Title

Rationale, Procedures, and Response Rates for a Pilot Study to Sample Cancer Survivors for NCIs Health Information National Trends Survey: HINTS-SEER 2021.

Permalink

https://escholarship.org/uc/item/6c82v316

Journal

Journal of Health Communication, 29(2)

Authors

Blake, Kelly

Moser, Richard

Murray, Ashley

et al.

Publication Date

2024-02-01

DOI

10.1080/10810730.2023.2290550

Peer reviewed

Rationale, procedures, and response rates for a pilot study to sample cancer survivors for NCI's Health Information National Trends Survey: HINTS-SEER 2021

Kelly D. Blake, ScD[1], **Richard P. Moser, PhD**[2], **Ashley B. Murray, PhD**[2], **Terisa Davis, MPH**[3], **David Cantor, PhD**[3], **Andrew Caporaso, PhD**[3], **Michele West, PhD**[4], **Suzanne Bentler, PhD**[4], **Meg McKinley, MPH**[5], **Salma Shariff-Marco, PhD, MPH**[6], **Chuck Wiggins, PhD, MSPH**[7], **Robin C. Vanderpool, DrPH**[1]

[1.] KD Blake, RC Vanderpool: Health Communication and Informatics Research Branch, Behavioral Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD

[2.] RP Moser, A Murray: Behavioral Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD

[3.] T Davis, D Cantor, A Caporaso: Westat Inc., Rockville, MD

[4.] West M, Bentler S: Department of Epidemiology, University of Iowa College of Public Health, Iowa Cancer Registry, Iowa City, IA

[5.] McKinley M: Department of Epidemiology and Biostatistics, University of California San Francisco, Greater Bay Area Cancer Registry, San Francisco, CA

[6.] Shariff-Marco S: Department of Epidemiology and Biostatistics, University of California San Francisco, UCSF Helen Diller Family Comprehensive Cancer Center, Greater Bay Area Cancer Registry, San Francisco, CA

[7.] Wiggins C: Department of Internal Medicine, University of New Mexico, University of New Mexico Comprehensive Cancer Center, New Mexico Tumor Registry, Albuquerque, NM

Abstract

The National Cancer Institute's (NCI) Health Information National Trends Survey (HINTS) is a nationally representative survey of U.S. adults in which 12–17% of respondents report a cancer history. To increase representation from adult cancer survivors, in 2021, NCI sampled survivors from three Surveillance, Epidemiology, and End Results (SEER) program cancer registries: Iowa, New Mexico, and Greater Bay Area. Sampling frames were stratified by time since diagnosis and race/ethnicity, with non-malignant tumors and non-melanoma skin cancers excluded. Participants completed a self-administered postal questionnaire. The overall response rate for HINTS-SEER (N=1,234) was 12.6%; a non-response bias analysis indicated few demographic differences between respondents and the pool of sampled patients in each registry. Most of the sample

Corresponding Author: Kelly D. Blake, 9609 Medical Center Drive, Rm 3E532, MSC 9671, Bethesda, MD 20892-9671, (p) 240.281.5934, kelly.blake@nih.gov.

was 10+ years since diagnosis (n=722; 60.2%); 392 respondents were 5 to <10 years since diagnosis (29.6%); and 120 were <5 years since diagnosis (10.2%). Common cancers included male reproductive (n=304; 24.6%), female breast (n=284; 23.0%), melanoma (n=119; 9.6%), and gastrointestinal (n=106; 8.6%). Tumors were mostly localized (67.8%; n=833), with 22.4% (n=282) regional, 6.2% (n=72) distant, and 3.7% (n=47) unknown. HINTS-SEER data are available by request and may be used for secondary analyses to examine a range of social, behavioral, and healthcare outcomes among cancer survivors.

Keywords

cancer survivors; survey research methods; Health Information National Trends Survey; HINTS

INTRODUCTION

The National Cancer Institute (NCI) developed the Health Information National Trends Survey (HINTS) in 2003 to monitor population trends in cancer communication practices, information preferences, risk behaviors, and cancer knowledge. The HINTS program provides unique population-level data on changing communication patterns and information support needs, changing health communication and technology trends and practices, cancer information access and usage, and cancer risk perceptions. The purpose of conducting a nationally-representative, probability-based survey to assess health communication processes and needs has been to provide public health, behavioral, and communication scientists and practitioners with data to assess the prevalence of cancer-relevant knowledge, attitudes, information-seeking practices, and health behaviors in the U.S. adult population. HINTS has been fielded 16 times since 2003, surveying more than 60,000 U.S. adults. Details about HINTS methodology for the general population survey have been published elsewhere (Nelson, 2004; Hesse, 2006; Finney Rutten, 2012; Blake, 2016; Finney Rutten, 2020; Blake, 2022).

Rationale: HINTS-SEER

An abbreviated description of HINTS data collection efforts from 2003–2020, including the number and percent of cancer survivors represented in each HINTS dataset, can be found in Table 1. To date, cancer survivors make up approximately 14 percent of the total unweighted HINTS sample across administrations from 2003–2020, with 12–17 percent of respondents from each individual administration reporting a history of cancer.

Despite being limited by the small sample sizes of cancer survivors in prior individual HINTS administrations, HINTS data have been used to characterize the experiences of cancer survivors. In a 2019 profile of the HINTS program, Rutten et al. conducted a literature review to identify and describe survivor-specific studies that utilized HINTS from 2003–2018 (Rutten, 2020). Among the themes of the survivor-specific published articles were trends regarding information seeking experiences and sources; patient-centered communication and clinical care; use of internet and mobile technology for health; and health status and health outcomes. The paper also presented the sociodemographic and self-reported cancer diagnosis and treatment characteristics of the cross-sectional subsample

of survivors in prior HINTS administrations and discussed the need for a larger sample of cancer survivors to support disaggregated inferences by cancer type and time since diagnosis (Rutten, 2020). To address this need, NCI developed a pilot project to sample cancer survivors using selected cancer registries from the Surveillance, Epidemiology, and End Results (SEER) Program (https://seer.cancer.gov) as a sampling frame. 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 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.

