



Published in final edited form as:

*Med Care*. 2013 February ; 51(2): e9–15. doi:10.1097/MLR.0b013e318222a711.

## Representativeness of Participants in the Cancer Care Outcomes Research and Surveillance (CanCORS) Consortium Relative to the Surveillance, Epidemiology and End Results (SEER) Program

**Paul J. Catalano, ScD,**

Dana-Farber Cancer Institute, Department of Biostatistics and Computational Biology and Harvard School of Public Health, Department of Biostatistics, Boston MA

**John Z. Ayanian, MD, MPP,**

Harvard Medical School, Department of Health Care Policy and Brigham and Women's Hospital, Division of General Medicine, Boston MA

**Jane C. Weeks, MD, MSc,**

Dana-Farber Cancer Institute, Department of Medical Oncology, Boston MA

**Katherine L. Kahn, MD,**

RAND Corporation, Santa Monica, CA and Division of General Internal Medicine and Health Services Research at the David Geffen School of Medicine at UCLA, Los Angeles CA

**Mary Beth Landrum, PhD,**

Harvard Medical School, Department of Health Care Policy, Boston MA

**Alan M. Zaslavsky, PhD,**

Harvard Medical School, Department of Health Care Policy, Boston MA

**Jeannette Lee, PhD,**

University of Arkansas for Medical Sciences, Department of Biostatistics, Little Rock AR

**Jane Pendergast, PhD, and**

---

*Author for Correspondence:* Paul J. Catalano Department of Biostatistics and Computational Biology Dana-Farber Cancer Institute 44 Binney Street Boston, MA 02115-6084 Tel: 617-632-2441 Fax: 617-632-2444 pcata@jimmy.harvard.edu.

Paul J. Catalano, ScD pcata@jimmy.harvard.edu 617-632-2441 617-632-2444f Department of Biostatistics and Computational Biology Dana-Farber Cancer Institute 44 Binney St Boston, MA 02115

John Z. Ayanian, MD, MPP ayanian@hcp.med.harvard.edu 617-432-3455 617-432-0173f Department of Health Care Policy Harvard Medical School 180 Longwood Ave Boston, MA 02115

Jane C. Weeks, MD, MSc jane\_weeks@dfci.harvard.edu 617-632-2509 617-582-8560f Center for Outcomes and Policy Research Dana-Farber Cancer Institute 44 Binney St Boston, MA 02115

Katherine L. Kahn, MD kahn@rand.org 310-393-0411 310-451-6917f RAND PO Box 2138 1700 Main St M-10 Santa Monica, CA 90409-2138

Mary Beth Landrum, PhD landrum@hcp.med.harvard.edu 617-432-2460 617-432-0173f Department of Health Care Policy Harvard Medical School 180 Longwood Ave Boston, MA 02115

Alan M. Zaslavsky, PhD zaslavsk@hcp.med.harvard.edu 617-432-2441 617-432-0173f Department of Health Care Policy Harvard Medical School 180 Longwood Ave Boston, MA 02115

Jeannette Lee, PhD jylee@uams.edu 501-526-6712 501-526-6729f Department of Biostatistics University of Arkansas for Medical Sciences 4301 West Markam St #781 Little Rock, AR 72205

Jane Pendergast, PhD jane\_pendergast@uiowa.edu 319-384-5028 319-384-5018f Department of Biostatistics University of Iowa College of Public Health 200 Hawkins Drive C22K, General Hospital Iowa City, IA 52242

David P. Harrington, PhD dph@jimmy.harvard.edu 617-632-2450 617-632-2444f Department of Biostatistics and Computational Biology Dana-Farber Cancer Institute 44 Binney St Boston, MA 02115

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

University of Iowa College of Public Health, Biostatistics Department, Iowa City IA

**David P. Harrington, PhD**

Dana-Farber Cancer Institute, Department of Biostatistics and Computational Biology and Harvard School of Public Health, Department of Biostatistics, Boston MA

**for the Cancer Care Outcomes Research and Surveillance Consortium**

## Abstract

**Background**—The research goals of the Cancer Care Outcomes Research and Surveillance (CanCORS) Consortium are to determine how characteristics and beliefs of patients, providers, and health-care organizations influence the treatments and outcomes of individuals with newly diagnosed lung and colorectal cancers. Because CanCORS results will inform national policy, it is important to know how they generalize to the United States population with these cancers.

