CrossMark

ORIGINAL PAPER

The Cancer Research Network: a platform for epidemiologic and health services research on cancer prevention, care, and outcomes in large, stable populations

Jessica Chubak[1] · Rebecca Ziebell[1] · Robert T. Greenlee[2] · Stacey Honda[3] ·
Mark C. Hornbrook[4] · Mara Epstein[5] · Larissa Nekhlyudov[6] · Pamala A. Pawloski[7] ·
Debra P. Ritzwoller[8] · Nirupa R. Ghai[9] · Heather Spencer Feigelson[8] ·
Heather A. Clancy[10] · V. Paul Doria-Rose[11] · Lawrence H. Kushi[10]

Abstract

Purpose The ability to collect data on patients for long periods prior to, during, and after a cancer diagnosis is critical for studies of cancer etiology, prevention, treatment, outcomes, and costs. We describe such data capacities within the Cancer Research Network (CRN), a cooperative agreement between the National Cancer Institute (NCI) and organized health care systems across the United States.

Methods Data were extracted from each CRN site's virtual data warehouse using a centrally written and locally executed program. We computed the percent of patients continuously enrolled ≥ 1, ≥ 5, and ≥ 10 years before cancer diagnosis in 2012–2015 (year varied by CRN site). To describe retention after diagnosis, we computed the cumulative percentages enrolled, deceased, and disenrolled each year after the diagnosis for patients diagnosed in 2000.

Results Approximately 8 million people were enrolled in ten CRN health plans on December 31, 2014 or 2015 (year varied by CRN site). Among more than 30,000 recent cancer diagnoses, 70 % were enrolled for ≥ 5 years and 56 % for ≥ 10 years before diagnosis. Among 25,274 cancers diagnosed in 2000, 28 % were still enrolled in 2010, 45 % had died, and 27 % had disenrolled from CRN health systems.

Conclusions Health plan enrollment before cancer diagnosis was generally long in the CRN, and the proportion of patients lost to follow-up after diagnosis was low. With long enrollment histories among cancer patients pre-diagnosis and low post-diagnosis disenrollment, the CRN provides an excellent platform for epidemiologic and

✉ Jessica Chubak
chubak.j@ghc.org

1 Group Health Research Institute, 1730 Minor Avenue, Suite 1600, Seattle, WA 98101, USA

2 Marshfield Clinic Research Foundation, Marshfield, WI, USA

3 Hawaii Permanente Group, Kaiser Permanente Hawaii, Honolulu, HI, USA

4 Kaiser Permanente Center for Health Research, Portland, OR, USA

5 Department of Medicine, Meyers Primary Care Institute, University of Massachusetts Medical School, Worcester, MA, USA

6 Department of Population Medicine, Harvard Pilgrim Health Care Institute, Harvard Medical School, Boston, MA, USA

7 HealthPartners Institute, Minneapolis, MN, USA

8 Kaiser Permanente Colorado, Institute for Health Research, Denver, CO, USA

9 Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, CA, USA

10 Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA

11 Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD, USA

🌀 Springer

health services research on cancer incidence, outcomes, and costs.

Introduction

Clinicians, researchers, and funders are increasingly recognizing the importance of conducting cancer research outside of the highly selected populations of controlled trials and specialty, referral-based clinics [1–5]. Research in community-based settings can provide valuable information on cancer incidence, survival, and costliness, as well as on the effectiveness of cancer-related prevention, screening, treatment, and health care services as delivered in practice.

Studies that can advance knowledge of cancer risk factors, preventive services, treatment effectiveness, and outcomes in "real-world" health care settings include: retrospective studies of cancer etiology with adequate periods for exposure assessment; comparative effectiveness analyses of therapeutic and health care delivery interventions; pragmatic cancer treatment trials; and prospective and retrospective cohort studies of associations between post-diagnosis exposures and cancer prognosis. Research capabilities that enhance the validity of such studies include:

1. The ability to identify and enumerate the underlying population giving rise to cases (i.e., "the denominator") for calculating and comparing incidence rates across exposure groups, and, in case control studies, for selecting controls from the same population from which cases arose.
2. The ability to assess exposures many years before a cancer diagnosis to study factors whose association with cancer may have long induction/latent periods.
3. The ability to follow cancer patients for many years after diagnosis to study clinical outcomes, long-term effects of treatment, and post-diagnosis exposures.

