2020.

Has the COVID-19 pandemic affected either your cancer treatment or any follow-up medical appointments related to your cancer? Do not include routine cancer screening or preventive care appointments.

Mark all that apply.

- ☐ I have not had any scheduled cancer treatment or any follow-up medical appointments related to my cancer during the pandemic
- ☐ Yes, some or all of my cancer treatment or follow-up medical appointments related to my cancer were cancelled or delayed
- ☐ Yes, some or all of my cancer treatment or follow-up medical appointments related to my cancer were done by phone or video conference instead of in-person (telehealth)
- ☐ No, my cancer treatment or follow-up medical appointments related to my cancer have not been affected by the COVID-19 pandemic

F2. Has the COVID-19 pandemic affected any of your appointments for routine cancer screening or preventive care (e.g., mammography, colonoscopy, etc.)?

Mark all that apply.

- ☐ I have not had any scheduled appointments for routine cancer screening or preventive care during the pandemic
- ☐ Yes, some or all of my appointments for routine cancer screening or preventive care were cancelled or delayed
- ☐ Yes, some or all of my appointments for routine cancer screening or preventive care were done by phone or video conference instead of in-person (telehealth)
- ☐ No, my appointments for routine cancer screening or preventive care have not been affected by the COVID-19 pandemic

F3. Has your cancer treatment plan changed as a result of the COVID-19 pandemic?

- ☐ Yes, my cancer treatment plan changed because of the COVID-19 pandemic
- ☐ No, my cancer treatment plan has not changed because of the COVID-19 pandemic
- ☐ I have not been undergoing cancer treatment during the COVID-19 pandemic

F4. Have any of your healthcare providers discussed, or provided you with information about your risk for COVID-19 complications due to your cancer history?

- ☐ Yes
- ☐ No
- ☐ Don't know

10488

F5. During the COVID-19 pandemic, have you done any of the following things more, less, or about the same as you normally do?

	I've done this MORE	I've done this the SAME	I've done this LESS	I don't do this at all
a. Slept	☐	☐	☐	
b. Ate food in general	☐	☐	☐	
c. Ate high fat or sugary foods	☐	☐	☐	☐
d. Ate healthy food	☐	☐	☐	☐
e. Exercised	☐	☐	☐	☐
f. Drank alcohol	☐	☐	☐	☐
g. Smoked cigarettes or vaped	☐	☐	☐	☐
h. Used cannabis, marijuana, or CBD	☐	☐	☐	☐
i. Used prescription drugs	☐	☐	☐	☐
j. Used non-prescription drugs	☐	☐	☐	☐
k. Connected with others, including talking with people you trust about your concerns and how you are feeling	☐	☐	☐	☐
l. Contacted a healthcare provider	☐	☐	☐	☐
m. Looked for health information	☐	☐	☐	☐
n. Took breaks from watching, reading, or listening to news stories, including social media	☐	☐	☐	☐

F6. How much would you trust each of the following for reliable information about COVID-19?

	Not at all	A little	Some	A lot
a. CDC – Centers for Disease Control and Prevention	☐	☐	☐	☐
b. NIH – National Institutes of Health	☐	☐	☐	☐
c. Your state government	☐	☐	☐	☐
d. Your local government	☐	☐	☐	☐
e. News media	☐	☐	☐	☐
f. Your healthcare provider	☐	☐	☐	☐
g. Your family and friends	☐	☐	☐	☐
h. Social media	☐	☐	☐	☐
i. WHO – The World Health Organization	☐	☐	☐	☐

F7. How much do you agree or disagree with each of the following statements about your feelings towards COVID-19?

	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
a. There's not much people can do to lower their chances of getting COVID-19	☐	☐	☐	☐
b. There are so many different recommendations about preventing COVID-19, it's hard for people to know which ones to follow	☐	☐	☐	☐

10488

G: Genetic Testing

G1. Genes are inherited from your parents and are passed down from one generation to the next. Genetic tests can determine your genetic makeup.

Which of the following types of genetic tests have you heard of?

Mark all that apply.

- ☐ **Ancestry testing:** To determine the background or geographic/ethnic origin of an individual's ancestors (for example, Ancestry.com and 23andMe)
- ☐ **Genetic health risk testing:** To determine health risk for a variety of health conditions (for example, 23andMe)
- ☐ **Cancer genetic testing** (for example, testing for inherited cancer syndromes like BRCA1/2 or Lynch Syndrome)
- ☐ Other - Specify → ☐
- ☐ Not sure
- ☐ I have not heard of any of these types of genetic tests → **GO TO G8 on the next page**

G2. From which of the following sources did you read or hear anything about genetic tests?

Mark all that apply.

- ☐ Newspaper
- ☐ Magazine
- ☐ Radio
- ☐ Your primary health care provider
- ☐ Oncologist/cancer surgeon
- ☐ Genetic counselor
- ☐ Family member
- ☐ Friend
- ☐ Social media
- ☐ Television
- ☐ Internet
- ☐ Other - Specify → ☐
- ☐ Have not heard of such tests → **GO TO G8 on the next page**
- ☐ Not sure

G3. Have you ever had any of the following types of genetic tests?

Mark all that apply.

- ☐ **Ancestry testing:** To determine the background or geographic/ethnic origin of an individual's ancestors (for example, Ancestry.com and 23andMe)
- ☐ **Genetic health risk testing:** To determine health risk for a variety of health conditions (for example, 23andMe)
- ☐ **Cancer genetic testing** (for example, testing for inherited cancer syndromes like BRCA1/2 or Lynch Syndrome)
- ☐ Other - Specify → ☐
- ☐ Not sure
- ☐ None of the above → **GO TO G8 on the next page**

G4. If you had a **cancer genetic test for inherited cancer syndromes**, where did you get information about this type of testing?

Mark all that apply.

- ☐ I did not have cancer genetic testing → **GO TO G6 on the next page**
- ☐ Your primary health care provider
- ☐ Oncologist/cancer surgeon
- ☐ Genetic counselor
- ☐ Genetic testing companies
- ☐ Someplace else. Specify → ☐

G5. Who ordered your cancer genetic test for inherited cancer syndromes?

Mark only one.

- ☐ Your primary health care provider
- ☐ Oncologist/cancer surgeon
- ☐ Genetic counselor
- ☐ I ordered it directly from a genetic testing company
- ☐ I don't know

10488

G6. If you had any genetic test, who helped you understand the results?

Mark all that apply.

- ☐ Your primary health care provider
- ☐ Oncologist/cancer surgeon
- ☐ Genetic counselor
- ☐ Spouse/partner
- ☐ Parents
- ☐ Siblings
- ☐ Children
- ☐ Friend
- ☐ Other
- ☐ No one helped me understand the results

G7. If you had any genetic test, who did you share the results with?

Mark all that apply.

- ☐ Your primary health care provider
- ☐ Oncologist/cancer surgeon
- ☐ Genetic counselor
- ☐ Spouse/partner
- ☐ Parents
- ☐ Siblings
- ☐ Children
- ☐ Friend
- ☐ Other
- ☐ Did not share the results

G8. How much do you think genes that are inherited determine whether or not a person will develop each of the following conditions?

	Not at all	A little	Somewhat	A lot
a. Obesity	☐	☐	☐	☐
b. Cancer	☐	☐	☐	☐
c. Cardiovascular disease	☐	☐	☐	☐
d. Diabetes	☐	☐	☐	☐

G9. How important is knowing a person's genetic information for…

	Not at all	A little	Somewhat	Very
a. Preventing cancer?	☐	☐	☐	☐
b. Detecting cancer early?	☐	☐	☐	☐
c. Treating cancer?	☐	☐	☐	☐

G10. "Precision medicine" is an approach for disease treatment and prevention that takes into account individual differences in genes, environment, and lifestyle.

Before completing this survey, had you ever heard of approaches like precision medicine?

- ☐ Yes
- ☐ No

G11. Precision medicine in the cancer treatment setting may involve doing genetic testing on the cancer tumor or tissue. This is different from genetic testing to look at genes that are inherited from your parents.

Was this type of genetic testing on your cancer tumor or tissue ever discussed with you?

- ☐ Yes
- ☐ No
- ☐ I don't know

G12. Was this type of testing done as part of your cancer diagnosis and/or treatment?

- ☐ Yes
- ☐ No
- ☐ I don't know

10488

H: Clinical Trials

H1. Clinical trials are research studies that involve people. They are designed to compare new kinds of health care with the standard health care people currently get. For example, a new drug or a new way for patients to track their diets.

How would you describe your level of knowledge about clinical trials?

- ☐ I don't know anything about clinical trials
- ☐ I know a little bit about clinical trials
- ☐ I know a lot about clinical trials

H2. Has a doctor or other member of your medical team discussed clinical trials as a possible treatment option for your cancer?

- ☐ Yes
- ☐ No

H3. If you had a need to get information about clinical trials. Which of the following would you go to first to get information about clinical trials?

Mark only one.

- ☐ My health care provider
- ☐ My family and friends
- ☐ Government health agencies
- ☐ Health organizations or groups (for example, the American Cancer Society, American Lung Association)
- ☐ Disease-specific patient support groups
- ☐ Drug companies
- ☐ Internet search

H4. If you had a need to get information about clinical trials. Which of the following would you most trust as a source of information about clinical trials?

Mark only one.

- ☐ My health care provider
- ☐ My family and friends
- ☐ Government health agencies
- ☐ Health organizations or groups (for example, the American Cancer Society, American Lung Association)
- ☐ Disease-specific patient support groups
- ☐ Drug companies

H5. Have you ever heard of the website clinicaltrials.gov?

- ☐ Yes
- ☐ No

H6. Have you ever participated in a clinical trial for treatment of your cancer?

- ☐ Yes
- ☐ No → **GO TO J1 on the next page**
- ☐ Don't know → **GO TO J1 on the next page**

10488

H7. If you participated in a clinical trial, how much did each of the following influence your decision to participate?