**Research Design**—This study assessed the representativeness of the CanCORS cohort of 10,547 patients with lung cancer (LC) or colorectal cancer (CRC) enrolled between 2003 and 2005. We compared characteristics (gender, race, age and disease stage) to the Surveillance, Epidemiology and End Results (SEER) population of 234,464 patients with new onset of these cancers during the CanCORS recruitment period.

**Results**—The CanCORS sample is well matched to the SEER Program for both cancers. In CanCORS, 41% LC / 47% CRC were female versus 47% LC / 49% CRC in SEER. African American, Hispanic and Asian cases differed by no more than 5 percentage points between CanCORS and SEER. The SEER population is slightly older, with the percentage of patients over 75 years 33.1% LC / 37.3% CRC in SEER versus 26.9% LC / 29.4% in CanCORS, and also has a slightly higher proportion of early stage patients. We also found that the CanCORS cohort was representative within specific SEER regions that map closely to CanCORS sites.

**Conclusions**—This study demonstrates that the CanCORS Consortium was successful in enrolling a demographically representative sample within the CanCORS regions.

## Keywords

Lung Cancer; Colorectal Cancer; Cancer Populations

## Introduction

The Cancer Care Outcomes Research and Surveillance (CanCORS) Consortium was funded and organized in 2001 to study lung and colorectal cancers through a cooperative agreement between the National Cancer Institute (NCI) and the Veterans' Administration and 7 Primary Data Collection and Research (PDCR) sites as well as a central Statistical Coordinating Center (SCC) in Boston, MA<sup>1</sup>. The seven PDCR sites include five sites for geographically defined regions, one site with 5 integrated health-care delivery systems in the NCI-funded Cancer Research Network (CRN), and one site with 15 Veterans' Administration hospitals. Taken together, these sites represent a total population of approximately 30 million people.<sup>1</sup>

The overall goals of the primary CanCORS study are to determine how the characteristics and beliefs of lung and colorectal cancer patients, physicians and health-care organizations influence treatments and outcomes spanning the continuum of cancer care from diagnosis to recovery or death, and to evaluate the effects of specific therapies on patients' survival, quality of life, and satisfaction with care. The project has three major strengths that distinguish it from prior studies of cancer care. First, diverse patient cohorts have been

prospectively enrolled from multiple regions and health-care systems. Second, patients or their surrogates have been surveyed relatively soon after diagnosis, so their beliefs, symptoms, and health-care experiences could be assessed in a timely manner. Third, by blending extensive surveys of patients and physicians with detailed clinical data from medical records, the study has unparalleled breadth and depth.

The sites participating in CanCORS were selected by peer-review from among applicants to an NCI Request for Applications, without specific attention to geographic representativeness. And although the study design mandated random selection of potential study subjects within these sites, those eligible for participation were newly-diagnosed cancer patients, some of them quite ill, who were asked to cooperate with extensive data gathering. For these reasons, achieving a representative sample was challenging. If the participants can be shown to be representative of the general population, the findings of CanCORS can inform decisions about patient management and health care policy for the United States as a whole. Therefore, it is important to know the extent to which the CanCORS cohort is representative of the broader population of patients diagnosed with lung and colorectal cancers.

The purpose of this analysis was to assess the representativeness of the CanCORS cohort relative to the broader population of patients diagnosed with lung cancer or colorectal cancer during the CanCORS recruitment period of 2003 through 2005. While there are no large, nationally representative population-based cohorts of cancer patients to serve as a comparison group, cancer registries participating in the NCI SEER Program have been previously compared to the overall United States population<sup>2,3</sup> and the representativeness of SEER has been discussed.<sup>4</sup> Thus to achieve our objective, we compared the full CanCORS cohort with total population diagnosed in all 17 SEER regions and also compared the subset of CanCORS patients residing in SEER areas (three geographic sites plus three large CRN health plans) with the specific patient populations in these SEER regions. A similar approach has been used to assess the representativeness of research programs with respect to population measures in Denmark<sup>5</sup> and China<sup>6</sup>.