Integrated health care delivery systems are particularly well-suited to conduct health and health care outcomes research in community care settings, and their success in doing so depends, in part, on having the capabilities described above. This manuscript illustrates the Cancer Research Network's (CRN's) capacity to support "real-world" epidemiologic and health services research in the CRN. Specifically, the purpose of this manuscript is to describe patients' enrollment in CRN health care systems before and after their cancer diagnoses—capacities crucial

for research on cancer risk factors, prevention, outcomes, and costs.

Methods

Setting

The CRN, established in 1999 [6, 7], is a cooperative agreement between the National Cancer Institute (NCI) and integrated non-profit health care delivery systems. It currently consists of eight supported and four affiliate research centers within these integrated health care systems (http://crn.cancer.gov/). The goal of the CRN is to conduct and foster public domain research across the cancer care continuum, from primary prevention to treatment, survivorship, and end-of-life care. The CRN has been a long-time contributor to epidemiologic and health services research on cancer etiology, prevention, screening, costs, utilization, communications, recurrence, effects of treatment, and survivorship (for examples, see references [8–17]).

In addition to having research scientists embedded in these health care settings, one of the CRN's most important resources is its standardized, distributed data network. The distributed data network is now managed by the Health Care Systems Research Network (HCSRN; formerly HMO Research Network (HMORN)). The HCSRN is an umbrella organization of public domain research centers embedded in or affiliated with health care systems; it encompasses several collaborative research networks including the CRN. The distributed data network supports the Virtual Data Warehouse (VDW), a standardized de-centralized data model that was developed for research use, in which each HCSRN site maintains its data locally, but a programmer at one site can write data-extraction and analysis programs that can be run at other sites [7, 18]. The VDW includes tables that capture information on demographics of health plan enrollees and/or health system patients, health plan enrollment, health care encounters, diagnoses, procedures, pharmacy fills, vital signs, social history, vital status, and laboratory values. Within the VDW, the tumor table (also known as the Virtual Tumor Registry [VTR]) contains data consistent with the North American Association of Central Cancer Registries standards (www.naaccr.org/) [13]. Tumor data are obtained from manual review of cancer patients' medical records by trained medical record abstractors.

All eight funded CRN sites contributed data to this analysis. Sites, listed by research center and associated health plan, are: Meyers Primary Care Institute, Reliant Medical Group and Fallon Community Health Plan (FCHP), Massachusetts; Group Health Research Institute, Group Health (GH), Washington; Department of Public

Health Sciences, Henry Ford Health System and Health Alliance Plan (HFHS), Michigan; Institute for Health Research, Kaiser Permanente Colorado (KPCO); Center for Health Research-Hawaii, Kaiser Permanente Hawaii (KPHI); Center for Health Research, Kaiser Permanente Northwest (KPNW), Oregon; Division of Research, Kaiser Permanente Northern California (KPNC); and Marshfield Clinic Research Foundation, Marshfield Clinic Health System and Security Health Plan (MCHS), Wisconsin. Three affiliate sites also provided data (HealthPartners Institute, HealthPartners (HP), Minnesota; Harvard Pilgrim Health Care (HPHC), Department of Population Medicine, Harvard Medical School, Massachusetts; Department of Research and Evaluation, Kaiser Permanente Southern California (KPSC)).

Data collection

In July 2016, data were extracted from the tumor, enrollment, demographics, and death tables in the VDW via a SAS program developed by the CRN Informatics Core. The program was designed to query standard VDW data elements and required only minimal site-specific customization (e.g., to specify the path to local site data storage locations). Program output consisted of tabulations (for review by local site investigators) and aggregated SAS data sets, the latter of which were returned to and compiled by the Informatics Core programmer at Group Health. Ten of 11 participating sites received the SAS program and returned results via the CRN installation of PopMedNet™, a software application designed to facilitate data sharing activities in distributed networks; the remaining site did so via secure email. As the lead institution, KPNC's IRB approved the aforementioned CRN infrastructure activities. CRN funded and affiliate sites either ceded IRB authority to KPNC's IRB (6 sites) or submitted materials to their own institution for review and approval (1 site) or exemption (2 sites); one site's IRB determined these data were preparatory to research and IRB approval was not required. Additional IRB approval was received from the Massachusetts State Tumor Registry to use cancer data from the two CRN sites in Massachusetts that maintain VDW tumor tables.