	Not at all	A little	Somewhat	A lot	Not Applicable
a. My participation will help other people.....	☐	☐	☐	☐	☐
b. I was paid to participate.....	☐	☐	☐	☐	☐
c. I was given support to participate such as transportation, childcare, or paid time off from work...	☐	☐	☐	☐	☐
d. My doctor encouraged me to participate.....	☐	☐	☐	☐	☐
e. My family and friends encouraged me to participate.....	☐	☐	☐	☐	☐
f. I thought that participating would help me get better....	☐	☐	☐	☐	☐
g. I wanted the chance to try a new kind of care.....	☐	☐	☐	☐	☐
h. The standard care was not covered by my insurance.....	☐	☐	☐	☐	☐

J: Your Overall Health

J1. In general, would you say your health is...

- ☐ Excellent,
- ☐ Very good,
- ☐ Good,
- ☐ Fair, or
- ☐ Poor?

J2. Overall, how confident are you about your ability to take good care of your health?

- ☐ Completely confident
- ☐ Very confident
- ☐ Somewhat confident
- ☐ A little confident
- ☐ Not confident at all

J3. Are you deaf or do you have serious difficulty hearing?

- ☐ Yes
- ☐ No

J4. Are you blind or do you have serious difficulty seeing, even when wearing glasses?

- ☐ Yes
- ☐ No

J5. Because of a physical, mental, or emotional condition, do you have serious difficulty concentrating, remembering, or making decisions?

- ☐ Yes
- ☐ No

J6. Do you have serious difficulty walking or climbing stairs?

- ☐ Yes
- ☐ No

J7. Do you have difficulty dressing or bathing?

- ☐ Yes
- ☐ No

J8. Because of a physical, mental, or emotional condition, do you have difficulty doing errands alone such as visiting a doctor's office or shopping?

- ☐ Yes
- ☐ No

J9. Is there anyone you can count on to provide you with emotional support when you need it - such as talking over problems or helping you make difficult decisions?

- ☐ Yes
- ☐ No

10488

J10. Do you have friends or family members that you talk to about your health?

- ☐ Yes
- ☐ No

J11. If you needed help with your daily chores, is there someone who can help you?

- ☐ Yes
- ☐ No

J12. How often do you feel that you lack companionship?

- ☐ Never
- ☐ Rarely
- ☐ Sometimes
- ☐ Always

J13. How often do you feel that you have a lot in common with the people around you?

- ☐ Never
- ☐ Rarely
- ☐ Sometimes
- ☐ Always

J14. How often do you feel close to people?

- ☐ Never
- ☐ Rarely
- ☐ Sometimes
- ☐ Always

J15. Please respond to each item by marking one box per row.

	Never	Rarely	Sometimes	Usually	Always
a. I feel left out	☐	☐	☐	☐	☐
b. I feel that people barely know me	☐	☐	☐	☐	☐
c. I feel isolated from others	☐	☐	☐	☐	☐
d. I feel that people are around me but not with me	☐	☐	☐	☐	☐

J16. Has a doctor or other health professional ever told you that you had any of the following medical conditions:

	Yes	No
a. Diabetes or high blood sugar?	☐	☐
b. High blood pressure or hypertension?	☐	☐
c. A heart condition such as heart attack, angina, or congestive heart failure?	☐	☐
d. Chronic lung disease, asthma, emphysema, or chronic bronchitis?	☐	☐
e. Depression or anxiety disorder?	☐	☐

J17. About how tall are you without shoes?

☐ Feet ***and*** ☐☐ Inches

J18. About how much do you weigh, in pounds, without shoes?

☐☐☐ Pounds

10488

J19. Over the past 2 weeks, how often have you been bothered by any of the following problems?

	Nearly every day	More than half the days	Several days	Not at all
a. Little interest or pleasure in doing things	☐	☐	☐	☐
b. Feeling down, depressed, or hopeless	☐	☐	☐	☐
c. Feeling nervous, anxious, or on edge	☐	☐	☐	☐
d. Not being able to stop or control worrying	☐	☐	☐	☐

J20. How much do you agree or disagree with the following statements?

	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
a. When I feel threatened or anxious I find myself thinking about my values	☐	☐	☐	☐
b. When I feel threatened or anxious I find myself thinking about my strengths	☐	☐	☐	☐

J21. How much do you agree or disagree with the following statement?

I go to medical appointments expecting the worst.

- ☐ Strongly agree
- ☐ Agree
- ☐ Neither agree nor disagree
- ☐ Disagree
- ☐ Strongly disagree

J22. From the set of values below, which ONE is most important to you in your day-to-day life?

Mark only one.

- ☐ Making my own decisions
- ☐ Being happy
- ☐ Helping people
- ☐ Being loyal to family and friends
- ☐ Having a deep connection to my religion
- ☐ Keeping myself in good health
- ☐ Assuring my family is safe and secure

K: Health and Nutrition

K1. Thinking about the last time you ordered food in a fast food or sit down restaurant, did you notice calorie information listed next to the food on the menu or menu board?

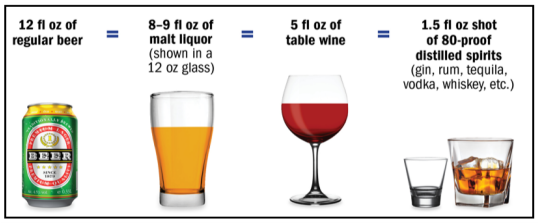
- ☐ Yes
- ☐ No

K2. To what extent would you support or oppose the following?

Junk food products, including candy, chips, soda, and flavored sports drinks, should not be advertised to children on social media.

- ☐ Strongly oppose
- ☐ Oppose
- ☐ Neither support nor oppose
- ☐ Support
- ☐ Strongly support

K3. These are examples of one drink of alcohol:



During the past 30 days, how many days per week did you have at least one drink of any alcoholic beverage?

☐ Days per week

→ **(IF 0 THEN GO TO K6 in the next column)**

K4. During the past 30 days, on the days when you drank, about how many drinks did you drink on average?

☐☐ Average drinks per day

K5. ***For males:*** During the past 30 days, how many times did you have 5 or more alcoholic drinks on one occasion?

For females: During the past 30 days, how many times did you have 4 or more alcoholic drinks on one occasion?

- ☐ Never
- ☐ 1 or 2 times
- ☐ 3 to 5 times
- ☐ 6 to 10 times
- ☐ 11 or more times

K6. In your opinion, how much does drinking the following types of alcohol affect the risk of getting cancer?

	Decreases risk a lot	Decreases risk a little	No effect	Increases risk a little	Increases risk a lot	Don't know
a. Beer	☐	☐	☐	☐	☐	☐
b. Wine	☐	☐	☐	☐	☐	☐
c. Liquor	☐	☐	☐	☐	☐	☐

K7. In your opinion, how much does drinking the following types of alcohol affect the risk of getting heart disease?

	Decreases risk a lot	Decreases risk a little	No effect	Increases risk a little	Increases risk a lot	Don't know
a. Beer	☐	☐	☐	☐	☐	☐
b. Wine	☐	☐	☐	☐	☐	☐
c. Liquor	☐	☐	☐	☐	☐	☐

K8. To reduce the problems associated with excessive alcohol use, to what extent would you support or oppose...

	Strongly oppose	Oppose	Neither support nor oppose	Support	Strongly support
a. Banning outdoor advertising of alcohol such as on billboards and bus stops?	☐	☐	☐	☐	☐
b. Requiring specific health warnings on alcohol containers?	☐	☐	☐	☐	☐
c. Requiring alcohol containers to show the recommended drinking guidelines for keeping health risks low?	☐	☐	☐	☐	☐

10488

L: Physical Activity and Exercise

L1. In a typical week, how many days do you do any physical activity or exercise of at least moderate intensity, such as brisk walking, bicycling at a regular pace, and swimming at a regular pace (do not include weightlifting)?

- ☐ None → **GO TO L3 below**
- ☐ 1 day per week
- ☐ 2 days per week
- ☐ 3 days per week
- ☐ 4 days per week
- ☐ 5 days per week
- ☐ 6 days per week
- ☐ 7 days per week

L2. On the days that you do any physical activity or exercise of at least moderate intensity, how long do you typically do these activities?

☐☐☐ **Minutes per day**

L3. In a typical week, outside of your job or work around the house, how many days do you do leisure-time physical activities specifically designed to strengthen your muscles such as lifting weights or circuit training (do not include cardio exercise such as walking, biking, or swimming)?

- ☐ None
- ☐ 1 day per week
- ☐ 2 days per week
- ☐ 3 days per week
- ☐ 4 days per week
- ☐ 5 days per week
- ☐ 6 days per week
- ☐ 7 days per week

L4. During the past 7 days, how much time did you spend sitting on a typical day at home or at work? This may include time spent sitting at a desk, visiting friends, reading, driving or riding in a car, or sitting or lying down to watch television.

☐☐ **Hours per day**

M: Tobacco Products

M1. Have you smoked at least 100 cigarettes in your entire life?

- ☐ Yes
- ☐ No → **GO TO M3 below**

M2. How often do you now smoke cigarettes?

- ☐ Every day
- ☐ Some days
- ☐ Not at all

M3. New types of cigarettes are now available called electronic cigarettes or e-cigarettes (also known as vapes, vape-pens, tanks, mods or pod-mods). These products deliver nicotine through a vapor. Compared to smoking cigarettes, would you say that electronic cigarettes are…

- ☐ Much less harmful,
- ☐ Less harmful,
- ☐ Just as harmful,
- ☐ More harmful,
- ☐ Much more harmful, or
- ☐ I don't know

10488

M4. Have you ever used an e-cigarette, even one or two times?

- ☐ Yes
- ☐ No → **GO TO M6 below**

M5. Do you now use an e-cigarette every day, some days, or not at all?

- ☐ Every day
- ☐ Some days
- ☐ Not at all

M6. Heated tobacco products, also known as heat-not-burn tobacco products, use a technology that heats tobacco instead of burning it. These are NOT the same as e-cigarettes. Some brands of heated tobacco products include IQOS and Eclipse.

Thinking about heated tobacco products, which of the following statements BEST applies to you?

- ☐ I have never heard of heated tobacco products
- ☐ I have heard of heated tobacco products but have never tried them
- ☐ I have tried heated tobacco products but do not use them anymore
- ☐ I currently use heated tobacco products
- ☐ Don't know

M7. To what extent would you support or oppose the following measures related to cigarettes?