## Methods

### CanCORS Recruitment and Enrollment

The scientific goals of CanCORS rely heavily on data sets that link self-reported experiences with medical record data on cancer diagnosis, care and outcomes. The Consortium used a variety of instruments to obtain these linked data, including an array of patient interview options for both the baseline and the follow-up instruments, a standardized computerized medical record abstraction tool, and self-administered surveys of physicians involved in patient care and of caregivers who provided support to the patients.

Each PDCR site identified patients with newly diagnosed lung cancer or colorectal cancer for aggregate population-based cohorts of approximately 5,000 patients for each cancer. Five of the PDCR sites (Northern California Cancer Center [NCCC], University of Alabama [UAB], University of Iowa [UIOWA], University of North Carolina, Chapel Hill [UNC] and University of California, Los Angeles [UCLA]) identified and enrolled participants using rapid case ascertainment (RCA) from cancer registries based on the geographic area in which the site was located<sup>7</sup>. The two other sites (the CRN and the Veterans' Administration [VA]) enrolled participants using RCA from cancer registries within the provider organization in which the participant was a member. We henceforth refer to the CRN and VA as "provider-organization-based" sites, and the other 5 sites as "geographically-based" sites. In some sites, the study team attempted to oversample or enroll all patients from certain demographic subgroups (e.g., African-American colorectal cancer patients in

Alabama). In instances where the number of eligible patients in a demographic subgroup exceeded the number needed for enrollment, participants were chosen randomly from the sequence of incident patients. The total populations and expected incident lung and colorectal cancer during the enrollment period by PDCR site have been previously described.<sup>1</sup>

## Eligibility

Inclusion criteria required a histologically or cytologically confirmed new diagnosis of invasive colorectal or lung cancer from a specific list of eligible histologic types that included over 95% of all cancers of the lung and colon/rectum. Participants had to be at least 21 years of age at diagnosis and the participant or surrogate must have been able to complete the interview in English, except in Los Angeles County and Northern California, where interviews could be conducted in Spanish and Chinese (Mandarin or Cantonese). There were no exclusions based on race, gender or ethnicity; patients incarcerated in correctional facilities were not eligible. The CanCORS study protocol was approved by institutional review boards (IRBs) at all seven PDCR sites and at the Statistical Coordinating Center at the Dana-Farber Cancer Institute.

Patients were contacted by mail 4 months after diagnosis and invited to participate in a telephone survey. Any patient who consented to the survey and provided responses beyond the introductory script (or whose surrogate provided responses) was considered an enrolled participant. For patients alive at initial contact, interviewers were directed to offer the full telephone interview first, in the language chosen by the participant. A brief patient interview was the next option to be used if the patient was too sick to complete the full survey; a surrogate interview was conducted if the patient was unable to participate at all or was deceased. If an appropriately selected surrogate provided responses, the participant was considered enrolled. A 30 minute follow-up interview of patients alive at the baseline interview was conducted approximately 14 months after diagnosis. Living patients were asked to participate in the interview themselves; a surrogate follow-up survey was performed for patients who had died since the baseline interview. A patient who provided a signed medical record consent form but did not consent for any form of the survey was also considered an enrolled case. The medical record abstraction was conducted for all living patients who authorized the chart review and for many of the deceased patients after appropriate IRB waiver.

## Target Population and Determination of Sample

To provide adequate statistical power for key research questions, CanCORS investigators sought to enroll sufficient numbers of patients within each cancer diagnosis for completion of either the patient survey or the surrogate respondent survey, based on a predicted 55% response rate for both cancers.

Certain racial and/or ethnic subgroups were oversampled at PDCR sites where feasible, within constraints imposed by incidence and participation assumptions. The identified “over-sampled groups” were sampled at higher rates than whites in order to achieve desired sample sizes for statistical comparisons. Further details of the sampling plan are available upon request from the lead author.