Each CRN site contributed population demographic characteristics (gender, age, race, and ethnicity) and insurance coverage data as of 12/31/2014 or 12/31/2015, depending on when their data were last refreshed. Each CRN site also sent the number of new malignant cancer diagnoses by year of diagnosis, age at diagnosis (<13, 13–18, 19–39, 40–49, 50–59, 60–64, 65–69, 70–79, ≥80 years), and cancer site. Cancer site was categorized based on the Surveillance, Epidemiology, and End Results

program's (SEER's) ICD-O3/WHO 2008 Site Recode definition [19]. Only cases who were enrolled at a CRN site at the time of diagnosis or whose diagnosis was within 45 days after disenrollment or 60 days before enrollment were included [20]. Persons with multiple primary tumors diagnosed during a given year were counted only once per year but were not de-duplicated across diagnosis years.

Analysis

For cancers diagnosed in the most recent year of available data (2012 at one site, 2013 at one site, and 2014 at six sites, and 2015 at two sites), we computed the percent continuously enrolled for ≥1, ≥5, and ≥10 years before diagnosis. We stratified analyses by CRN site, cancer site, and age at diagnosis.

In analyses of disenrollment and death after cancer diagnosis, we focused on cases diagnosed in 2000 because all CRN sites could provide at least 10 years of follow-up. As of each 12-month interval after diagnosis, we computed the following cumulative percentages: continuously enrolled and presumed alive (no indicator of death); continuously enrolled until death; and disenrolled. These categories were mutually exclusive and added to 100 %. Continuous enrollment was defined as no gaps in enrollment >60 days. We conducted stratified analyses by CRN site, cancer site, and age at diagnosis.

Because KPSC restricted their submitted data to a subset of cancer types (breast [$n = 1,968$], colon and rectum [$n = 1,166$], prostate [$n = 2,171$], lymphoma [$n = 528$], lung and bronchus [$n = 1,165$], and urinary bladder [$n = 210$]), we excluded KPSC from our analyses. Including these data would have skewed our overall proportions toward specific cancers. However, we verified in sensitivity analyses restricted to breast, colon and rectum, prostate, lymphoma, lung and bronchus, and urinary bladder that including KPSC data would not meaningfully have altered cancer-specific results.

No statistical testing was performed as our goal was to characterize available data, not to draw inferences or test hypotheses. Analyses were conducted in SAS® version 9.3 (Cary, North Carolina) and Stata® version 12 (College Station, Texas).

Results

Approximately 8 million people were enrolled in participating CRN health plans (Table 1) as of 12/31/2014 or 12/31/2015 (date varied by CRN site). About half of enrollees were aged 40 and older. Race was known in

Table 1 Selected characteristics of Cancer Research Network health plan enrollees (n = 8,039,569)

	Percent of total
Sex	
Female	51.8
Male	48.2
Age (years)	
<13	13.3
13–18	8.4
19–39	28.0
40–49	13.8
50–59	15.1
60–69	12.3
70–79	5.8
≥80	3.3
Race	
American Indian or Alaska Native	0.3
Asian	10.8
Black or African American	5.4
Native Hawaiian or Other Pacific Islander	0.9
White	46.2
Multiple races	2.0
Other	0.6
Unknown	33.9
Ethnicity	
Hispanic	11.9
Non-Hispanic	19.4
Unknown	68.7
Cancer Research Network site	
Group Health (as of 12/31/2015)	7.3
Henry Ford Health System (as of 12/31/2014)	1.8
Harvard Pilgrim Health Care (as of 12/31/2014)	10.6
HealthPartners (as of 12/31/2015)	12.0
Kaiser Permanente Colorado (as of 12/31/2015)	7.7
Kaiser Permanente Hawaii (as of 12/31/2015)	3.0
Kaiser Permanente Northern California (as of 12/31/2015)	46.7
Kaiser Permanente Northwest (as of 12/31/2015)	6.4
Marshfield Clinic Health System (as of 12/31/2014)	2.1
Meyers-Reliant-Fallon Health (as of 12/31/2014)	2.3
Insurance type[a]	
Commercial	85.1
Private pay	10.7
Self-funded	12.5
Medicaid	7.3
Medicare	14.8
State-subsidized	0.9