MATERIALS AND METHODS

HINTS-SEER was planned as a way to oversample cancer survivors for HINTS 5 Cycle 4 (2020), which was given a non-human subjects research designation from the NIH Office of Human Subjects Research in April 2016 via exemption number 13204. HINTS-SEER was approved by the Westat Institutional Review Board in February 2020 (Project Number 6048.14), with a subsequent amendment approved in May 2020. The participating SEER registries obtained independent IRB approvals for providing case listings for the study.

To determine which SEER registries would participate in HINTS-SEER, NCI presented the project to the SEER Research Group, which facilitates research across SEER, in July 2019, after which Westat sent out a short survey to SEER registry principal investigators in the fall of 2019. Responses identified the registries that were interested in participating in HINTS-SEER, and provided initial information about each registry's procedures, including their participant 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, the research team initially selected four registries 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. The team also considered the availability of registry staff to participate, and the costs associated with sampling and consenting participants. Three SEER registries ultimately participated: Iowa Cancer Registry, New Mexico Tumor Registry, and Greater Bay Area Cancer Registry.

Sampling Procedures

Survivors with records meeting four eligibility criteria, identified by the research team and designed to enable broad participation, were included in the sampling frame for each participating SEER registry: cancer survivors with a vital status of alive at the time the samples were drawn; survivors with a cancer diagnosis specified to invasive cancers; survivors age 18 or older as of December 2020; and survivors for whom the last date of contact by the registry was no earlier than January 1, 2016 (the latest date available, included to reduce the likelihood of nonresponse due to incorrect mailing addresses). Ineligibility criteria included non-malignant tumors and survivors whose only diagnosis was non-melanoma skin cancer. Non-melanoma skin cancers were excluded in order to focus on

more invasive tumors, as 22 percent of the cancers self-reported in HINTS probability-based samples are non-melanoma skin cancer, and HINTS-SEER sought to identify survivors for whom information support needs about treatment and follow-up may be more significant than would be for non-melanoma skin cancer.

For cancer survivors that had more than one tumor, one eligible tumor was chosen for reporting using SEER cancer sequence numbers (CSEQ). The CSEQ indicates the order in which a reportable primary tumor is discovered in relation to the total number of primary tumors 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 the first of multiple tumors). For participants with multiple tumors, if the first (or second) cancer did not meet the eligibility criteria for the study (for example, a non-melanoma skin cancer), then the next tumor was selected, and its corresponding CSEQ is therefore higher (e.g., 02, 03). For the small percent of participants with multiple eligible tumors, one eligible tumor was chosen for reporting in the HINTS-SEER meta data. In the Iowa and New Mexico SEER registries, 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 wherein, for people with multiple eligible tumors, the tumor was selected *randomly* rather than systematically, which resulted in slightly more respondent cases having a sequence number higher than 01. However, the overall percent of cases with a first primary tumor (CSEQ 00 or 01) selected was high among respondents across the three registries (Iowa 98.5%, Greater Bay Area 90.2%, New Mexico, 96.5%).

For each registry, implicit stratification was used to generate a representative sample of its eligible survivors by years since diagnosis and race/ethnicity. Each participating registry undertook a unique consent process before survivors' mailing addresses were provided to Westat to be included in the sample for HINTS-SEER. The Iowa Cancer Registry used an active consent process wherein the registry contacted eligible participants in advance and obtained a signed release form before their contact information was shared with Westat. Internal policies required further exclusions, removing any cases that were non-Iowa residents at the time of diagnosis, survivors seen at a Veterans Affairs facility only, cases acquired from the state data exchange only, and survivors flagged as "do not contact." Based on the HINTS request for a sample of 447 Iowa participants to be included in the sample, the Iowa Cancer Registry determined that they needed to send out 6,433 consent forms based on a 7% anticipated agreement rate. In December 2020, the Iowa Cancer Registry delivered to Westat a file of 482 addresses of survivors who consented to be contacted to participate in HINTS-SEER.

The Greater Bay Area Cancer Registry did not pre-consent study participants to provide mailing addresses but expected respondents to be consented during the study procedures. As such, a consent form was included on the inside cover of the HINTS-SEER survey instrument for the Greater Bay Area cohort, and as required by the state of California's Committee for the Protection of Human Subjects, data from any respondent who responded to the survey without signing the consent form was discarded unless they could be re-consented. HINTS requested 923 cases from the Greater Bay Area Cancer Registry, and in

January 2021, they delivered to Westat a file of 2000 addresses of survivors, based on the maximum number that can be requested from the registry at a single point in time. The list excluded those with a "do not contact" flag and survivors seen at Veterans Affairs facilities only, as is standard practice for the Greater Bay Area Cancer Registry.

The New Mexico Tumor Registry's procedures included a passive consent process that required the registry to contact potential study participants and collect study refusals. The registry also employed internal policies that require further exclusions, removing those with only carcinoid tumors, those who identify as Native Americans, survivors seen at a Veterans Affairs facility only, survivors flagged as "do not contact" in the registry, and survivors with invalid addresses such as a correction center, nursing home, assisted living, hospice, or social service. 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 HINTS-SEER. Based on the HINTS request for a sample of 240 New Mexico participants, the New Mexico Tumor Registry selected the number of survivors for the consent mailing, 1400, based on their experience with similar requests wherein approximately 20% were ultimately accrued for studies. The consent mailing was sent out to 1056 addresses in January 2021 after further vetting for ineligibility. In February 2021, the New Mexico Tumor Registry delivered to Westat a file of 850 addresses of survivors who did not refuse to participate in the study and whose addresses were not deemed undeliverable.