Initial determination of race and ethnicity for sampling purposes occurred during RCA at most sites, using routinely collected data. Estimates from previous studies have shown that the RCA race data are accurate at most sites. Concordance of patients’ race between RCA and medical records was almost 99% in Alabama and over 96% in North Carolina; in both of these sites, almost all participants were either white or African-American.

The sample for medical record abstraction was identical to that for the patient survey, since consent for abstraction was obtained as part of the survey process. It was expected that medical record consent would be obtained for approximately 85% of the patients who participated in the patient survey and that medical record data would be available for only a small fraction of patients without surveys. The total sample size goals for participant survey and/or medical record abstraction were 5,714 lung and 5,304 colorectal cancers.

During recruitment and data collection, PDCRs were responsible for ascertaining diagnosed cases in partnership with a collaborating cancer registry for a state, region, health plan or hospital. In some instances, this was done by having CanCORS staff visit registry offices (Iowa, Northern California), and in some instances PDCRs received electronic data files from registries. Sites determined eligibility of potential participants from registry information. This work involved close contact with registries and detailed checking of pathology reports to ensure eligibility according to the histologic categories listed in the protocol.

Consortium investigators recognized the importance of timely and accurate data reporting and furthermore, that any differences in participation across sites would seriously damage the representativeness of the sample. The project used several mechanisms to track and maintain quality of PDCR performance: enrollment reports by PDCR site and survey instrument, on-site audits conducted by the SCC, centralized training of all interviewers, and random review of interview audiotapes to ensure compliance with recruitment and interview question scripts.

### Comparison to SEER Population

SEER is a multi-regional program funded by the NCI and CDC to collect cancer incidence and survival statistics from population-based cancer registries covering roughly 28% of the US population.<sup>2,8</sup> SEER data in combination with other national databases on health care delivery have been extensively used in many observational health services research studies.<sup>3</sup> We compared the characteristics of CanCORS enrollees to the full SEER data set and to specific SEER regions covered by CanCORS. The SEER data included invasive lung and colorectal cancers identified during the calendar years 2003 – 2005 at all 17 SEER registries, nearly matching the CanCORS diagnosis dates of May 2003 to December 2005. A SEER\*Stat<sup>8</sup> version 6.4.4 query run in May 2008 retrieved 134,635 Lung and 106,299 CRC cases. We then applied CanCORS exclusions: age < 21 years, stage 0 and occult cancers, and certain disallowed histologies (e.g., lipomatous neoplasms, blood vessel tumors, etc). The final SEER data set included 132,758 patients with lung cancer (98% of the SEER cases retrieved) and 101,706 patients with colorectal cancer (96% of the retrieved cases).

### Statistical Analysis

All enrolled CanCORS participants were included in the statistical comparisons with weighting inversely proportional to the sampling rate. Unweighted percentages were calculated for our final SEER data set. We compared distributions of variables available in both data sets: gender, age at diagnosis, race/ethnicity and stage of disease. Two main comparisons were investigated: the entire CanCORS cohort versus all SEER sites and region-specific comparisons for six CanCORS sites that corresponded closely to specific SEER areas. Response rates to the baseline patient survey were calculated using standard American Association of Public Opinion Research<sup>9</sup> formulae in two ways: baseline participant survey responders as a percentage of all patients sampled, not known to be ineligible, and for whom we had physician consent to contact the patient (an overall response rate); and as a proportion of the set of patients described above and for whom the Consortium obtained verifiably correct contact information, e.g., confirmation by a relative

or by information on a telephone voice mail (termed a "cooperation rate"). We also computed the corresponding contact rates, calculated as the percentage of eligible households that were reached by survey staff.

## Results

Between September 2003 and December 2005, CanCORS obtained baseline interviews for 5,150 eligible participants with lung cancer and 4,911 participants with colorectal cancer. This interview was the primary mode of enrollment, conducted approximately 4 months after diagnosis. In 416 lung and 70 colorectal cases, permission for medical record abstraction was obtained from patients or their surrogates after interview participation was declined. The total numbers of enrolled participants were therefore 5,566 with lung cancer and 4,981 with colorectal cancer.