	Strongly oppose	Oppose	Neither support nor oppose	Support	Strongly support
a. Just like with violence and sex, movies with cigarette smoking should be rated "R" to protect children and youth from seeing cigarette smoking in movies.....	☐	☐	☐	☐	☐
b. Cigarette packs should be required to have warning labels that use both images and words to show the negative health effects of smoking.....	☐	☐	☐	☐	☐

M8. To what extent would you support or oppose the following measures related to all tobacco products, including cigarettes, e-cigarettes, smokeless tobacco, hookah, and cigars?

	Strongly oppose	Oppose	Neither support nor oppose	Support	Strongly support
a. Stores should be required to keep tobacco products out of customers' view at the checkout counter.....	☐	☐	☐	☐	☐
b. Stores should be required to keep advertisements for tobacco products away from cash registers and out of windows.....	☐	☐	☐	☐	☐
c. Tobacco products should not be advertised on social media.....	☐	☐	☐	☐	☐

N: Cancer Screening and Awareness

N1. ***For males:*** GO TO N3 below

For females: How long ago did you have your most recent Pap test to check for cervical cancer?

- ☐ A year ago or less
- ☐ More than 1, up to 2 years ago
- ☐ More than 2, up to 3 years ago
- ☐ More than 3, up to 5 years ago
- ☐ More than 5 years ago
- ☐ I have never had a Pap test

N2. When did you have your most recent mammogram to check for breast cancer, if ever?

- ☐ A year ago or less
- ☐ More than 1, up to 2 years ago
- ☐ More than 2, up to 3 years ago
- ☐ More than 3, up to 5 years ago
- ☐ More than 5 years ago
- ☐ I have never had a mammogram

N3. There are a few different tests to check for colon cancer. These tests include:

A **colonoscopy** - For this test, a tube is inserted into your rectum and you are given medication that may make you feel sleepy. After the procedure, you need someone to drive you home.

A **sigmoidoscopy** - For this test, you are awake when the tube is inserted into your rectum. After the test you can drive yourself home.

A **stool blood test** - For this test, you collect a stool sample at home, and then provide it to a doctor or lab for testing.

Have you ever had one of these tests to check for colon cancer?

- ☐ Yes
- ☐ No

N4. Have you ever heard of **HPV**? HPV stands for Human Papillomavirus. It is not HCV, HIV, HSV, or herpes.

- ☐ Yes
- ☐ No → **GO TO N6 below**

N5. Do you think **HPV** can cause...

	Yes	No	Not sure
a. Cervical Cancer?	☐	☐	☐
b. Penile Cancer?	☐	☐	☐
c. Anal Cancer?	☐	☐	☐
d. Oral Cancer?	☐	☐	☐

N6. A vaccine to prevent **HPV** infection is available and is called the HPV shot, cervical cancer vaccine, GARDASIL®.

Before today, have you ever heard of the cervical cancer vaccine or HPV shot?

- ☐ Yes
- ☐ No

O: Beliefs About Cancer

O1. How worried are you about getting cancer again?

- ☐ Not at all
- ☐ Slightly
- ☐ Somewhat
- ☐ Moderately
- ☐ Extremely

10488

O2. How much do you agree or disagree with each of the following statements?

	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
a. It seems like everything causes cancer.....	☐	☐	☐	☐
b. There's not much people can do to lower their chances of getting cancer	☐	☐	☐	☐
c. There are so many different recommendations about preventing cancer, it's hard to know which ones to follow.....	☐	☐	☐	☐

O3. How much do you think that each of the following can influence whether or not a person will develop cancer?

	A lot	A little	Not at all	Don't know
a. Being overweight or obese.....	☐	☐	☐	☐
b. Gaining weight in adult life.....	☐	☐	☐	☐
c. Eating too much red meat.....	☐	☐	☐	☐

P: You and Your Household

P1. What is your age?

☐☐☐ Years old

P4. In the past 30 days, did you usually work 35 hours or more per week in total at all jobs or businesses?

- ☐ Yes
- ☐ No

P5. Which of the following best describe your current occupational status?

Mark <u>all that apply</u>.

- ☐ Employed
- ☐ Unemployed for 1 year or more
- ☐ Unemployed for less than 1 year
- ☐ Homemaker
- ☐ Student
- ☐ Retired
- ☐ Disabled
- ☐ Other-Specify → ☐

P6. What is your marital status?

- ☐ Married
- ☐ Living as married or living with a romantic partner
- ☐ Divorced
- ☐ Widowed
- ☐ Separated
- ☐ Single, never been married

10488

P7. What is the highest grade or level of schooling you completed?

- ☐ Less than 8 years
- ☐ 8 through 11 years
- ☐ 12 years or completed high school
- ☐ Post high school training other than college (vocational or technical)
- ☐ Some college
- ☐ College graduate
- ☐ Postgraduate

P8. Are you of Hispanic, Latino/a, or Spanish origin? One or more categories may be selected.

Mark all that apply.

- ☐ No, not of Hispanic, Latino/a, or Spanish origin
- ☐ Yes, Mexican, Mexican American, Chicano/a
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino/a, or Spanish origin

P9. What is your race? One or more categories may be selected.

Mark all that apply.

- ☐ White
- ☐ Black or African American
- ☐ American Indian or Alaska Native
- ☐ Asian Indian
- ☐ Chinese
- ☐ Filipino
- ☐ Japanese
- ☐ Korean
- ☐ Vietnamese
- ☐ Other Asian
- ☐ Native Hawaiian
- ☐ Guamanian or Chamorro
- ☐ Samoan
- ☐ Other Pacific Islander

P10. How much do you agree or disagree with the following statement?

I have a strong sense of belonging to my own ethnic group.

- ☐ Strongly agree
- ☐ Agree
- ☐ Neither agree nor disagree
- ☐ Disagree
- ☐ Strongly disagree

P11. Do you think of yourself as…

- ☐ Heterosexual, or straight
- ☐ Homosexual, or gay or lesbian
- ☐ Bisexual
- ☐ Something else – Specify →

P12. Including yourself, how many people live in your household?

☐☐ Number of people

P13. How many children under the age of 18 live in your household?

☐☐ Number of children under 18

P14. Thinking about politics these days, how would you describe your own political viewpoint?

- ☐ Very Liberal
- ☐ Liberal
- ☐ Somewhat Liberal
- ☐ Moderate
- ☐ Somewhat Conservative
- ☐ Conservative
- ☐ Very Conservative

10488

P15. Thinking about members of your family living in this household, what is your combined annual income, meaning the total pre-tax income from all sources earned in the past year?

- ☐ $0 to $9,999
- ☐ $10,000 to $14,999
- ☐ $15,000 to $19,999
- ☐ $20,000 to $34,999
- ☐ $35,000 to $49,999
- ☐ $50,000 to $74,999
- ☐ $75,000 to $99,999
- ☐ $100,000 to $199,999
- ☐ $200,000 or more

P16. Which one of these comes closest to your own feelings about your household's income?

- ☐ Living comfortably on present income
- ☐ Getting by on present income
- ☐ Finding it difficult on present income
- ☐ Finding it very difficult on present income

P17. We invite you to participate in future research studies. These studies are voluntary and will involve answering surveys similar to this one a few times a year.

Can we send you a request to participate in additional studies?

- ☐ Yes
- ☐ No → **GO TO END**

P18. To make it easier to contact you, could you provide your e-mail address in the box below? This is voluntary and we will follow-up by mail if you do not provide an e-mail address.

E-mail: ☐

Thank you!

Please return this survey in the postage-paid envelope within 2 weeks.

If you have lost the envelope, mail the completed survey to:

HINTS Study, TC 1046F
Westat
1600 Research Boulevard
Rockville, MD 20850

10488

Appendix B

HINTS-SEER Greater Bay Area Consent Form

Appendix B
HINTS-SEER Greater Bay Area Consent Form

CONSENT FORM

Before you start the HINTS survey, we need to obtain your consent to participate. Please read the statements below and initial at the bottom of the page if you agree to participate.

- HINTS is a study about experiences with cancer, health in general, and how people get health information. For example, we will ask how you usually get information about health and what sources of information you most trust. We will also ask about your beliefs on what contributes to good health, how best to prevent disease, and other health related topics. You can find out more about HINTS at hints.cancer.gov.

- Your participation is completely voluntary. You can skip any questions you do not wish to answer.

- The survey will take around 30 minutes to complete.

- Some of the questions ask about topics that may be considered sensitive, such as alcohol use, tobacco use, and mental health, and completing the survey may therefore cause some discomfort and anxiety. You can skip any question or set of questions that you do not feel comfortable answering.

- By completing the survey, there is a small risk to your privacy that may result from linking your survey responses to information from the cancer registry. However, we have taken measures to ensure that your private information will not be disclosed:
  - Once received, your completed survey will be given an anonymous code that will prevent it from being linked to your name, address or other personal information.
  - Your name will not appear in any written reports or publications stemming from this research.
  - Your answers will be combined with those of other people who complete the survey.

- There are no direct benefits to you for taking part in this survey, but your answers will help us understand the information needs of people who have had cancer.

- We are including a postage-paid envelope for you to return your completed survey.

- If you have questions about this research, please contact Kelly Blake, the Principal Investigator at 240-281-5934 or kelly.blake@nih.gov.

- If you have questions about your rights and welfare as a research participant, please call the Westat Human Subjects Protections office at 1-888-920-7631. Please leave a message with your full name, the name of the research study that you are calling about (HINTS), and a phone number beginning with the area code. Someone will return your call as soon as possible.

If you have read this consent form and agree to participate in this survey, please initial here: _____

Be sure to initial here!

Appendix C

Cover Letters for Iowa and New Mexico Cohorts, Reminder Postcard for all Cohorts

Appendix C
Cover Letters for Iowa and New Mexico Cohorts, Reminder Postcard for all Cohorts

NIH NATIONAL CANCER INSTITUTE

FIRST MAILING

Dear {name}:

We received your name and address from the {state} cancer registry. We invite you to take part in an important national survey sponsored by the National Cancer Institute: the Health Information National Trends Survey (HINTS). The goal of HINTS is to learn about how people find and use health and medical information. HINTS collects information from adults all over the country. By completing this survey, you will help us learn what health information you need and how to make that information available to you, your family, and your community.

Your participation in HINTS is voluntary and your responses will not be linked to your name or household. We have enclosed $2 as a token of our appreciation for your participation.

You can find out more about HINTS at hints.cancer.gov. Westat, a research firm, is conducting the survey. If you have any questions about HINTS, please call Westat toll-free at 1-888-738-6805.