[a] Categories are not mutually exclusive

66.1 % of the enrollees; approximately 30 % of persons with known race were non-white. Demographic characteristics differed across CRN sites (Online Resource S1), primarily with respect to race, ethnicity, and insurance.

Among more than 30,000 cancers diagnosed in the most recently available year of data, the majority occurred in patients aged 60 and older (Table 2). The most common were cancers of the breast (17.8 %), lung, bronchus, and pleura (10.4 %), prostate (9.8 %), colon and rectum (8.7 %), and skin (excluding basal cell and squamous cell carcinoma) (7.8 %) (Table 3). However, there were large numbers of even relatively uncommon cancers, such as those of the ovary (n = 424) or urinary bladder (n = 740). The relative incidence of different cancers within the CRN generally matched SEER program data (Fig. 1), except that prostate, bladder, and lung and bronchus cancers accounted for a smaller percentage of cases within the CRN and breast and melanoma accounted for higher percentages. Of note, CRN data included both male and female breast cancers, while SEER data were limited to female breast cancers.

Most patients were continuously enrolled in their health plan for many years before cancer diagnosis (Table 4). More than two thirds were enrolled for ≥5 years and more than half for ≥10 years before diagnosis (69.5 and 55.6 %, respectively). Of the specific cancer sites examined, patients with thyroid cancer had the lowest ≥10 year enrollment. Prior enrollment varied by CRN site (Online Resource S2), with the percent enrolled for ≥10 years ranging from 30.1 to 67.1 %. Age at diagnosis was also related to pre-cancer enrollment (Online Resource S3), with patients age 40 years and older having longer enrollment before diagnosis and persons under 40 having a shorter pre-diagnosis enrollment duration.

Among cancers diagnosed in the year 2000 (n = 25,274), the majority either stayed enrolled (28 %) or died (45 %) within 10 years (Fig. 2). Only 27 % disenrolled within 10 years. Nearly all CRN sites had similar patterns of disenrollment in the 5 years following cancer diagnosis (Online Resource S4). However, occurrences of death and disenrollment differed by cancer site (Online Resource S5). Disenrollment tended to be lower among people with cancers that usually affect older persons or are more rapidly fatal (e.g., lung and bronchus) and higher among less fatal cancers affecting younger persons (e.g., thyroid). Patients under age 40 years at cancer diagnosis were more likely to disenroll than older patients, who were more likely to stay enrolled or, in the case of patients aged 70 years or more, to have died during the observation period (Online Resource S6).

Table 2 Age and Cancer Research Network site of cancers diagnosed in most recent year of available data ($n = 30,420$)

	n	%
Age at diagnosis (years)		
<13	156	0.5
13–18	87	0.3
19–39	1,445	4.8
40–49	2,346	7.7
50–59	5,929	19.5
60–69	9,217	30.3
70–79	6,727	22.1
≥80	4,513	14.8
Cancer Research Network site		
Group Health	2,693	8.9
Henry Ford Health System	691	2.3
Harvard Pilgrim Health Care	2,628	8.6
HealthPartners	712	2.3
Kaiser Permanente Colorado	2,578	8.5
Kaiser Permanente Hawaii	1,200	3.9
Kaiser Permanente Northern California	15,660	51.5
Kaiser Permanente Northwest	2,767	9.1
Marshfield Clinic Health System	908	3.0
Meyers-Reliant-Fallon Health	583	1.9