The ascertainment and enrollment process for the baseline interview is summarized in Table 1(a) beginning with the number of patients identified using RCA in each cancer and ending in the number of patients enrolled. During the enrollment period, the Consortium identified 27,631 potential participants (14,327 with lung cancer, 13,304 with colorectal cancer). Of these, 21,872 were sampled from among those not known to be ineligible because of stage (i.e., noninvasive) or disease type, and the Consortium obtained physician consent to contact 21,335 of these individuals. As shown in Table 1(b), the overall response rates for the Consortium were 49% in lung cancer and 53% in colorectal cancer. The numbers of participants in the two cohorts correspond nearly exactly to the design goals for the study, so all the power and sample size goals of the study were met. Enrollment by PDCR site is provided in Table 2 and demographic characteristics by cancer are shown in Table 3 (CanCORS columns).

The baseline telephone interviews were available in 4 different versions: 2,478 lung and 3,089 colorectal cancer patients participated in a full interview of approximately 1 hour; 607 lung and 713 colorectal cancer patients participated in a shorter, structured 20 minute interview because they did not feel well enough to participate in the longer interview; surrogates completed one of two versions, depending on whether the patient was alive but too ill to conduct a phone interview (506 lung and 523 colorectal cancer surrogates) or deceased (1366 lung and 380 colorectal cancer surrogates). A limited number of self-administered paper surveys (98 in total) were completed by patients who agreed to participate, but did not wish to be interviewed by telephone.

Among the participants eligible for a follow-up interview (participants who undertook the living patient baseline survey), we obtained either a participant or surrogate interview from 80% of patients with lung cancer and 82% of those with colorectal cancer. Despite the length and complexity of the interviews, item response rates were very high for items on all instruments, generally in excess of 99% and rarely as low as 95%. Of particular interest is the cohort for whom both self-reported survey data and medical record data from the chart abstraction were obtained. Medical records were abstracted for 78% of lung cancer participants with a baseline interview, and 72% of colorectal cancer patients with a baseline interview.

Taking into account the CanCORS sampling rates described above, Table 3 shows the concordance between CanCORS as a whole and all SEER registry-diagnosed cases for the major categories of race/ethnicity, age and stage of disease. Comparisons of the proportion of patients with non-white race and ethnicity by age group are shown in Figure 1. These analyses demonstrate that the CanCORS sample is well matched to SEER-diagnosed cases in both cancers, although the CanCORS cohort is somewhat younger (e.g., median age of 72

for colorectal cancer in SEER vs 67 in CanCORS) and has a slightly higher proportion of earlier stage patients.

One factor contributing to representativeness of the CanCORS cohort may be the stability of the population throughout the enrollment process, from initial ascertainment to final enrollment. Notably the composition of the population changed only minimally as subjects were ascertained, sampled, contacted, consented and ultimately enrolled. For example, the proportion of the sample that was female changed by less than 2% throughout the enrollment process for either cancer. Other demographic and clinical characteristics exhibited similar stability (data not shown).

In addition to the national comparisons, we also investigated representativeness within specific SEER regions that corresponded to CanCORS PDCR populations. Table 4 shows within-region comparisons by gender, race and age greater than 75 years for 6 CanCORS sites that enrolled patients from corresponding SEER regions. In general, the characteristics were well matched. Sites having the largest discrepancies with SEER were also the sites with smallest enrollment (e.g., Hawaii and Detroit, individual sites within the CRN) making definitive conclusions about representativeness in these sites difficult since confidence intervals associated with their estimates are wide and include the SEER rates.