Demographic distributions of survivors in each registry's sampling frame, including the CSEQ for primary tumors, can be found in the HINTS-SEER methodology report.

Instrument

The foundation for the HINTS-SEER instrument was the HINTS 5 Cycle 4 instrument that underwent cognitive testing in 2019 and was fielded using the standard HINTS probability-based methods in early 2020. Core constructs included health information seeking; technology and internet use; wearable and medical device use; social media use; health care access, utilization, and satisfaction; medical record/patient portal access and use; family history of cancer and personal cancer history; genetic testing; clinical trial awareness and utilization; health status; diet; physical activity; tobacco product use; and cancer screening, among others. Edits were made to the instrument both to make some of the items more appropriate for cancer survivors versus the general population, as well as to add topics of specific concern to those with 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 included: an item added about whether the respondent had spoken to a mental health professional (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) to assess COVID-19's impact on cancer treatment, follow-up cancer care, cancer screening, and preventive care, as well as questions about patient-provider discussions and trust in sources of COVID-19 information. In addition,

questions were added to determine respondents' 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 (items J4-J11); and a series of items to measure social isolation were added (items J12-J15). Most of the items added to the HINTS-SEER instrument had been cognitively tested and fielded on prior HINTS cycles, and others had only minor edits to their original wording. The COVID-19 items were pre-tested in an online survey that was being conducted as part of another NCI study. The questionnaire was administered in English only and the final HINTS-SEER instrument is available here: https://hints.cancer.gov/docs/Instruments/HINTS_SEER_AnnotatedInstrumentEnglish.pdf

Data Collection

Data collection for HINTS-SEER started on January 11, 2021 and concluded on August 20, 2021. The survey was administered exclusively by mail using an address-based frame with a $2 pre-paid monetary incentive to encourage participation. The mailing protocol for all three registry cohorts followed a modified Dillman approach (Dillman, 2009) with a total of four contacts: an initial mailing, a reminder postcard, and two follow-up mailings. All sampled individuals received the first mailing and reminder postcard, while only non-respondents received the subsequent survey mailings. Respondents from the Greater Bay Area Cancer Registry who completed and returned the survey but did not sign the consent form (n=102) were sent an additional mailing containing the consent form and a postage-paid return envelope, in an attempt to consent them. Of those, 51.96% (n=53) returned the signed consent form and were included in the study.

A toll-free telephone number was provided in all mailings for participants who had questions or concerns about the study. The toll-free line received 26 calls throughout the HINTS-SEER field period. Most calls were refusals or calls to let the study team know that the recipient addressed in the mailings had passed away or was incapacitated.

In 19 returned surveys, the respondent reported that they had never been diagnosed as having cancer. These cases were brought to the registries' attention and registry staff determined that although these individuals were diagnosed with cancer, the tumors were often a less aggressive or early stage of cancer, and the patient may not have been aware of the diagnosis. Regardless, these 19 surveys were treated as ineligible and removed from the sample. Five questionnaires (0.4%) were discarded because they did not meet the survey completion criteria (i.e., answering at least 50 percent of the required questions in Sections A and B). Item non-response rates varied by survey item and were generally below 5%. These survey completion and item non-response rates were consistent with prior HINTS general population surveys.

Imputation

Five questions were imputed because several respondents incorrectly selected more than one response when just one response was required. The imputed responses were based on the distribution of answers among those who provided a valid (single-answer) response to each question. If a respondent selected two responses, for example, where their first response comprised 40 percent of the valid 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 imputed items are: *If you had a strong need to get information about cancer, where would you go first?* (SEERStrongNeedCancerInfo_IMP); *Who ordered your cancer genetic test for inherited cancer syndromes?* (WhoOrderedCATest_IMP); *If you had a need to get information about clinical trials, which of the following would you go to* first? (FirstInfoClinTrials2_IMP); *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?* (TrustInfoClinTrials2_IMP); and *From the set of values below, which ONE is most important to you in your day-to-day life?* (MostImportantValues_IMP).

Weighting and Variance Estimation

Weighting was conducted using demographic control totals reflecting the eligible population within each of the three SEER registries. Although not nationally representative, the HINTS-SEER samples are probability samples from the frames from which they were drawn, and weights for each registry scale up to the registry population, thereby making the samples representative of the respective registries. As described later in this paper, comparing HINTS-SEER estimates to other HINTS data collections should be done with caution as other HINTS data collections are weighted to represent the full U.S. population.

Every sampled cancer survivor who completed a questionnaire for HINTS-SEER received a full-sample weight and a set of 50 replicate weights. The full-sample weight is used to calculate population and subpopulation estimates and the replicate weights are used to compute accurate standard errors for these estimates. Replicate weights were calculated using the 'delete one' jackknife (JK1) replication method. 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 previously.

Although replication is the recommended method for variance estimation for HINTS-SEER, not all software packages have a replication option to produce variance estimates. To accommodate SPSS users or other data users who would like to produce variance estimates using Taylor Series methods, a variance-estimation stratum variable (VarStratum) is provided in the dataset, as is a VarCluster variable that represents the primary sampling unit or cluster within the variance-estimation stratum.