Diagnosed in 2015: Group Health, Kaiser Permanente Hawaii; 2014: Henry Ford Health System, HealthPartners, Kaiser Permanente Colorado, Kaiser Permanente Northern California, Kaiser Permanente Northwest, Marshfield Clinic Health System; 2013: Harvard Pilgrim Health Care; 2012: Meyers-Reliant-Fallon Health

Discussion

Combined, CRN sites have a large, diverse population with relatively high numbers of common cancers as well as certain rare ones. We observed that about half of incident cancer cases had at least 10 years of enrollment before diagnosis, and only one in five were lost to follow-up in the first 10 years after diagnosis. Some variation by CRN site was observed, as shown in Online Resource S2 and Online Resource S4. Our overall results are consistent with a report from a decade ago showing high retention rates among cancer survivors in a subset of CRN sites [20]. In the current analysis, we noted differences in loss to follow-up by cancer type, but these were generally consistent with what we would expect based on age at diagnosis and known survival statistics [21]. Stratification of length of enrollment by CRN site, cancer site, and age at diagnosis provides important information for researchers planning studies within the CRN. Estimates in this manuscript can be used to plan future, large-scale studies of cancer incidence and outcomes within the CRN. For example, these statistics can help investigators select CRN sites for their research and conduct power analyses. Of note, when data

from all CRN sites are combined, the average is weighted toward larger sites like KPNC. We, therefore, provided CRN site-specific estimates in Online Resources S2 and S4. However, investigators planning studies may need more granular information (e.g., stratified by cancer site and CRN site), which can be obtained through the Preparatory-to-Research process (https://crn.cancer.gov/collaboration/process.html).

All data in this manuscript were obtained using a distributed SAS program written at one site and distributed to the other sites. This process illustrates one of the main efficiencies in conducting research within the CRN. Researchers should be aware, however, of several limitations in using CRN data. First, not all CRN sites have data available over the same years. Second, most CRN sites experience a lag in the availability of tumor data. If newly diagnosed cases are needed for a study, researchers should take this into account when selecting CRN sites to be included in their studies. Other methods for quickly identifying cases (e.g., via pathology reports) may still be possible but employing them may require different approaches at each CRN site. A third important consideration is that some data are collected differently across CRN sites. For example, at some CRN sites (i.e., KPNC, HP), everyone who is not verified as Hispanic is classified as unknown ethnicity, whereas at other sites, persons may be classified as non-Hispanic. Fourth, the catchment area for a CRN site's VDW tumor table may not always include the entire health plan enrollment. For example, Group Health's VDW tumor table is populated for Group Health enrollees who reside in the western Washington SEER catchment area (about 80 % of enrollees). For HP, cancer diagnoses are not available for health plan members who are not also patients of their medical group (about half of enrollees). These subtleties highlight the importance of incorporating local expertise and knowledge of how each site's VDW is populated when conducting multi-site research.

This manuscript illustrates two important strengths of the CRN that make it an excellent setting for conducting population-based research along the cancer care continuum from prevention to survivorship and end of life: long pre-diagnosis enrollment and low loss to follow-up post-diagnosis. Pre-diagnosis information is important for studies of cancer incidence in relation to exposures like medication use (e.g., Boudreau et al. [22]) or screening (e.g., Doubeni et al. [11]). Low rates of disenrollment after cancer diagnosis are important for studies of outcomes requiring extended follow-up periods, such as second malignancies (e.g., Clough-Gorr et al. [23]), late-effects of cancer (e.g., Bowles et al. [15]), and health care utilization in survivors (e.g., Buist et al. [24]).

Because the CRN is based on community practice, it is positioned to facilitate multi-site research on comparative

Table 3 Types of cancer diagnosed in most recent year of available data ($n = 30{,}420$)