## Discussion

This study demonstrates that the CanCORS Consortium was successful in enrolling a demographically and clinically representative cohort in the regions covered by the CanCORS population that was reflective of newly diagnosed patients with lung or colorectal cancer in all SEER regions. The target population in the five geographic-based sites consisted of all patients diagnosed with colorectal or lung cancer living in the geographical area in which the site was located during the enrollment period for the study. The target population in the other sites (CRN and the VA) consisted of all patients diagnosed with colorectal or lung cancer during the enrollment period for the study who were treated by these provider organizations. The CanCORS cohort was also internally consistent in demographic and clinical makeup with respect to eligible, contacted, and enrolled subjects.

Design factors that may have contributed to the success of enrollment and retention of the target population include the provision of incentives with a monetary value of \$10 to \$20 to patients agreeing to participate in the study, close alignment of participating sites with their corresponding cancer registries, and rigorous, ongoing monitoring of ascertainment and enrollment by a central statistical coordinating center. The use of multiple survey instruments in three languages also contributed substantially to the Consortium's ability to obtain data on a representative cohort. The fact that a decedent survey was performed for 27% of all lung cancer patients highlights the inherent bias in enrollment strategies that miss patients diagnosed with late stage, aggressive cancer. The questions on the four survey versions were designed so that all instruments would provide consistent data in the most important research domains.

While the CanCORS sample is somewhat younger than the SEER population, the Consortium was notably more successful in enrolling the elderly than, for example, NCI-sponsored cancer cooperative groups. Of note, 12.6% of colorectal cancers cases in SEER were over the age of 85 whereas 8.2% of the CanCORS cohort was in this age group. In contrast, only 0.4% of patients with colorectal cancer enrolled on Eastern Cooperative Oncology Group clinical trials are older than 85 years of age (lead author personal communication).

Strengths of the CanCORS Consortium include investigating important research questions, and a large, population-based sample of patients enrolled soon after diagnosis from multiple regions and health care systems representing about 10% of the U.S. population. The Consortium also over-sampled minorities and collected a rich set of variables bearing on cancer care and outcomes. Strengths of the present analysis lie in the robustness, scope and sample size of SEER in describing the U.S. population of incident lung and colorectal cancers.

This analysis also has several limitations. Only a limited set of characteristics was available in SEER for assessment and comparison and the presumption is that SEER represents a gold standard for U.S. cancer cases diagnosed during the study period. Previous studies have demonstrated that the SEER population is representative of the U.S. population in terms of age and sex, although SEER areas are more urban and more affluent than non-SEER areas<sup>2</sup>. Finally, the representativeness of specific subsets with relatively small sample sizes was difficult to assess.

Representativeness is fundamental to generalizability of research findings in CanCORS and other observational studies of cancer treatment and outcomes. Recruitment and retention strategies applied in this project were successful in achieving this goal and thus could be used in future research. With representative cohorts, observational studies in cancer health services and outcomes research can be analyzed with greater confidence to guide clinical decision-making and health policy.

## Acknowledgments

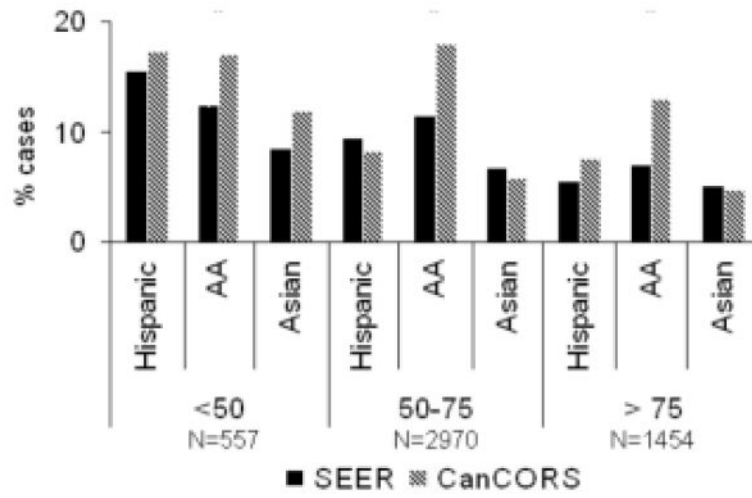
The Cancer Care Outcomes Research and Surveillance (CanCORS) Consortium was supported by grants from the National Cancer Institute (NCI) to the Statistical Coordinating Center at the Dana-Farber Cancer Institute (U01 CA093344) and the Primary Data Collection and Research Centers at the Harvard Medical School and Northern California Cancer Center (U01 CA093324); Dana-Farber Cancer Institute and Cancer Research Network (U01 CA093332); RAND and University of California, Los Angeles (U01 CA093348); University of Alabama at Birmingham (U01 CA093329); University of Iowa (U01 CA093339); University of North Carolina (U01 CA093326); and by a Department of Veteran's Affairs grant to the Durham VA Medical Center (U01 CDA093344 (MOU) and HARQ 03-438MO-03).