Thank you in advance for your participation.

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
National Cancer Institute
U.S. Dept. of Health and Human Services

The Health Information National Trends Survey is authorized under 42 USC, Section 285A.

Health Information National Trends Survey hints

U.S. Department of Health & Human Services | National Institutes of Health

POSTCARD TEXT

A few days ago, you should have received a questionnaire packet asking for your participation in the Health Information National Trends Survey (HINTS). By participating in HINTS, you can help the National Cancer Institute determine the best ways of communicating important health information to members of your community.

If you have already completed the survey and returned it to us, please accept my sincere thanks. If you have not yet completed and returned the survey, we ask that you please do so as soon as possible.

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
National Cancer Institute
U.S. Dept. of Health and Human Services

NIH NATIONAL CANCER INSTITUTE

SECOND AND THIRD MAILINGS

Dear {name}:

We received your name and address from the {state} cancer registry. We recently invited you to take part in an important national survey sponsored by the National Cancer Institute: the Health Information National Trends Survey (HINTS). The goal of HINTS is to learn about how people find and use health and medical information. HINTS collects information about adults all over the country. Your responses will help us keep you, your family, and members of your community better informed on the health issues that matter to you.

We have not yet received your completed survey. To make sure HINTS provides accurate information, we need everyone invited to participate in this study to complete the survey. If you did send back your survey and it crossed in the mail with this letter, thank you for the time you took to help make this study a success. In the event that your survey was misplaced, an additional copy is enclosed.

Additional information about HINTS is available at hints.cancer.gov. Westat, a research firm, is conducting the survey. If you have any questions about HINTS, please call Westat toll free at 1-888-738-6805.

Thank you in advance for contributing to this important national study.

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
National Cancer Institute
U.S. Dept. of Health and Human Services

The Health Information National Trends Survey is authorized under 42 USC, Section 285A.

Health Information National Trends Survey hints

U.S. Department of Health & Human Services | National Institutes of Health

Appendix D

Cover Letters for Greater Bay Area Cohort

Appendix D
Cover Letters for Greater Bay Area Cohort

NIH NATIONAL CANCER INSTITUTE

FIRST MAILING

Dear {name}:

We received your name and address from the California Cancer Registry (please see the enclosed brochure). We invite you to take part in an important national survey sponsored by the National Cancer Institute: the Health Information National Trends Survey (HINTS). The goal of HINTS is to learn about how people find and use health and medical information. HINTS collects information from adults all over the country. By completing this survey, you will help us learn what health information you need and how to make that information available to you, your family, and your community.

Your participation in HINTS is voluntary. In addition to asking about your opinions, the HINTS survey will also ask about some things that people consider sensitive such as alcohol use, tobacco use, and mental health. Your responses will not be linked to your name or household and you can skip any questions you do not want to answer. We have enclosed $2 as a token of our appreciation for your participation.

You can find out more about HINTS at hints.cancer.gov. Westat, a research firm, is conducting the survey. If you have any questions about HINTS, please call Westat toll-free at 1-888-738-6805.

Please be sure to read and put your initials on page 1 of the survey to indicate your consent to participate. Once you fill out the rest of the survey, you can return it using the enclosed postage-paid enveloped. Thank you in advance for your participation.

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
National Cancer Institute
U.S. Dept. of Health and Human Services

The Health Information National Trends Survey is authorized under 42 USC, Section 285A.

Health Information National Trends Survey hints

U.S. Department of Health & Human Services | National Institutes of Health

NIH NATIONAL CANCER INSTITUTE

SECOND AND THIRD MAILINGS

Dear {name}:

We received your name and address from the California Cancer. We recently invited you to take part in an important national survey sponsored by the National Cancer Institute: the Health Information National Trends Survey (HINTS). The goal of HINTS is to learn about how people find and use health and medical information. HINTS collects information about adults all over the country. Your responses will help us keep you, your family, and members of your community better informed on the health issues that matter to you.

We have not yet received your completed survey. To make sure HINTS provides accurate information, we need everyone invited to participate in this study to complete the survey. If you did send back your survey and it crossed in the mail with this letter, thank you for the time you took to help make this study a success. In the event that your survey was misplaced, an additional copy is enclosed.

Your participation in HINTS is voluntary. In addition to asking your opinions, the HINTS survey will also ask about some things that some people consider sensitive such as alcohol use, tobacco use and mental health. Your responses will not be linked to your name or household and you can skip any questions you do not want to answer.

Additional information about HINTS is available at hints.cancer.gov. Westat, a research firm, is conducting the survey. If you have any questions about HINTS, please call Westat toll free at 1-888-738-6805.

Please be sure to read and put your initials on page 1 of the survey to indicate your consent to participate. Once you fill out the rest of the survey, you can return it using the enclosed postage-paid envelope. Thank you in advance for contributing to this important national study.

Sincerely,

Kelly D. Blake, ScD
Director, HINTS
National Cancer Institute
U.S. Dept. of Health and Human Services

The Health Information National Trends Survey is authorized under 42 USC, Section 285A.

Health Information National Trends Survey hints

U.S. Department of Health & Human Services | National Institutes of Health

NIH NATIONAL CANCER INSTITUTE

FOURTH MAILING

Dear {name}:

Thank you for taking the time to respond to the National Cancer Institute's Health Information National Trends Survey (HINTS). Your responses will help us learn how people use health information and technology, and how to make health information available to people like you, your family, and your community.

Although we received your completed survey, we did not receive your signed consent to participate. The consent form was on the first page of the survey and the space to initial may have been difficult to see. We are re-sending it here. **Please read and initial the consent form included with this letter and return it using the enclosed postage-paid envelope.** By doing so, we will be able to use your responses to the survey and make your opinions count.

If you have any questions about HINTS, please call Westat toll-free at 1-888-738-6805. You can also find out more about HINTS at hints.cancer.gov.

Thank you for participating in this important national study.

Sincerely,

Kelly D. Blake, ScD

Director, HINTS
National Cancer Institute
U.S. Dept. of Health and Human Services

The Health Information National Trends Survey is authorized under 42 USC, Section 285A.

Health Information National Trends Survey hints

U.S. Department of Health & Human Services | National Institutes of Health

Appendix E

FAQs for Iowa and New Mexico Cohorts

Appendix E
FAQs for Iowa and New Mexico Cohorts

Some Frequently Asked Questions about the Health Information National Trends Survey

Q: What is the study about? What kind of questions do you ask?

A: You can find out more about HINTS at hints.cancer.gov. The study concerns health and how people get health information. For example, we will ask how you usually get information about health and what sources of information you most trust. We will also ask about your beliefs on what contributes to good health, how best to prevent disease, and other health related topics.

Q: How will the study results be used?

A: The results will help the National Cancer Institute promote good health and prevent disease by determining the best ways to communicate accurate health information.

Q: How did you get my address?

A: Your name and address were provided to us by your state's cancer registry.

Q: Why should I take part in this study? Do I have to do this?

A: Getting answers from everyone chosen for the study is the best way to make sure the study results reflect the thoughts and opinions of people across the United States. Your participation is voluntary, and you may refuse to answer any questions or leave the study at any time. However, your answers are very important to the success of this study and will represent thousands of others.

Q: Will my answers to the survey be kept private?

A: Yes. Your answers will be kept private under the Privacy Act. Your answers cannot be linked to any information that could identify you to the extent provided by law. Your completed survey will be stored in a secure file with restricted access. All contact information (such as your name and mailing address) will be destroyed shortly after the research is finalized.

Q: How long will it take to answer the questions?

A: About 20 to 30 minutes.

Q: Who is sponsoring the study?

A: The study is sponsored by the National Cancer Institute, a part of the National Institutes of Health.

Q: Who is Westat?

A: Westat is a research company located in Rockville, Maryland. Westat is conducting this survey under contract to the U.S. Department of Health and Human Services.

Appendix F

FAQs for Greater Bay Area Cohort, CCR Patient Notification Brochure

Appendix F
FAQs for Greater Bay Area Cohort, CCR Patient Notification Brochure

Some Frequently Asked Questions about the Health Information National Trends Survey

Q: What is the study about? What kind of questions do you ask?

A: You can find out more about HINTS at hints.cancer.gov. The study concerns health and how people get health information. For example, we will ask how you usually get information about health and what sources of information you most trust. We will also ask about your beliefs on what contributes to good health, how best to prevent disease, and other health related topics.

Q: How will the study results be used?

A: The results will help the National Cancer Institute promote good health and prevent disease by determining the best ways to communicate accurate health information.

Q: How did you get my address?

A: Your name and address were provided to us by your state's cancer registry.

Q: Why should I take part in this study? Do I have to do this?

A: Getting answers from everyone chosen for the study is the best way to make sure the study results reflect the thoughts and opinions of people across the United States. Your participation is voluntary, and you may refuse to answer any questions or leave the study at any time. However, your answers are very important to the success of this study and will represent thousands of others.

Q: Will my answers to the survey be kept private?

A: Yes. Your answers will be kept private under the Privacy Act. Your answers cannot be linked to any information that could identify you to the extent provided by law. Your completed survey will be stored in a secure file with restricted access. All contact information (such as your name and mailing address) will be destroyed shortly after the research is finalized.

Q: What are the risks of participating in the survey?

A: Some of the questions ask about topics that may be considered sensitive, such as alcohol use, tobacco use, and mental health, and completing the survey may therefore cause some discomfort and anxiety. You can skip any question or set of questions that you do not feel comfortable answering. There is a small risk to your privacy that may result from linking your survey responses to information from the cancer registry. However, we have taken measures to ensure that your private information will not be disclosed. Once received, your completed survey will be given an anonymous code that will prevent it from being linked to your name, address, or other personal information. Your name will not appear in any written reports or publications stemming from this research. Your answers will be combined with those of other people who complete the survey.

Q: How long will it take to answer the questions?

A: About 20 to 30 minutes.

Q: Who is sponsoring the study?

A: The study is sponsored by the National Cancer Institute, a part of the National Institutes of Health.

Q: Who is Westat?

A: Westat is a research company located in Rockville, Maryland. Westat is conducting this survey under contract to the U.S. Department of Health and Human Services.

Source: California Cancer Registry Patient Notification Brochure

What is the California Cancer Registry?