Cancer type	n	%	Cancer type	n	%	Cancer type	n	%
Oral cavity and pharynx	880	2.9	Gallbladder	66	0.2	*Urinary system*	1,805	5.9
Lip	85	0.3	Other biliary	120	0.4	Urinary bladder	740	2.4
Tongue	295	1	Pancreas	828	2.7	Kidney and renal pelvis	1,005	3.3
Salivary gland	88	0.3	Retroperitoneum	33	0.1	Ureter	42	0.1
Floor of mouth	27	0.1	Peritoneum, omentum, and mesentery	55	0.2	Other urinary organs	18	0.1
Gum and other mouth	111	0.4	Other digestive organs	28	0.1	*Eye and orbit*	64	0.2
Nasopharynx	49	0.2	*Respiratory system*	3,366	11.1	*Brain and other nervous system*	450	1.5
Tonsil	159	0.5	Nose, nasal cavity and middle ear	46	0.2	Brain	426	1.4
Oropharynx	29	0.1	Larynx	138	0.5	Cranial nerves other nervous system	24	0.1
Hypopharynx	28	0.1	Lung, bronchus, and pleura	3,171	10.4	*Endocrine system*	948	3.1
Other oral cavity and pharynx	9	<0.1	Trachea, mediastinum and other respiratory organs	11	<0.1	Thyroid	893	2.9
Digestive system	5,492	18.1	*Bones and joints*	49	0.2	Other endocrine including thymus	55	0.2
Esophagus	314	1	*Soft tissue including heart*	248	0.8	*Lymphoma*	1,637	5.4
Stomach	464	1.5	*Skin excluding basal and squamous*	2,381	7.8	Hodgkin lymphoma	189	0.6
Small intestine	149	0.5	Melanoma of the skin	2,231	7.3	Non-hodgkin lymphoma	1,448	4.8
Colon and rectum	2,644	8.7	Other non-epithelial skin	150	0.5	Nodal	897	3
Colon excluding rectum	1,872	6.2	*Breast*	5,417	17.8	Extranodal	551	1.8
Cecum	426	1.4	*Female genital system*	2,085	6.9	*Myeloma*	441	1.5
Appendix	80	0.3	Cervix uteri	173	0.6	*Leukemia*	929	3.1
Ascending colon	348	1.1	Corpus and uterus, NOS	1,296	4.3	Lymphocytic leukemia	517	1.7
Hepatic flexure	98	0.3	Corpus uteri	1,270	4.2	Acute lymphocytic	101	0.3
Transverse colon	189	0.6	Uterus, NOS	26	0.1	Chronic lymphocytic	388	1.3
Splenic flexure	58	0.2	Ovary	424	1.4	Other lymphocytic	28	0.1
Descending colon	99	0.3	Vagina	33	0.1	Myeloid and monocytic leukemia	382	1.3
Sigmoid colon	495	1.6	Vulva	104	0.3	Acute myeloid	267	0.9
Large intestine, NOS	79	0.3	Other female genital organs	55	0.2	Acute monocytic	9	<0.1
Rectum and rectosigmoid junction	772	2.5	*Male genital system*	3,245	10.7	Chronic myeloid	96	0.3
Rectosigmoid junction	139	0.5	Prostate	2,990	9.8	Other myeloid/monocytic	10	<0.1
Rectum	633	2.1	Testis	219	0.7	Other leukemia	30	0.1
Anus, anal canal and anorectum	153	0.5	Penis	24	0.1	Other acute leukemia	9	<0.1
Liver and intrahepatic bile duct	638	2.1	Other male genital organs	12	<0.1	Aleukemic, subleukemic and NOS	21	0.1
Liver	537	1.8				*Mesothelioma*	71	0.2
Intrahepatic bile duct	101	0.3				*Kaposi Sarcoma*	32	0.1
						Miscellaneous	868	2.9
						Invalid per SEER site recode	12	<0.1

Diagnosed in 2015: Group Health, Kaiser Permanente Hawaii; 2014: Henry Ford Health System, HealthPartners, Kaiser Permanente Colorado, Kaiser Permanente Northern California, Kaiser Permanente Northwest, Marshfield Clinic Health System; 2013: Harvard Pilgrim Health Care; 2012: Meyers-Reliant-Fallon