SEER Meta Data

For each survivor sampled for HINTS-SEER, the following meta data about each selected tumor are included in the dataset as both raw and recoded variables (where applicable):

- Primary cancer site (Registry_Cancer_Site, Registry_Cancer_Site_StdRecode, Registry_Cancer_Site_Group, Registry_Cancer_Site_OrganSys): where the cancer was located in the body;

- Cancer histology (Registry_Histology, Registry_Histology_Recode): the type of tissue from which the cancer originated;

- SEER summary stage (SEER_Summary_Stage): tumor stage at diagnosis; and

- Date of diagnosis (Registry_Year_Of_Diagnosis): the date that the cancer survivor was diagnosed based on the selected tumor.

The CSEQ variable that was used as part of the registries' sampling procedures is not provided on the HINTS-SEER dataset. As noted above, for people with multiple eligible tumors, the Iowa and New Mexico registries selected the first eligible primary tumor while the Greater Bay Area Cancer Registry used random selection of eligible primary tumors. Across all three registry cohorts, approximately 95% of survey respondents' tumor information is from a first primary cancer diagnosis. Further details about respondents' CSEQ variables are provided in the HINTS-SEER Methodology Report in section 4.6.

RESULTS

Sample Size and Response Rate

Although each registry population was stratified by years since diagnosis and race/ethnicity, all survivors were systematically sampled with the same selection probability and therefore given the same base weight (i.e., the response rates are not weighted). In addition, because the participating SEER registries had different consent procedures for addresses to be included in the HINTS-SEER sample, the response rate was calculated in two stages. The first stage accounts for the consent rate and is calculated as the proportion of sampled registrants who consented to participate. The second stage accounts for the survey completion rate and is computed as the proportion of those who consented who returned a complete survey. The overall response rate is the product of the consent and completion rates (Stage 1 * Stage 2).

The final overall response rate for HINTS-SEER was 12.6 percent, with a sample of 1,234 respondents. Sample characteristics for the full sample and by each individual SEER registry can be found in Table 2. The Greater Bay Area Cancer Registry and New Mexico Tumor Registry achieved similar overall response rates of 24.1 and 24.6 percent, respectively. The Iowa Cancer Registry 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 (Sakshaug et al, 2016).

Non-Response Bias Analysis

To evaluate whether the HINTS-SEER respondents were different from non-respondents, we compared their basic demographic composition using aggregated data available from their registry. Generally, HINTS-SEER respondents were not substantially different from the overall sample of survivors from which they were sampled. For most of the demographic comparisons, the difference in proportions was five percentage points or less. In Iowa, there was a larger proportion of individuals aged 60 or older in the respondents (84%) than in

the overall sample (78%), and there was a smaller proportion of respondents who were diagnosed with cancer prior to 2010 (47% compared to 54%). In the Greater Bay Area, the respondents included a larger proportion of individuals over the age of 60 (88%) and Non-Hispanic Whites (74%) than the overall sample (82% and 65%, respectively). In New Mexico, there was a larger proportion of Hispanics among the respondents (31%) than in the overall sample (26%).

Tumor Characteristics

Common cancers represented among HINTS-SEER respondents include male reproductive (n=304; 24.6%), female breast (n=284; 23.0%), melanoma (n=119; 9.6%), and gastrointestinal (n=106; 8.6%). Cancer sites, by organ system recodes, can be found in Table 3 for the overall sample and by individual registry. Among HINTS-SEER respondents, tumors were mostly localized (67.8%), with 22.4% regional, 6.2% distant, and 3.7% unknown. Summary stage characteristics for the total sample and by individual registry can be found in Table 4.

DISCUSSION

HINTS is a routine data collection effort that uses probability methods to produce national estimates for health and cancer communication-related constructs and variables. HINTS-SEER 2021 was a unique pilot study intended to garner and assess a larger sample of cancer survivors than typically responds to probability-based administrations of HINTS. Cancer survivors were sampled from three SEER registries, and while not nationally representative, the HINTS-SEER samples are representative of the frames from which they were drawn. This pilot study will serve to inform future efforts to oversample cancer survivors for national surveys.

HINTS-SEER data users may conduct analyses to generate estimates that represent each individual SEER registry by using the registry flag on the dataset (SEERREGISTRY_FLAG) or combine across the three SEER samples to benefit from the full sample (N=1,234). Some caution should be taken when combining samples across the registries given the different consent procedures used and response rates achieved. Before conducting analyses across all sites, analysts should test whether there are differences in the outcome of interest between the three registries. Testing for differences can be completed using simple bivariate tests (e.g., t-tests) that compare the outcomes across the sites. If differences are found between the registries, they might represent differences in the methods used to recruit respondents and the resulting samples. In this case, models could include the registry flag as a covariate when analyzing the full sample. If there are differences across the registries, investigators should also 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 help to control for differences between registries.

It is possible to compare HINTS-SEER results to cancer survivors that are captured in the larger HINTS data collection program. Comparing the demographic distributions, the types of cancers, and participant responses to the health and communication items on the

surveys can provide users of the HINTS-SEER data a way to assess how the datasets differ along key outcomes of interest. For example, this might involve examining the percent of survivors who are satisfied with the care they are getting from their primary care physician, comparing the HINTS-SEER sample to the sample of cancer survivors from the national HINTS population survey. Data users can statistically compare HINTS and HINTS-SEER (i.e., conduct significance tests, run models) using the same methods analysts use when comparing data across separate HINTS cycles (Rizzo, 2008). However, *combining* HINTS-SEER with prior HINTS administrations to generate a single estimate is discouraged because the samples differ in important ways, including probability and non-probability sampling methods, geographic representation, sampling individual survivors versus households, and time of cancer diagnosis. The HINTS-SEER sample also excluded non-melanoma skin cancer, which is the most frequently reported cancer diagnosis on the general HINTS population survey.