Every state has a cancer registry, and the California Cancer Registry (CCR) is California's cancer tracking system. CCR has collected information about cancer in California since 1988 and is part of the California Department of Public Health.

It is the law in California that all cancer cases be reported to CCR.*

CCR collects information about cancer in California. Information is used to learn more about new cancer cases, cancer treatment, cancer screening programs, and cancer outcomes.

Health care providers, hospitals, and cancer treatments facilities are required by law to submit report of cancer diagnoses and treatment to CCR.

Cancer research is important.

Every person in California benefits from cancer research. With the help of people like you, hundreds of research studies using CCR information have helped us understand the causes of cancer, and improve cancer treatment and outcomes.

Why collect cancer cases?

Cancer cases are collected to monitor the number of new cancer cases, track cancer outcomes, respond to public concerns, and to invite people to join research studies.

Your information is safe.**

Information about cancer patients is kept private and secure. Outpatient information is protected by law. Researchers may ask you to join a study. They never share names or personal details with others.

You may be invited to join a research study.

If researchers have approval from the California Committee for the Protection of Human Subjects, they can use your record to contact you to join a research study. The California Committee for the Protection of Human Subjects protects you by reviewing research activities.

Your Rights.

You do not have to join a research study. Medical care and cancer treatment will not be affected if patients do not want to take part in a research study.

You may report any concerns about cancer research to the California Committee for the Protection of Human Subjects. You can contact CCR and ask that your contact information not be shared with researchers.

Where can I go for more information?

California Cancer Registry

website: www.ccrcal.org

phone: 916-731-2500

Brochure available for download on www.ccrcal.org

CALIFORNIA CANCER REGISTRY

*California Health and Safety Sections 103885
**California Health and Safety Sections 100330

Source: California Cancer Registry Patient Notification Brochure

Appendix G

Variable Values and Data Editing Procedures

Appendix G
Variable Values and Data Editing Procedures

Missing Value Definitions

Values identifying types of nonresponse or indeterminate responses:

- -1 = Valid skips or appropriately missing data following a dependent question (correctly skipped). Example: If SeekCancerInfo=2 'no' and CancerLotOfEffort was missing, CancerLotOfEffort was assigned the value -1.

- -2 = Question was answered, but respondent should not have answered the question. The question was answered in error by the respondent. Example: If SeekCancerInfo=2 'no' and CancerLotOfEffort was not missing, CancerLotOfEffort was assigned the value -2.

- -4 = Question was answered, but data was removed because the entry of the number or character could not be determined (e.g. unreadable or non-conforming numeric response).

- -5 = Respondent selected more response options than appropriate for the question. Example: If CancerTrustDoctor had values 3 'a little' and 2 'some', CancerTrustDoctor was assigned the value -5. In cases where both -2 and -5 values could be assigned, the -2 value was assigned.

- -6 = Missing data in variables following a missing filter question. Example: If filter question (e.g., SeekCancerInfo) was missing and variables up until the next applicable question (e.g. CancerConfidentGetHealthInf) were missing (e.g., CancerLotOfEffort = missing and CancerFrustrated = missing), variables with missing values were assigned the value -6.

- -9 = Missing data. Not ascertained. Question should have been answered, but no response was recorded. Example: If CancerConfidentGetHealthInf was missing, it was assigned the value -9.

Data Editing Procedures

Variable	Editing rule	Description of rule
SeekCancerInfo UseInternet WearableDevTrackHealth UndergoCancerTreatment Smoke100 UsedECigEver HeardHPV FutureStudies	Recoding filter/skip questions	For these filter questions (questions containing a skip instruction associated with the particular response that was selected), response patterns following the question were examined if the filter question was not answered. The 'yes' value (in the majority of cases where a 'yes' response instructed a respondent to continue answering the subsequent questions) was substituted for the missing filter question when any of the subsequent questions were answered. Similarly (when a 'no' response instructed a respondent to skip subsequent questions), the 'no' value was substituted for the missing filter question when all of the subsequent questions that a 'no' response would have directed the respondent to skip were left unanswered and the respondent answered the next applicable question all respondents were supposed to answer. Please note that if neither condition was met, the missing response code values were retained.
SEERStrongNeedCancerInfo_IMP WhoOrderedCaTest_IMP FirstInfoClinTrials2_IMP TrustInfoClinTrials2_IMP MostImportantValues_IMP SexualOrientation_I	Imputation for multiple responses	Imputation was carried out when multiple responses were selected, resulting in one unique response for these "mark only one" variables. Respondent's multiple answers were replaced with a single imputed answer that had the same distribution over the multiple answers as occurred in the single-answer responses. Imputation was not performed on missing values for this question. The suffixes "_IMP" and "_I" indicate that these variables include imputed values. Flags (indicated by suffix '_IFlag') indicate which values were imputed.

Variable	Editing rule	Description of rule
Internet_DialUp Internet_BroadBnd Internet_Cell Internet_WiFi Electronic_SelfHealthInfo Electronic_TalkDoctor Electronic_TestResults Electronic_MadeAppts Tablet_AchieveGoal Tablet_MakeDecision Tablet_DiscussionsHCP WillingShareData_HCP WillingShareData_YourFamily WillingShareData_YourFriends IntRsn_VisitedSocNet IntRsn_SharedSocNet IntRsn_SupportGroup IntRsn_YouTube SEER_ProbCare_BringTest SEER_ProbCare_WaitLong SEER_ProbCare_RedoTest SEER_ProbCare_ProvideHist HealthIns_InsuranceEmp HealthIns_InsurancePriv HealthIns_Medicare HealthIns_Medicaid HealthIns_Tricare HealthIns_VA HealthIns_IHS HealthIns_Other NotAccessed_SpeakDirectly NotAccessed_NoInternet NotAccessed_NoNeed NotAccessed_ConcernedPrivacy NotAccessed_NoRecord NotAccessed_LogInProb NotAccessed_Uncomfortable NotAccessed_MultipleRec RecordsOnline_ViewResults RecordsOnline_MessageHCP RecordsOnline_DownloadHealth ESent_AnotherHCP ESent_Family ESent_HealthApp CancerCond_CogImpair CancerCond_Neuropathy CancerCond_Fatigue CancerCond_Nausea CancerCond_Other MedConditions_Diabetes MedConditions_HighBP MedConditions_HeartCondition MedConditions_LungDisease MedConditions_Depression	Recoding missing responses for items with forced-choice response formats	Respondents were asked to select 'yes' or 'no' to a series of sub-items, allowing them to select as many responses as would apply. These 'forced-choice' response formats sometimes result in respondents indicating which sub-items apply to them by selecting the 'yes' response option for some and leaving the others unanswered. To allow the data to reflect this practice, if respondents did check one or more 'yes' response options within the group, but did not check a 'no' response option for any sub-item in the question, the sub-items that were missing a response were set to 'no.' However, if a respondent, in addition to leaving other sub-items unanswered, did select a 'no' response option for at least one sub-item, the unanswered sub-items were not assumed to be 'no' responses and instead remained missing.

Variable	Editing rule	Description of rule
TabletHealthWellnessApps FreqGoProvider AccessOnlineRecord ClinicalTrialCancerTx2 TimesModerateExercise	Recoding filter/skip questions	For these filter questions (questions containing a skip instruction associated with the particular response that was selected), response patterns following the question were examined if the filter question was not answered. The value representing the skip response was substituted for the missing filter question if all of the subsequent questions that the response directed the respondent to skip were left unanswered, and the respondent answered the next applicable question. However, missing values were not substituted with other values if the filter question was not answered but a follow-up question was answered.

Variable	Editing rule	Description of rule
Height_Feet Height_Inches	Edits for implausible values	The rules that were applied minimized the number of out-of-range values by accounting for response measurements in incorrect boxes, responses using metric, responses using only one unit of measurement and other response errors. Rules Applied to Edit Height Variables: If HEIGHT_Feet was 0 or missing and HEIGHT_Inches>48 and HEIGHT_Inches<=60, then the first digit was taken as the feet value and the second digit was taken as the inches value (to correct for respondents expressing both feet and inches in the inches box). If HEIGHT_Feet was 0 or missing and HEIGHT_Inches>61 and HEIGHT_Inches<=83, then the inches value was converted to its feet-and-inches equivalent (to correct for respondents expressing height in inches, resulting in heights from 5'1" to 6'11"). If HEIGHT_Feet was 1 and HEIGHT_Inches>=3 and HEIGHT_Inches<=9 (or HEIGHT_Inches>=30 and HEIGHT_Inches<=90) then this metric value was converted to feet-and-inches (to correct for respondents using meters and tenths and hundredths of a meter to express height). If HEIGHT_Feet>3 and HEIGHT_Feet<7 and HEIGHT_Inches = 20, 30, etc. thru 90 then the trailing 0 was removed. If HEIGHT_Feet>3 and HEIGHT_Feet<7 and HEIGHT_Inches = 15, 25, etc. thru 95 then the trailing 5 was removed (to correct for respondents expressing values in tenths of an inch). If HEIGHT_Feet>3 and HEIGHT_Feet<7 and HEIGHT_Inches = 12, 23, 34, 45 etc. thru 89 then the first digit was taken (to correct for respondents giving an inch value as a range, e.g., 1-2 or 8-9 inches). If HEIGHT_Feet>3 and HEIGHT_Feet<7 and HEIGHT_Inches = a two digit value whereby the first digit equaled the feet value the second digit was taken as the inches value (to correct for respondents expressing the height in inches as well as in feet, e.g., 5'58" resulted in value 5'8") If HEIGHT_Feet>6 and HEIGHT_Feet<12 and HEIGHT_Inches>3 and HEIGHT_Inches<7, then the values were switched (to correct for respondents putting measurements in the wrong boxes, resulting in edited values from 4'7" to <7 feet). If none of the preceding height editing rules were applicable: <u>Height_Feet (Height in Feet):</u> Any responses greater than 7 feet were recoded to "-4", which is the code for non-conforming responses.