## References

1. Ayanian JZ, Chrischilles EA, Fletcher RH, et al. Understanding cancer treatment and outcomes: the Cancer Care Outcomes Research and Surveillance Consortium. *J Clin Oncol.* 2004; 22:2992–6. [PubMed: 15284250]
2. Nattinger AB, McAuliffe TL, Schapira MM. Generalizability of the SEER registry population: factors relevant to epidemiologic and health care research. *J Clin Epidemiol.* 1997; 50:939–945. [PubMed: 9291879]
3. Warren JL, Klabunde CN, Schrag D, et al. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care.* 2002; 40(suppl):IV-3–IV-18.
4. Frey DA, McMillen MM, Cowan CD, et al. Representativeness of the Surveillance, Epidemiology, and End Results Program Data: recent trends in cancer mortality rates. *J NCI.* 1992; 84(11):872–877.
5. Jensen AR, Storm HH, Møller S, et al. Validity and representativity in the Danish Breast Cancer Cooperative Group. *Acta Oncologica.* 2003; 42(3):179–185. [PubMed: 12852693]
6. Li GL, Chen WQ. Representativeness of population-based cancer registration in China – comparison of urban and rural areas. *Asian Pacific J Cancer Prevention.* 2009; 10:559–564.
7. Pearson ML, Ganz PA, McGuigan K, et al. The case identification challenge in measuring quality of cancer care. *J Clin Oncol.* 2002; 20(21):4353–4360. [PubMed: 12409335]



8. Surveillance, Epidemiology, and End Results (SEER) Program Populations (1969-2007). National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch; 2011.  
[www.seer.cancer.gov](http://www.seer.cancer.gov)
9. The American Association for Public Opinion Research (AAPOR). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 5th edition. AAPOR; Lenexa, Kansas: 2008.



**Figure 1.** Distribution of Non-White Patients by Race/Ethnicity and Age. Sample sizes refer to the CanCORS cohorts.

**Table 1**

Ascertainment and Enrollment of CanCORS Participants and Response Rates to the Baseline Participant Survey

	<b>Lung Cancer</b>	<b>Colorectal Cancer</b>	<b>Total</b>
<b>Part (a)</b>			
Ascertained	14327	13304	27631
Not known to be ineligible at Ascertainment	13455	12067	25522
Sampled	11630	10242	21872
Physician Consent to Enroll	11301	10034	21335
Successful Contact of Household	8340	7843	16183
Eligible after contact	7901	7303	15204
Enrolled to baseline interview and eligible	5150	4911	10061
Baseline survey and/or medical record review	5566	4981	10547
<b>Part (b)</b>			
Response Rate *	49.1%	53.2%	51.0%
Cooperation Rate *	58.9%	61.0%	59.9%
Contact Rate *	83.4%	87.3%	85.2%
Refusal Rate *	20.1%	20.2%	20.2%

\* Calculated according to standards of the American Association of Public Opinion Research (AAPOR, 2008)