Limitations

The study has several limitations, including low response rates and lack of national representation. In addition, the use of English-only survey materials may have discouraged limited- or non-English speaking survivors from participating in HINTS-SEER. Further, the SEER meta data available on the dataset are limited to four tumor-related variables (cancer site, date of diagnosis, histology, and summary stage) and therefore do not allow for a full SEER record linkage. Because of the different consent procedures used and response rates achieved across the participating registries, some caution should be taken when deciding how to combine and compare HINTS-SEER estimates to those from survivor responses to the HINTS probability-based surveys.

Conclusion

HINTS-SEER data provide analysts with a large pool of cancer survivors to conduct inferential statistics and can be used to examine a range of social, behavioral, and healthcare outcomes among cancer survivors. NCI encourages secondary data analyses using HINTS-SEER to identify and address extant information support needs of cancer survivors. The HINTS-SEER instrument, supporting documentation, and full methodology report are available for public use as of November 2022, at http://hints.cancer.gov. Because the HINTS-SEER dataset contains medical diagnosis data and may pose disclosure risk, the data are available as a restricted-use data file that is available only by request at https://hints.cancer.gov/data/restricted-data.aspx and users who are approved to obtain the data must sign a stringent data use agreement.

Data Availability Statement:

HINTS-SEER (2021) data are available by request by filling out a restricted-use data request form (https://hints.cancer.gov/data/restricted-data.aspx) and signing a Data Use Agreement.

REFERENCES

Finney Rutten LJ, Davis T, Beckjord EB, Blake K, Moser RP, Hesse BW. Picking Up the Pace: Changes in Method and Frame for the Health Information National Trends Survey (2011 – 2014). Journal of Health Communication 2012;17(7):979–989. [PubMed: 23020763]

Nelson DE, Kreps GL, Hesse BW, et al. The Health Information National Trends Survey (HINTS): Development, design, and dissemination. Journal of Health Communication. 2004;9(5):443–460. [PubMed: 15513791]

Hesse BW, Moser RP, Rutten LJ, Kreps GL. The health information national trends survey: research from the baseline. J Health Commun. 2006;11 Suppl 1:vii–xvi. doi: 10.1080/10810730600692553. [PubMed: 16641070]

Blake KD, Portnoy DB, Kaufman AR, Lin CJ, Lo SC, Backlund E, Cantor D, Hicks L, Lin A, Caporaso A, Davis T, Moser RP, Hesse BW. Rationale, Procedures, and Response Rates for the 2015 Administration of NCI's Health Information National Trends Survey: HINTS-FDA 2015. J Health Commun. 2016 Dec;21(12):1269–1275. doi: 10.1080/10810730.2016.1242672. Epub 2016 Nov 28. [PubMed: 27892827]

Finney Rutten LJ, Blake KD, Skolnick VG, Davis T, Moser RP, Hesse BW. Data Resource Profile: The National Cancer Institute's Health Information National Trends Survey (HINTS). Int J Epidemiol. 2020 Feb 1;49(1):17–17j. doi: 10.1093/ije/dyz083. [PubMed: 31038687]

Blake KD, Rutten L, Moser RP, Hesse B. The U.S. National Cancer Institute's Health Information National Trends Survey (HINTS). In Ho EY, Bylund CL, van Weert JCM (Eds). International Encyclopedia of Health Communication (2022). Wiley-Blackwell. 10.1002/9781119678816.iehc0745

Dillman DA, Smyth JD, Christian LM. Internet, mail, and mixed-mode surveys: The tailored design method. Hoboken, NJ: John Wiley & Sons; 2009.

Sakshaug JW, Schmucker A, Kreuter F, Couper MP, Singer E. Evaluating Active (Opt-In) and Passive (Opt-Out) Consent Bias in the Transfer of Federal Contact Data to a Third-Party Survey Agency. Journal of Survey Statistics and Methodology 2016 Sept;4(3):382–416.

Rizzo L, Moser RP, Waldron W, Wang Z, and Davis WW (2008). 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

Table 1.

Data Collection History. Health Information National Trends Survey (HINTS) 2003–2022

Administration (Year)	Data Collection Period	Mode	N and Response Rate (RR)	Number of Cancer Survivors in Sample	% of Cancer Survivors in Sample
HINTS 1 (2003)	Oct 2002-Apr 2003	RDD	N=6369, RR=33%	763	12
HINTS 2 (2005)	Feb 2005-Aug 2005	RDD (plus online pilot)	N=5586, RR=21%	873	16
HINTS 3 (2008)	Jan 2008-Apr 2008	Postal and RDD	N=3582 (mail), RR=30.9% N=4092 (RDD), RR=24.2%	1001	13
HINTS Puerto Rico (2009)	Apr 2009-June 2009	RDD	N=639, RR=76%	44	<1
HINTS 4 Cycle 1 (2011–2012)	Oct 2011-Feb 2012	Postal	N=3565, RR=37.91%	563	16
HINTS 4 Cycle 2 (2012–2013)	Oct 2012-Jan 2013	Postal	N=3630, RR=39.97%	464	13
HINTS 4 Cycle 3 (2013)	Sept 2013-Nov 2013	Postal	N=3185, RR=35.19%	459	14
HINTS 4 Cycle 4 (2014)	Aug 2014-Nov 2014	Postal	N=3677, RR=34.44%	542	15
HINTS-FDA (2015)	May-Sept 2015	Postal	N=3738, RR=33.04%	Question not asked	NA
HINTS-FDA Cycle 2 (2017)	Jan-May 2017	Postal	N=1736, RR=34.05%	Question not asked	NA
HINTS 5 Cycle 1 (2017)	Jan-May 2017	Postal	N=3285, RR=32.4%	504	15
HINTS 5 Cycle 2 (2018)	Jan-May 2018	Postal	N=3527, RR=32.85%	593	17
HINTS 5 Cycle 3 (2019)	Jan-Apr 2019 Jan-May 2019	Postal Web pilot	N=4573 (mail), RR=30.2% N=865 (Web), RR=30.6% N=5438 (total), RR=30.3% (overall)	856	16
HINTS 5 Cycle 4 (2020)	Feb-June 2020	Postal	N=3865, RR=37%	626	16
HINTS 6 (2022)	Mar-Nov 2022	Postal and Push to Web	N=6252, RR=28.07%	926	15

RDD= Random digit dial

Table 2.