Variable	Editing rule	Description of rule
Height_Feet Height_Inches	Edits for implausible values	Height_Inches (Height in Inches): Any responses greater than 11 inches were recoded to "-4", which is the code for non-conforming responses.
HaveDevice_Cat SEER_Cancer_Cat COVIDCa_Cat COVIDRoutine_Cat HeardGenTest_Cat TestSource_Cat HadTest2_Cat CaTest_Cat UndGenTest2_Cat SharedRes3_Cat Occupation_Cat Hisp_Cat Race_Cat2	Summarized distribution of 'mark all that apply' responses	A variable was created to indicate each response selection a respondent made for these 'mark all that apply' variables. The derived variable with the suffix '_cat' summarized the response selected or indicated that multiple responses were selected.
Education IncomeRanges	Edits for multiple responses	The highest order (e.g., education level or income range) was taken when multiple responses were selected.
QDisp	Derived variable	A variable was created to indicate the proportion of items respondents answered in the first two sections. This was used to determine incompletely-filled out questionnaires.
FullTimeOcc_Cat	Derived variable	A variable was created which combines the responses to P4 and P5, which aims to give a more comprehensive idea of a respondent's full time occupation.
WhenDiagnosedCancer Weight DrinkDaysPerWeek AverageTimeSitting Age SexualOrientation_OS	Recoding out of range responses	WhenDiagnosedCancer (Age at Time of Cancer Diagnosis): Any responses greater than the age of the respondent were recoded to "-4", which is the code for non-conforming responses. Weight: Any responses less than 40 pounds or greater than 500 pounds were recoded to "-4", which is the code for non-conforming responses. DrinkDaysPerWeek Any responses greater than 7 days per week were recoded to "-4", which is the code for non-conforming responses. AverageTimeSitting Any responses greater than 24 hours were recoded to "-4", which is the code for non-conforming responses. Age Responses were examined for out of range or unlikely ages (those listing their age as < 18 and > 105).

Variable	Editing rule	Description of rule
WhenDiagnosedCancer Weight DrinkDaysPerWeek AverageTimeSitting Age SexualOrientation_OS	Recoding out of range responses	<u>SexualOrientation_OS</u> Review of verbatim responses – Responses of "none of your business" and other similar phraseology were reviewed for scanning accuracy and recoded to "-4", which is the code for nonconforming responses.
HaveDevice_CellPh HaveDevice_None HeardGenTest_None TestSource_NotHeard HadTest2_None CaTest_NotHad	Recoding filter/skip questions	For these "mark all that apply" filter questions ("mark all that apply" type questions where one or more response option contains a skip instruction at the "No" or "None" response), when the "No" or "None" response was selected, all responses within the question group were examined. If other responses were checked, the "No" or "None" response was recoded to "Not selected", and the other responses were retained.
COVIDCa_NoAppts COVIDRoutine_NoAppts SharedRes3_NotShared UndGenTest2_NoOne NotHisp	Recoding illogical response combinations	For these "mark all that apply" questions ("mark all that apply" type questions where one or more response options do not contain a skip instruction at the "No" or "None" response, but keeping a "No" or "None" response in combination with other responses does not make logical sense), when the "No" or "None" response was selected, all responses within the question group were examined. If other responses were checked, the "No" or "None" response was recoded to "Not selected", and the other responses were retained.

Deriving the FullTimeOcc_Cat Variable

Fulltimeocc_cat combines responses to P4 (WorkFullTime) and P5 (Occupation_Cat) in to a single indicator of occupation status with the response options listed below.

Respondents are assigned to the category they selected in P5 which appears highest in the list below. For participants who chose 'Employed' for P5, their answer to P4 is used to determine whether they are coded as 'Employed full time' or 'Employed part time.' In some instances participants open-ended response to the P5 'Other' category were used to re-categorize them in to a different category than the highest one selected on the list. Respondents who mentioned a COVID-19 related work disruption were assigned to the 'Other' category. Participants who chose both 'Employed' and an Unemployed category in P5 were coded as 'Illogical response combination.'

Category	Value
P4 or P5 are missing	-9
Illogical response combination	-4
Employed full time	1
Employed part time	2
Homemaker	3
Student	4
Retired	5
Disabled	6
Unemployed less than 1 year	7
Unemployed 1 year or more	8
Other	9

Appendix H

Cancer Site Group Recode

Appendix H
Cancer Site Group Recode

Recode value	Label for Registry_Cancer_ Site group	Included Registry_Site codes	Registry_Histology inclusions and exclusions
1	Lip – Excludes: Skin of Lip	C000-C006, C008, C009	Excludes: 9050-9055, 9140, 9590-9993
2	Anterior Tongue	C020-C023, C028-C029	Excludes: 9050-9055, 9140, 9590-9993
3	Gum and Other Mouth	C030-C031, C039-C041, C048-C049, C060-C062, C068-C069	Excludes: 9050-9055, 9140, 9590-9993
4	Palate – Excludes: Soft and Uvula	C050, C058-C059	Excludes: 9050-9055, 9140, 9590-9993
5	Oropharynx – including Base of Tongue and Tonsils	C019, C024, C051-C052, C090-C091, C098-C100, C102-C104, C108-C109	Excludes: 9050-9055, 9140, 9590-9993
6	Other Pharynx and Other Oral Cavity	C140, C142, C148	Excludes: 9050-9055, 9140, 9590-9993
7	Nasal Cavity and Paranasal Sinuses	C300, C310-C311	Excludes: 9050-9055, 9140, 9590-9993
8	Other Sinuses	C301, C312-C313, C318-C319	Excludes: 9050-9055, 9140, 9590-9993
9	Nasopharynx	C110-C113, C118-C119	Excludes: 9050-9055, 9140, 9590-9993
10	Hypopharynx	C129-C132, C138-C139	Excludes: 9050-9055, 9140, 9590-9993
11	Major Salivary Glands	C079-C081, C088-C089	Excludes: 9050-9055, 9140, 9590-9993
12	Larynx	C101, C320-C323, C328-C329	Excludes: 9050-9055, 9140, 9590-9993
13	Trachea	C339	Excludes: 9050-9055, 9140, 9590-9993
14	Esophagus	C150-C155, C158-C159	Excludes: 9050-9055, 9140, 9590-9993
15	Stomach	C160-C166, C168-C169	Excludes: 9050-9055, 9140, 9590-9993
16	Small Intestine	C170-C173, C178-C179	Excludes: 9050-9055, 9140, 9590-9993
17	Appendix	C181	Excludes: 9050-9055, 9140, 9590-9993
18	Colon and Rectum	C180, C182-C189, C199, C209	Excludes: 9050-9055, 9140, 9590-9993
19	Anus and Anal Canal	C210-C212, C218	Excludes: 9050-9055, 9140, 9590-9993
20	Liver	C220	Excludes: 9050-9055, 9140, 9590-9993
21	Intrahepatic Bile Ducts	C221	Excludes: 9050-9055, 9140, 9590-9993
22	Gallbladder	C239	Excludes: 9050-9055, 9140, 9590-9993

Recode value	Label for Registry_Cancer_ Site group	Included Registry_Site codes	Registry_Histology inclusions and exclusions
23	Extrahepatic Bile Ducts	C240	Excludes: 9050-9055, 9140, 9590-9993
24	Ampulla of Vater	C241	Excludes: 9050-9055, 9140, 9590-9993
25	Other Biliary Tract	C248-C249	Excludes: 9050-9055, 9140, 9590-9993
26	Exocrine Pancreas	C250-C253, C257-C259	Excludes: 9050-9055, 9140, 9590-9993
27	Digestive tube Other	C260, C268-C269	Excludes: 9050-9055, 9140, 9590-9993
28	Lower: Bronchus and Lung	C340-C343, C348-C349	Excludes: 9050-9055, 9140, 9590-9993
29	Thymus	C379	Excludes: 9050-9055, 9140, 9590-9993
30	Heart	C380	Excludes: 9050-9055, 9140, 9590-9993
31	Mediastinum Other	C381-C383, C388	Excludes: 9050-9055, 9140, 9590-9993
32	Pleura	C384	Excludes: 9050-9055, 9140, 9590-9993
33	Other and Overlapping	C390, C398-C399	Excludes: 9050-9055, 9140, 9590-9993
34	Long Bones and Other Bones/Jts	C400-C403, C408-C414, C418-C419	Excludes: 9050-9055, 9140, 9590-9731, 9733-9993 Including 9732
35	Retroperitoneum	C480	Excludes: 9050-9055, 9140, 9590-9993
36	Peritoneum	C481, C482	Excludes: 9050-9055, 9140, 9590-9993
37	Overlapping Peritoneal/Retroperitoneal	C488	Excludes: 9050-9055, 9140, 9590-9993
38	Connective, Subcutaneous, and other Soft Tissue	C490-C496, C498-C499	Excludes: 9050-9055, 9140, 9590-9993
39	Cutaneous Head and Neck Sites	C440-C444	Includes: 8720-8790
41	Cutaneous Trunk and Upper Extremities	C445-C446	Includes: 8720-8790
41	Cutaneous Hip and Lower Extremities	C447	Includes: 8720-8790
42	Melanoma – Overlapping and Other Cutaneous	C448-C449	Includes: 8720-8790
43	Cutaneous Head and Neck Sites	C440-C444	Excludes: 8720-8790, 9050-9055, 9140, 9590-9993
44	Cutaneous Trunk and Upper Extremities	C445-C446	Excludes: 8720-8790, 9050-9055, 9140, 9590-9993
45	Cutaneous Hip and Lower Extremities	C447	Excludes: 8720-8790, 9050-9055, 9140, 9590-9993
46	Non-Epithelial Skin – Overlapping and Other Cutaneous	C448-C449	Excludes: 8720-8790, 9050-9055, 9140, 9590-9993
47	Nipple	C500	Excludes: 9050-9055, 9140, 9590-9993