**Table 2**

Total Enrollment by CanCORS Primary Data Collection and Resarch (PDCR) Site

PDCR Site	Collaborating Organizations	Cancer	
		Lung	Colorectal
Alabama (UAB)	University of Alabama, Birmingham *	668	649
Cancer Research Network (CRN) Seattle, Washington Boston, Massachusetts Detroit, Michigan Hawaii Portland, Oregon	Dana-Farber Cancer Institute * Group Health Cooperative Harvard Pilgrim Health Care Henry Ford Health System Kaiser Permanente Hawaii Kaiser Permanente Northwest	770	741
Iowa (UIOWA)	University of Iowa *	1145	--
Los Angeles County, California (UCLA)	University of California, Los Angeles * RAND	1054	1064
North Carolina (UNC) 22 eastern/central counties <sup>‡</sup>	University of North Carolina *	--	1013
Northern California (NCCC) 8 counties <sup>‡</sup>	Harvard Medical School * Northern California Cancer Center Kaiser Permanente Northern California	984	1023
Veterans Health Administration (VA) 10 VA hospitals <sup>§</sup>	Department of Veterans Affairs Medical Center, Durham, North Carolina *	945	491
Total		5566	4981

\* Primary grantee organization.

<sup>‡</sup> Alameda, Contra Costa, Sacramento, San Francisco, San Joaquin, San Mateo, Santa Clara, and Solano counties.

<sup>‡</sup> Alamance, Bertie, Chatham, Durham, Edgecombe, Greene, Guilford, Harnett, Johnston, Lee, Lenoir, Martin, Montgomery, Moore, Nash, Orange, Pamlico, Pitt, Randolph, Wake, Wayne, and Wilson counties.

<sup>§</sup> Baltimore, MD; Biloxi, MS; Chicago, IL; Durham, NC; Minneapolis, MN; Nashville, TN; New York, NY; Portland, OR; Temple, TX; and Tucson, AZ.

**Table 3**

Comparison of CanCORS Enrollees to Patients Diagnosed with Lung or Colorectal Cancer in All SEER Regions

	Lung Cancer		Colorectal Cancer	
	SEER % N = 132,758	CanCORS % N = 5,566	SEER % N = 101,706	CanCORS % N = 4,981
Female	46.6	41.0	49.3	47.4
Race				
Hispanic	5.3	4.8	8.5	8.1
African American	9.8	11.8	9.9	14.8
Asian	4.7	3.6	6.3	5.5
Age (yrs)				
<50	5.3	5.1	9.3	11.2
50-75	61.6	68.1	53.4	59.4
>75	33.1	26.9	37.3	29.4
Stage				
I	17.8	21.0	24.5	21.1
II	3.7	6.8	25.2	24.6
III	25.2	24.8	23.5	24.7
IV	41.1	38.1	18.1	16.9

**Table 4**

Comparison of CanCORS Enrollees to Patients Diagnosed with Lung or Colorectal Cancer in Specific SEER Regions. Sample sizes refer to the CanCORS cohorts.

Region	Characteristic	Lung Cancer		Colorectal Cancer	
		SEER	CanCORS	SEER	CanCORS
Seattle (Lung n=221, Colorectal n=204)	% Female	48.9	53.1	48.8	52.6
	% African-American	2.9	2.5	3.2	1.8
	% > 75 years	35.3	33.7	37.4	32.5
Detroit (Lung n=184, Colorectal n=141)	% Female	47.5	49.5	50.2	51.8
	% African-American	22.6	29.4	25.0	34.8
	% > 75 years	31.8	26.1	37.7	21.3
Hawaii (Lung n=78, Colorectal n=90)	% Female	41.5	51.3	45.8	46.7
	% African-American	1.4	2.2	-	-
	% > 75 years	33.4	19.2	34.7	22.2
Northern California (Lung n=984, Colorectal n=1023)	% Female	49.1	51.6	50.2	53.8
	% African-American	9.7	8.6	7.7	6.5
	% > 75 years	37.0	29.2	37.9	30.1
Los Angeles (Lung n=1054, Colorectal n=1064)	% Female	47.1	52.4	50.1	36.3
	% African-American	15.1	10.6	13.1	9.2
	% > 75 years	35.6	28.2	36.6	31.8
Iowa (Lung n=1145)	% Female	42.7	44.3	-	-
	% African-American	1.7	1.1	-	-
	% > 75 years	34.6	29.4	-	-