Sample Characteristics. HINTS-SEER 2021, Unweighted Frequencies and Weighted Proportions by SEER Registry and Total Sample, N=1234

Variable	Iowa Unweighted Frequency	Iowa Weighted Percent	Iowa S.E. of Percent	Bay Area Unweighted Frequency	Bay Area Weighted Percent	Bay Area S.E. of Percent	New Mexico Unweighted Frequency	New Mexico Weighted Percent	New Mexico S.E. of Percent	Total Sample Unweighted Frequency	Total Sample Weighted Percent	Total Sample S.E. of Percent	N Missing
Years Since Cancer Diagnosis (M = 14.14, S.E. = .23)													
Less than 5 years	39	9.54	0.93	51	10.94	1.42	30	8.55	1.34	120	10.19	0.89	
5 to less than 10 years	141	30.35	1.07	130	27.87	1.97	121	34.81	1.80	392	29.59	1.13	
10 years or more	228	60.11	0.82	300	61.19	1.88	194	56.64	1.98	722	60.22	1.16	
Total	408	100		481	100		345	100		1234	100		0
Sex Assigned at Birth													
Male	184	43.01	0.68	220	46.54	0.50	153	43.33	0.84	557	45.06	0.34	
Female	215	56.99	0.68	248	53.46	0.50	182	56.67	0.84	645	54.94	0.34	
Total	399	100		468	100		335	100		1202	100		32
Age (M = 70.45, S.E. = .24)													
18–34	*	*	*	*	*	*	*	*	*	10	1.15**	0.38**	
35–49	16	5.66**	1.34**	11	4.11**	0.90**	9	*	*	36	4.42	0.73	
50–64	95	26.11	2.13	74	20.99	1.65	75	23.84	1.34	244	22.89	1.13	
65–74	142	29.74	1.99	158	32.08	2.03	115	30.36	1.75	415	31.15	1.26	
75+	141	37.13	1.93	217	41.78	1.55	133	41.44	1.57	491	40.38	1.10	
Total	398	100		463	100		335	100		1196	100		38
Race													
White only	393	98.67	0.64	364	76.80	2.62	291	93.81	1.32	1048	85.63	1.54	
Black only	*	*	*	12	5.41**	1.39**	*	*	*	17	3.29**	0.80**	

Variable	Iowa Unweighted Frequency	Iowa Weighted Percent	Iowa S.E. of Percent	Bay Area Unweighted Frequency	Bay Area Weighted Percent	Bay Area S.E. of Percent	New Mexico Unweighted Frequency	New Mexico Weighted Percent	New Mexico S.E. of Percent	Total Sample Unweighted Frequency	Total Sample Weighted Percent	Total Sample S.E. of Percent	N Missing
American Indian or Alaska Native only				*	*	*	6	*	*	7	*	*	
Multiple races selected	*	*	*	11	3.58 **	1.04 **	5	*	*	19	2.51 **	0.65 **	
Asian Indian only				5	*	*		*	*	5	*	*	
Chinese only				29	6.03	1.29	*	*	*	31	3.48	0.72	
Filipino only	*	*	*	16	3.22 **	0.79 **	*	*	*	18	1.94 **	0.45 **	
Japanese only				8	*	*				8	*	*	
Korean only				*	*	*				*	*	*	
Vietnamese only				*	*	*				*	*	*	
Other Asian only				*	*	*				*	*	*	
Native Hawaiian only					*	*	*	*	*	*	*	*	
Other Pacific Islander only				*	*	*				*	*	*	
Total	398	100		455	100		310	100		1163	100		71
Ethnicity													
Not Hispanic only	374	97.85	0.85	404	89.17	1.34	225	64.69	2.06	1003	88.06	0.75	
Mexican only	*	*	*	16	6.66 **	1.28 **	31	10.84	1.67	51	5.75	0.79	
Other Hispanic only	*	*	*	13	3.67 *	0.97 **	66	22.77	2.17	81	5.55	0.66	
Multiple Hispanic ethnicities selected	*	*	*	*	*	*	*	*	*	6	*	*	
Total	381	100		434	100		326	100		1141	100		93
Derived Race/Ethnicity													