Recode value	Label for Registry_Cancer_ Site group	Included Registry_Site codes	Registry_Histology inclusions and exclusions
48	All Other Breast	C501-C506, C508-C509	Excludes: 9050-9055, 9140, 9590-9993
49	Vulva	C510-C512, C518-C519	Excludes: 9050-9055, 9140, 9590-9993
50	Vagina	C529	Excludes: 9050-9055, 9140, 9590-9993
51	Cervix	C530-C531, C538-C539	Excludes: 9050-9055, 9140, 9590-9993
52	Uterine Corpus/Endometrial	C540-C543, C548-C549, C559	Excludes: 9050-9055, 9140, 9590-9993
53	Fallopian Tube	C570	Excludes: 9050-9055, 9140, 9590-9993
54	Ovary	C569	Excludes: 9050-9055, 9140, 9590-9993
55	Placenta	C589	Excludes: 9050-9055, 9140, 9590-9993
56	Other Adnexa	C571-C574, C577-C579	Excludes: 9050-9055, 9140, 9590-9993
57	Penis	C600-C602, C608-C609	Excludes: 9050-9055, 9140, 9590-9993
58	Prostate	C619	Excludes: 9050-9055, 9140, 9590-9993
59	Testis	C620-C621, C629	Excludes: 9050-9055, 9140, 9590-9993
60	Other Male Reproductive	C630-C632, C637-C639	Excludes: 9050-9055, 9140, 9590-9993
61	Kidney	C649	Excludes: 9050-9055, 9140, 9590-9993
62	Renal Pelvis	C659	Excludes: 9050-9055, 9140, 9590-9993
63	Ureter	C669	Excludes: 9050-9055, 9140, 9590-9993
64	Bladder	C670-C679	Excludes: 9050-9055, 9140, 9590-9993
65	Urethra	C680	Excludes: 9050-9055, 9140, 9590-9993
66	Urinary Other	C681, C688-C689	Excludes: 9050-9055, 9140, 9590-9993
67	Retina	C692	Excludes: 9050-9055, 9140, 9590-9993
69	Lacrimal Gland	C695	
69	Other Eye	C690-C691, C693-C694, C698-C699	Excludes: 9050-9055, 9140, 9590-9993
70	Orbit	C696	Excludes: 9050-9055, 9140, 9590-9993
71	Meninges	C700-C701, C709	Excludes: 9050-9055, 9140, 9590-9993
75	Brain (Malignant)	C710-C719, C728-C729	Excludes: 9050-9055, 9140, 9590-9993
76	Spinal Cord (Malignant)	C720-C721	Excludes: 9050-9055, 9140, 9590-9993
77	Cranial Nerves (Malignant)	C722-C725	Excludes: 9050-9055, 9140, 9590-9993

Recode value	Label for Registry_Cancer_ Site group	Included Registry_Site codes	Registry_Histology inclusions and exclusions
78	Thyroid	C739	Excludes: 9050-9055, 9140, 9590-9993
79	Adrenal Gland	C740-C741, C749	Excludes: 9050-9055, 9140, 9590-9993
80	Parathyroid	C750	Excludes: 9050-9055, 9140, 9590-9993
84	Pituitary Gland-Malignant	C751	Excludes: 9050-9055, 9140, 9590-9993
85	Craniopharyngeal Duct-Malignant	C752	Excludes: 9050-9055, 9140, 9590-9993
86	Pineal Gland-Malignant	C753	Excludes: 9050-9055, 9140, 9590-9993
87	Carotid Body, Paraganglia	C754	Excludes: 9050-9055, 9140, 9590-9993
88	Other and Overlapping Endocrine	C758-C759	Excludes: 9050-9055, 9140, 9590-9993
89	Endocrine Pancreas	C254	Excludes: 9050-9055, 9140, 9590-9993
90	Hodgkin Lymphomas	C024, C098-C099, C142, C379, C422-C424, C770-C775, C778-C779	Includes: 9650-9653, 9654-9656*, 9659, 9661-9662*, 9663, 9664-9665*, 9667*
91	Non-Hodgkin Lymphomas (Nodal and Extranodal)	C024, C098-C099, C142, C379, C422-C424, C770-C775, C778-C779 + other solid organ sites	Includes: 9590-9597, 9670-9671, 9673, 9675, 9678-9680, 9684, 9687-9691, 9695, 9698-9702, 9705, 9708-9709, 9712, 9714-9719, 9726, 9728-9729, 9735, 9737-9738, 9811-9819, 9823, 9826, 9835-9837
92	Acute Lymphocytic Leukemias	C421	Includes: 9811-9819, 9835-9837
93	Chronic Lymphocytic Leukemia	C421	Includes: 9823
94	Other Lymphocytic Leukemias	C421	Includes: 9591, 9670, 9820, 9832-9834, 9940
95	Acute Myelogenous Leukemias	C421	Includes: 9840, 9861, 9865-9867, 9869-9874, 9877-9879, 9891, 9895-9898, 9910-9912, 9920, 9931
96	Chronic Myelogenous Leukemias	C421	Includes: 9863, 9875-9876, 9945-9946
97	Other Myelogenous Leukemias	C421	Includes: 9860, 9930, 9950, 9961, 9975, 9983, 9989

Recode value	Label for Registry_Cancer_ Site group	Included Registry_Site codes	Registry_Histology inclusions and exclusions
98	Plasmacytomas	C019, C024, C051-C052, C079, C090-C091, C098-C104, C108-C113, C118-C119, C140, C142, C148, C300-C301, C310-C313, C318-C323, C328-C329, C400-C403, C408-C414, C418-C419, C440-C449, C501-C506, C508-C509, C620-C621, C629, C670-C679, C710-C725, C728-C729, C739, C770-C779	Includes: 9731, 9734
99	Myeloma	C421	Includes: 9732
100	Waldenstrom Macroglobulinemia (Non-Hodgkin Lymphoma)	C420	Includes: 9761
101	Kaposi Sarcoma	All Primary Sites	Includes: 9140
102	Non-Hematopoietic	C760-C765, C767-C768	Excludes: 9050-9055, 9140, 9590-9993
103	Cancer of Unknown Primary	C809	

*Obsolete in recent years, but valid histologic type in past.

Appendix I

Cancer Site Organ System Recodes

Appendix I
Cancer Site Organ System Recodes

Recode value	Variable label for Registry_Cancer_Site_Organsys	Registry_Cancer_Site codes included	Registry_Histology codes excluded and included
1	Head & Neck (for example: Lips, Tongue, Other Mouth, Pharyngeal, Other Nasal Cavities)	C000-C006, C008-C009, C019-C024, C028-C031, C039-C041, C048-C052, C058-C062, C068-C069, C079-C081, C088-C091, C098-C104, C108-C113, C118-C119, C129-C132, C138-C140, C142, C148, C300-C301, C310-C313, C318-C323, C328-C329, C339	Excluding 9050-9055, 9140, 9590-9993
2	GI Tube (for example: Esophagus, Stomach, Intestine, Colon, Rectum, Anus)	C150-C155, C158-C166, C168-C170-C173, C178-C189, C199, C209-C212, C218, C260, C268-C269	Excluding 9050-9055, 9140, 9590-9993
3	GI Solid Organs (for example: Liver, Pancreas)	C220, C250-C253, C257-C259	Excluding 9050-9055, 9140, 9590-9993
4	Biliary Tract (for example: Bile Ducts, Gallbladder)	C221, C239-C241, C248-C249	Excluding 9050-9055, 9140, 9590-9993
5	Respiratory System (for example: Bronchus, Lung)	C340-C343, C348-C349, C390, C398-C399	Excluding 9050-9055, 9140, 9590-9993
6	Mediastinum (for example: Thymus, Pleura)	C379-C384, C388	Excluding 9050-9055, 9140, 9590-9993
7	Bone & Soft Tissue (for example: Peritoneum, Subcutaneous tissue)	C400-C403, C408-C414, C418-C419, C480-C482, C488, C490-C496, C498-C499	Excluding 9050-9055, 9140, 9590-9731, 9733-9993
8	Skin-Melanoma	C440-C449	Includes: 8720-8790
9	Skin-Non-Epithelial	C440-C449	Excludes: 8720-8790, 9050-9055, 9140, 9590-9993
10	Breast	C500-C506, C508, C509	Excluding 9050-9055, 9140, 9590-9993
11	Female Reproductive System (for example: Vagina, Cervix, Uterus)	C510-C512, C518-C519, C529-C531, C538-C543, C548-C549, C559, C569-C574, C577-C579, C589	Excluding 9050-9055, 9140, 9590-9993
12	Male Reproductive System (for example: Penis, Prostrate, Testis)	C600-C602, C608-C609, C619-C621, C629-C632, C637-C639	Excluding 9050-9055, 9140, 9590-9993
13	Urinary Tract (for example: Kidney, Bladder)	C649, C659, C669-C681, C688-C689	Excluding 9050-9055, 9140, 9590-9993
14	Eye	C690-C696, C698-C699	Excluding 9050-9055, 9140, 9590-9993
15	Central Nervous System & Meninges (for example: Brain, Spinal Cord, Crainal Nerves)	C700-C701, C709-C725, C728-C729	Excluding 9050-9055, 9140, 9590-9993

Recode value	Variable label for Registry_Cancer_Site_Organsys	Registry_Cancer_Site codes included	Registry_Histology codes excluded and included
16	Endocrine	C254, C739-C741, C749-C754, C758-C759	Excluding 9050-9055, 9140, 9590-9993
17	Lymphocytic Lymphomas & Lymphoblastic Leukemias	C024, C098-C099, C142, C379, C420-C424, C770-C775, C778-C779 + other solid organ sites	Includes: 9590-9597, 9650-9653, 9654*, 9655*, 9656*, 9659, 9661-9662*, 9663, 9664-9665*, 9667*, 9670-9671, 9673, 9675, 9678-9680, 9684, 9687-9691, 9695, 9698-9702, 9705, 9708-9709, 9712, 9714-9719, 9726, 9728-9729, 9735, 9737-9738, 9811-9819, 9820, 9823, 9826, 9832-9837, 9940
18	Myelodysplastic/Myeloproliferative Neoplasms & Myeloid Leukemias	C421	Includes: 9840, 9860-9861, 9863, 9865-9867, 9869-9879, 9891, 9895-9898, 9910-9912, 9920, 9930-9931, 9945-9946, 9950, 9961, 9975, 9983, 9989
19	Myeloma & Plasma Cell Disorders	C019, C024, C051-C052, C079, C090-C091, C098-C104, C108-C113, C118-C119, C140, C142, C148, C300-C301, C310-C313, C318-C323, C328-C329, C400-C403, C408-C414, C418-C419, C421, C440-C449, C501-C506, C508-C509, C620-C621, C629, C670-C679, C710-C725, C728-C729, C739, C770-C779,	Includes: 9731, 9732, 9734
20	Kaposi Sarcoma	All Primary Sites	Includes: 9140
21	Ill-Defined	C760-C765, C767, C768	Excluding 9050-9055, 9140, 9590-9993
22	Cancer of Unknown Primary	C809	

*Obsolete in recent years, but valid histologic type in past.