Variable	Iowa Unweighted Frequency	Iowa Weighted Percent	Iowa S.E. of Percent	Bay Area Unweighted Frequency	Bay Area Weighted Percent	Bay Area S.E. of Percent	New Mexico Unweighted Frequency	New Mexico Weighted Percent	New Mexico S.E. of Percent	Total Sample Unweighted Frequency	Total Sample Weighted Percent	Total Sample S.E. of Percent	N Missing
Non-Hispanic White	370	97.01	0.94	324	69.68	2.04	212	61.22	2.20	906	76.43	1.36	
Non-Hispanic Black or African American	*	*	*	10	4.66**	1.45**	*	*	*	15	2.84**	0.82**	
Hispanic	7	*	*	30	10.92	1.35	101	35.40	2.06	138	12.00	0.76	
Non-Hispanic Asian	*	*	*	60	13.24	1.88	8	*	*	64	7.58	1.05	
Non-Hispanic Other	*	*	*	6	*	*	5	*	*	13	1.15**	0.42**	
Total	381	100		430	100		325	100		1136	100		98
Education													
Less than High School	6	*	*	14	3.46**	1.10**	16	5.27**	1.04**	36	3.19	0.66	
High School Graduate	64	15.97	2.03	36	7.92	1.13	52	16.24	1.94	152	11.46	0.82	
Some College	142	36.14	2.63	95	20.94	2.54	92	27.57	2.70	329	26.30	1.63	
College Graduate or More	186	46.26	2.74	321	67.68	2.44	178	50.92	3.07	685	59.05	1.65	
Total	398	100		466	100		338	100		1202	100		32
Household Income													
Less than $20,000	32	9.22	1.71	22	5.17**	1.07**	36	13.00	1.74	90	7.54	0.74	
$20,000 to < $35,000	47	11.70	1.65	32	7.85	1.31	46	15.36	2.12	125	10.11	0.95	
$35,000 to < $50,000	56	14.07	1.82	30	8.58	1.60	33	11.27	1.60	119	10.65	0.93	
$50,000 to < $75,000	73	18.74	2.13	61	14.16	1.77	54	17.31	2.11	188	16.02	1.23	
$75,000 or more	171	46.26	2.77	264	64.23	2.26	132	43.07	2.92	567	55.68	1.41	
Total	379	100		409	100		301	100		1089	100		145

Variable	Iowa Unweighted Frequency	Iowa Weighted Percent	Iowa S.E. of Percent	Bay Area Unweighted Frequency	Bay Area Weighted Percent	Bay Area S.E. of Percent	New Mexico Unweighted Frequency	New Mexico Weighted Percent	New Mexico S.E. of Percent	Total Sample Unweighted Frequency	Total Sample Weighted Percent	Total Sample S.E. of Percent	N Missing
Marital Status													
Married	287	72.79	2.09	297	64.02	1.96	210	61.97	2.57	794	66.26	1.38	
Living as married or living with a romantic partner	11	2.94**	0.84**	15	3.19**	1.02**	16	4.82**	1.33**	42	3.36	0.70	
Divorced	35	8.39	1.35	49	10.56	1.58	57	16.86	1.83	141	10.85	0.99	
Widowed	41	9.98	1.53	67	13.16	1.68	39	12.44	1.54	147	12.14	1.15	
Separated		*		5	*	*	*	*	*	7	*	*	
Single, never been married	23	5.90**	1.46**	31	7.38	1.31	11	*	*	65	6.36	0.90	
Total	397	100		464	100		335	100		1196	100		38
Sexual Orientation													
Heterosexual or straight	383	97.73	0.87	434	96.98	0.78	311	96.26	0.84	1128	97.10	0.52	
Homosexual, or gay or lesbian	*	*	*	11	2.49**	0.75**	10	2.65**	0.75**	25	2.10	0.47	
Bisexual	*	*	*	*	*	*	*	*	*	9	*	*	
Other	*	*	*	*	*	*	*	*	*	*	*	*	
Total	391	100		448	100		325	100		1164	100		70
Gender Identity													
Male	184	43.08	0.69	219	46.45	0.50	150	43.13	0.93	553	45.00	0.32	
Female	214	56.92	0.69	247	53.03	0.61	178	56.22	1.06	639	54.61	0.38	
Transgender/ Gender Non Conforming		*		*	*	*	*	*	*	*	*	*	
Total	398	100		467	100		330	100		1195	100		39
Sexual or Gender Minority													

Variable	Iowa Unweighted Frequency	Iowa Weighted Percent	Iowa S.E. of Percent	Bay Area Unweighted Frequency	Bay Area Weighted Percent	Bay Area S.E. of Percent	New Mexico Unweighted Frequency	New Mexico Weighted Percent	New Mexico S.E. of Percent	Total Sample Unweighted Frequency	Total Sample Weighted Percent	Total Sample S.E. of Percent	N Missing
Cisgender Heterosexual	382	97.73	0.87	432	96.44	0.90	305	95.84	0.89	1119	96.73	0.59	
Sexual or Gender Minority	8	*	*	15	3.56**	0.90**	15	4.16**	0.89**	38	3.27	0.59	77
Total	390	100		447	100		320	100		1157	100		
Smoking Status													
Current	17	4.52**	1.00**	16	3.37**	0.82**	15	4.29**	1.04**	48	3.84	0.57	
Former	161	39.77	2.53	167	34.21	2.25	128	37.13	2.66	456	36.25	1.54	
Never	225	55.71	2.91	288	62.42	2.48	194	58.58	2.75	707	59.91	1.73	
Total	403	100		471	100		337	100		1211	100		23
Rural/Urban Status													
Urban (RUCC 1–3)	223	55.56	2.72	476	99.06	.41	249	71.17	2.23	948	82.49	1.02	
Rural (RUCC 4–9)	185	44.44	2.72	5	*	*	96	28.83	2.23	286	17.50	1.02	
Total	408	100		481	100		345	100		1234	100		0

*
Note that for cell sizes from 5 to <10, unweighted frequencies are reported, but weighted proportions are suppressed due to instability of weighted estimates for small cells. For cell sizes <5, neither unweighted frequencies nor weighted proportions are reported.

**
Note that weighted estimates for small cell sizes (between 10 and 25) are reported for illustrative purposes, however these estimates are prone to instability and should not be used for prevalence estimates of cancer survivors, nor should they be used to create population estimates of health-related characteristics or needs of cancer survivors in the sample.

S.E.=Standard Error

Table 3.

Cancer Site Organ Systems. HINTS-SEER 2021. Unweighted Frequencies and Weighted Proportions by SEER Registry and Total Sample, N=1234

Cancer Site Organ Systems Recode	Iowa Unweighted Frequency	Iowa Weighted Percent	Iowa S.E. of Percent	Bay Area Unweighted Frequency	Bay Area Weighted Percent	Bay Area S.E. of Percent	New Mexico Unweighted Frequency	New Mexico Weighted Percent	New Mexico S.E. of Percent	Total Sample Unweighted Frequency	Total Sample Weighted Percent	Total Sample S.E. of Percent
Head and Neck	14	3.95 **	1.02 **	15	2.97 **	0.79 **	8	*	*	37	3.16	0.61
Gastrointestinal Tube	29	9.25	0.88	46	9.79	1.70	31	8.86	1.53	106	9.50	1.06
Gastrointestinal Solid Organs	*	*	*	6	*	*	*	*	*	10	0.97 **	0.36 **
Respiratory System	8	*	*	14	3.02 **	0.75 **	6	*	*	28	2.48	0.42
Mediastinum				*	*	*				*	*	*
Bone & Soft Tissue	*	*	*	*	*	*	5	*	*	13	1.06 **	0.31 **
Skin - Melanoma	32	9.35	1.00	56	11.17	1.61	31	9.40	1.62	119	10.39	1.07
Breast	98	24.07	0.39	113	23.60	1.96	73	22.21	1.83	284	23.53	1.16
Female Reproductive System	35	8.46	0.65	39	8.65	1.49	30	8.91	1.78	104	8.64	0.98
Male Reproductive System	90	18.73	0.69	123	24.73	1.82	91	25.51	1.92	304	23.11	1.08
Urinary Tract	36	5.11	0.07	12	2.46 **	0.88 **	20	5.70 **	1.19 **	68	3.69	0.57
Eye				*	*	*	*	*	*	*	*	*
Central Nervous System & Meninges							*	*	*	*	*	*
Endocrine System	15	5.36 **	0.86 **	11	2.56 **	0.76 **	19	5.88 **	1.15 **	45	3.85	0.56
Lymphocytic Lymphomas &/or Lymphoblastic Leukemias	35	8.56	0.75	30	6.64	1.18	19	5.27 **	1.30 **	84	7.00	0.73
Myelodysplastic/ Myeloproliferative Neoplasms & Myeloid Leukemias	6	*	*	*	*	*	5	*	*	15	1.14 **	0.32 **
Myeloma & Plasma Cell Disorders	*	*	*	*	*	*	*	*	*	9	*	*

Cancer Site Organ Systems Recode	Iowa Unweighted Frequency	Iowa Weighted Percent	Iowa S.E. of Percent	Bay Area Unweighted Frequency	Bay Area Weighted Percent	Bay Area S.E. of Percent	New Mexico Unweighted Frequency	New Mexico Weighted Percent	New Mexico S.E. of Percent	Total Sample Unweighted Frequency	Total Sample Weighted Percent	Total Sample S.E. of Percent
Cancer of Unknown Primary	*	*	*	*	*	*	*	*	*	*	*	*
Total	408	100		481	100		345	100		1234	100	

*
Note that for cell sizes from 5 to <10, unweighted frequencies are reported, but weighted proportions are suppressed due to instability of weighted estimates for small cells. For cell sizes <5, neither unweighted frequencies nor weighted proportions are reported.

**
Note that weighted estimates for small cell sizes (between 10 and 25) are reported for illustrative purposes, however these estimates are prone to instability and should not be used for prevalence estimates of cancer survivors, nor should they be used to create population estimates of health-related characteristics or needs of cancer survivors in the sample.

S.E.=Standard Error

Table 4.

SEER Summary Stage. HINTS-SEER 2021. Unweighted Frequencies and Weighted Proportions by SEER Registry and Total Sample, N=1234

Summary Stage	Iowa Unweighted Frequency	Iowa Weighted Percent	Iowa S.E. of Percent	Bay Area Unweighted Frequency	Bay Area Weighted Percent	Bay Area S.E. of Percent	New Mexico Unweighted Frequency	New Mexico Weighted Percent	New Mexico S.E. of Percent	Total Sample Unweighted Frequency	Total Sample Weighted Percent	Total Sample S.E. of Percent
Localized only	267	64.85	0.80	335	69.37	2.10	231	67.47	2.32	833	67.79	1.32
Regional by direct extension only	82	19.66	0.40	26	5.40	0.88	28	7.40	1.43	136	9.80	0.58
Regional lymph nodes only				45	9.76	1.46	32	9.92	1.76	77	6.97	0.88
Regional by BOTH direct extension AND lymph node involvement	29	7.48	0.40	18	3.59 **	0.81 **	8	*	*	55	4.51	0.49
Regional, Not Otherwise Specified				7	*	*	7	*	*	14	1.11 **	0.31 **
Distant site(s)/ node(s) involved				40	8.57	1.37	32	8.93	1.47	72	6.15	0.79
Unknown if extension or metastasis (unstaged, unknown, or unspecified)	30	8.01	0.77	10	1.88 **	0.66 **	7	*	*	47	3.67	0.46
Total	408	100		481	100		345	100		1234	100	

* Note that for cell sizes < 10, unweighted results are reported, but weighted results are suppressed due to instability of weighted estimates for small cells

** Note that weighted estimates for small cell sizes (between 10 and 25) are reported for illustrative purposes, however these estimates are prone to instability and should not be used for prevalence estimates of cancer survivors, nor should they be used to create population estimates of health-related characteristics or needs of cancer survivors in the sample