Appendix J

Standard Cancer Site Recode

Appendix J
Standard Cancer Site Recode

Recode value	Definition for Registry_Cancer_Site_StdRecode	Included Registry_Site Codes	Registry_Histology inclusions and exclusions
20010	Lip	C000-C006, C008, C009	Excludes: 9050-9055, 9140, 9590-9993
20020	Tongue	C019-C024, C028, C029	Excludes: 9050-9055, 9140, 9590-9993
20030	Salivary Gland	C079, C080, C081, C088, C089	Excludes: 9050-9055, 9140, 9590-9993
20040	Floor of Mouth	C040, C041, C048, C049	Excludes: 9050-9055, 9140, 9590-9993
20050	Gum & Other Mouth	C030, C031, C039, C050-C052, C058-C062, C068, C069	Excludes: 9050-9055, 9140, 9590-9993
20060	Nasopharynx	C110-C113, C118, C119	Excludes: 9050-9055, 9140, 9590-9993
20070	Tonsil	C090, C091, C098, C099	Excludes: 9050-9055, 9140, 9590-9993
20080	Oropharynx	C100-C104, C108, C109	Excludes: 9050-9055, 9140, 9590-9993
20090	Hypopharynx	C129, C130-C132, C138, C139	Excludes: 9050-9055, 9140, 9590-9993
20100	Other Oral Cavity & Pharynx	C140, C142, C148	Excludes: 9050-9055, 9140, 9590-9993
21010	Esophagus	C150-C155, C158, C159	Excludes: 9050-9055, 9140, 9590-9993
21020	Stomach	C160-C166, C168, C169	Excludes: 9050-9055, 9140, 9590-9993
21030	Small Intestine	C170-C173, C178, C179	Excludes: 9050-9055, 9140, 9590-9993
21041	Cecum	C180	Excludes: 9050-9055, 9140, 9590-9993
21042	Appendix	C181	Excludes: 9050-9055, 9140, 9590-9993
21043	Ascending Colon	C182	Excludes: 9050-9055, 9140, 9590-9993
21044	Hepatic Flexure	C183	Excludes: 9050-9055, 9140, 9590-9993
21045	Transverse Colon	C184	Excludes: 9050-9055, 9140, 9590-9993
21046	Splenic Flexure	C185	Excludes: 9050-9055, 9140, 9590-9993
21047	Descending Colon	C186	Excludes: 9050-9055, 9140, 9590-9993
21048	Sigmoid Colon	C187	Excludes: 9050-9055, 9140, 9590-9993
21049	Large Intestine, NOS	C188-C189, C260	Excludes: 9050-9055, 9140, 9590-9993
21051	Rectosigmoid Junction	C199	Excludes: 9050-9055, 9140, 9590-9993

Recode value	Definition for Registry_Cancer_Site_StdRecode	Included Registry_Site Codes	Registry_Histology inclusions and exclusions
21052	Rectum	C209	Excludes: 9050-9055, 9140, 9590-9993
21060	Anus, Anal Canal, and Anorectum	C210-C212, C218	Excludes: 9050-9055, 9140, 9590-9993
21071	Liver	C220	Excludes: 9050-9055, 9140, 9590-9993
21072	Intrahepatic Bile Duct	C221	Excludes: 9050-9055, 9140, 9590-9993
21080	Gallbladder	C239	Excludes: 9050-9055, 9140, 9590-9993
21090	Other Biliary	C240-C249	Excludes: 9050-9055, 9140, 9590-9993
21100	Pancreas	C250-C259	Excludes: 9050-9055, 9140, 9590-9993
21110	Retroperitoneum	C480	Excludes: 9050-9055, 9140, 9590-9993
21120	Peritoneum, Omentum, and Mesentery	C481-C482	Excludes: 9050-9055, 9140, 9590-9993
21130	Other Digestive Organs	C260, C268-C269, C488	Excludes: 9050-9055, 9140, 9590-9993
22010	Nose, Nasal Cavity, and Middle Ear	C300-C301, C310-C319	Excludes: 9050-9055, 9140, 9590-9993
22020	Larynx	C320-C329	Excludes: 9050-9055, 9140, 9590-9993
22030	Lung and Bronchus	C340-C349	Excludes: 9050-9055, 9140, 9590-9993
22050	Pleura	C384	Excludes: 9050-9055, 9140, 9590-9993
22060	Trachea, Mediastinum, and Other Respiratory Organs	C339, C381-C383, C388, C390, C398-C399	Excludes: 9050-9055, 9140, 9590-9993
23000	Bones and Joints	C400-C419	Excludes: 9050-9055, 9140, 9590-9731, 9733-9993 Includes: 9732
24000	Soft Tissue including Heart	C380, C490-C496, C498, C499	Excludes: 9050-9055, 9140, 9590-9993
25010	Melanoma of the Skin	C440-C449	Includes: 8720-8790
25020	Other Non-Epithelial Skin	C440-C449	Excludes: 8720-8790, 9050-9055, 9140, 9590-9993
26000	Breast	C500-C509	Excludes: 9050-9055, 9140, 9590-9993
27010	Cervix Uteri	C530, C531, C538, C539	Excludes: 9050-9055, 9140, 9590-9993
27020	Corpus Uteri	C540-C543, C548, C549	Excludes: 9050-9055, 9140, 9590-9993
27030	Uterus, NOS	C559	Excludes: 9050-9055, 9140, 9590-9993
27040	Ovary	C569	Excludes: 9050-9055, 9140, 9590-9993
27050	Vagina	C529	Excludes: 9050-9055, 9140, 9590-9993
27060	Vulva	C510-C512, C518, C519	Excludes: 9050-9055, 9140, 9590-9993

Recode value	Definition for Registry_Cancer_Site_StdRecode	Included Registry_Site Codes	Registry_Histology inclusions and exclusions
27070	Other Female Genital Organs	C570-C579, C589	Excludes: 9050-9055, 9140, 9590-9993
28010	**Prostate**	C619	Excludes: 9050-9055, 9140, 9590-9993
28020	**Testis**	C620, C621, C629	Excludes: 9050-9055, 9140, 9590-9993
28030	**Penis**	C600-C602, C608, C609	Excludes: 9050-9055, 9140, 9590-9993
28040	Other Male Genital Organs	C630-C632, C637-C639	Excludes: 9050-9055, 9140, 9590-9993
29010	Urinary Bladder	C670-C679	Excludes: 9050-9055, 9140, 9590-9993
29020	Kidney and Renal Pelvis	C649, C659	Excludes: 9050-9055, 9140, 9590-9993
29030	Ureter	C669	Excludes: 9050-9055, 9140, 9590-9993
29040	Other Urinary Organs	C680-C689	Excludes: 9050-9055, 9140, 9590-9993
30000	Eye and Orbit	C690-C699	Excludes: 9050-9055, 9140, 9590-9993
31010	Brain	C710-C719	Excludes: 9050-9055, 9140, 9590-9993
31040	Cranial Nerves Other Nervous System	C710-C719 C700-C709, C720-C729	Includes: 9530-9539 Excludes: 9050-9055, 9140, 9590-9993
32010	Thyroid	C739	Excludes: 9050-9055, 9140, 9590-9993
32020	Other Endocrine including Thymus	C379, C740-C749, C750-C759	Excludes: 9050-9055, 9140, 9590-9993
33011	Hodgkin Lymphoma – Nodal	C024, C098, C099, C142, C379, C422-C424, C770-C775, C778, C779	Includes: 9650-9653 9654-9656, 9659, 9661-9662, 9663, 9664-9665, 9667
33012	Hodgkin Lymphoma - Extranodal	All sites, except: C024, C098, C099, C142, C379, C422-C424, C770-C775, C778, C779	Includes: 9650-9653 9654-9656, 9659, 9661-9662, 9663, 9664-9665, 9667
33041	Non-Hodgkin Lymphoma - Nodal	C024, C098, C099, C142, C379, C422-C424, C770-C775, C778, C779 + other solid organ sites	Includes: 9590-9597, 9670-9671, 9673, 9675, 9678-9680, 9684, 9687-9691, 9695, 9698-9702, 9705, 9708-9709, 9712, 9714-9719, 9726, 9728-9729, 9735, 9737-9738, 9811-9819, 9823, 9826, 9835-9837
33042	Non-Hodgkin Lymphoma - Extranodal	All sites, except: C024, C098, C099, C142, C379, C422-C424, C770-C775, C778, C779	Includes: 9590-9597, 9670-9671, 9673, 9675, 9678-9680, 9684, 9687-9691, 9695, 9698-9702, 9705, 9708-9709, 9712, 9714-9719, 9726, 9728-9729, 9735, 9737-9738, 9761, 9811-9819, 9823, 9826-9827, 9835-9838
34000	Myeloma		Includes: 9731-9732, 9734

Recode value	Definition for Registry_Cancer_Site_StdRecode	Included Registry_Site Codes	Registry_Histology inclusions and exclusions
35011	Acute Lymphocytic Leukemia	C420, C421, C424	Includes: 9811-9819, 9826, 9835-9837
35012	Chronic Lymphocytic Leukemia	C420, C421, C424	Includes: 9823
35013	Other Lymphocytic Leukemia		Includes: 9591, 9670, 9820, 9832-9834, 9940
35021	Acute Myeloid Leukemia		Includes: 9840, 9861, 9865-9867, 9869-9874, 9877-9879, 9891, 9895-9897, 9898, 9910-9912, 9920, 9931
35022	Chronic Myeloid Leukemia		Includes: 9863, 9875-9876, 9945-9946
35023	Other Myeloid/Monocytic Leukemia		Includes: 9860, 9930, 9950, 9961, 9975, 9983, 9989
35041	Other Acute Leukemia		Includes: 9801, 9805-9809, 9931
35043	Aleukemic, subleukemic and NOS		Includes: 9733, 9742, 9800, 9827, 9831, 9870, 9948, 9963-9964
36010	Mesothelioma		Includes: 9050-9055
36020	Kaposi Sarcoma		Includes: 9140
37000	Miscellaneous	C760-C765, C767, C768, C809	Excludes: 9050-9055, 9140, 9590-9993