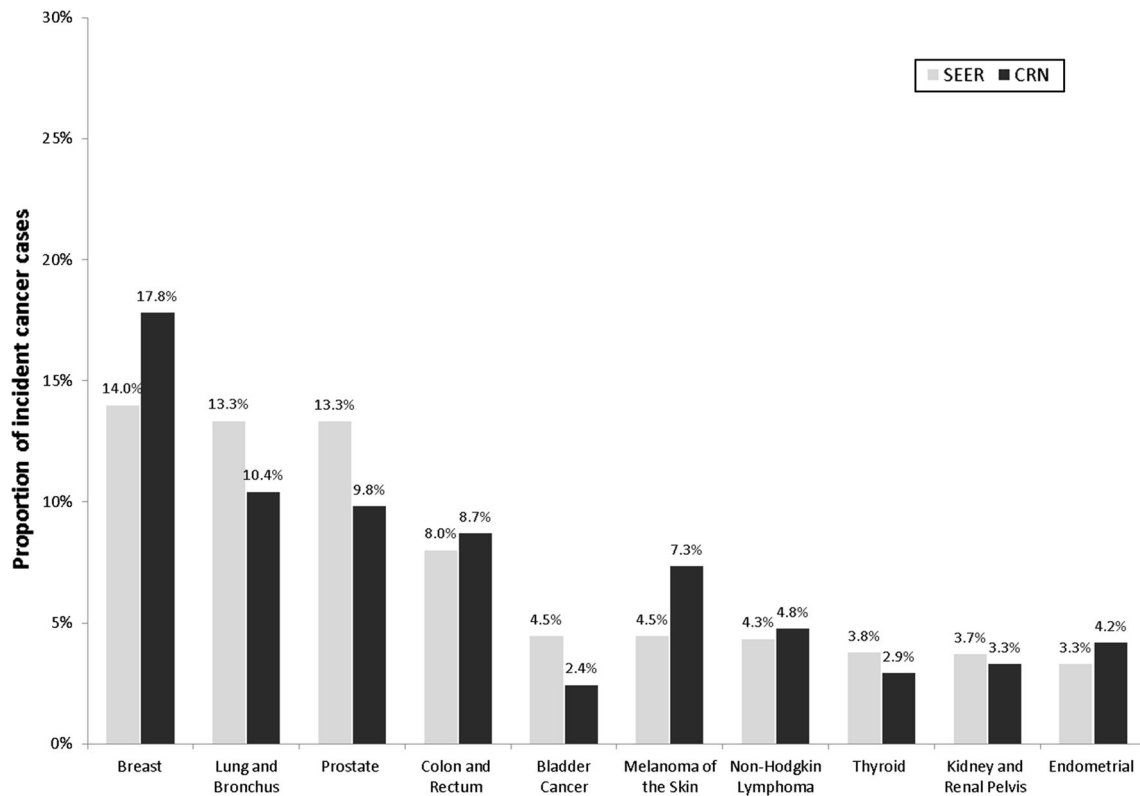*NOS* not otherwise specified; *SEER* surveillance, epidemiology, and end results

Fig. 1 Proportional incidence of common cancer types in the SEER program compared to the Cancer Research Network. *Note*: Breast cancer cases from SEER are in females only but are for both males and females in CRN data. SEER data are from 2015. CRN data are from the most recent year available (Table 2). *CRN* Cancer Research Network; *SEER* Surveillance, Epidemiology, and End Results

Table 4 Length of continuous enrollment before cancer diagnosis, overall and by cancer site

Cancer type	Total	≥1 year pre-diagnosis		≥5 years pre-diagnosis		≥10 years pre-diagnosis	
	n	*n*	*%*	*n*	*%*	*n*	*%*
Total	30,420	26,888	88.4	21,153	69.5	16,912	55.6
Prostate	2,990	2,619	87.6	1,998	66.8	1,560	52.2
Breast	5,417	4,728	87.3	3,686	68.1	2,908	53.7
Lung and bronchus	3,169	2,899	91.5	2,361	74.5	1,956	61.7
Colon and rectum	2,644	2,289	86.6	1,814	68.6	1,451	54.9
Melanoma of the skin	2,231	1,986	89.0	1,580	70.8	1,261	56.5
Urinary bladder	740	680	91.9	557	75.3	476	64.3
Non-hodgkin lymphoma	1,448	1,316	90.9	1,069	73.8	872	60.2
Kidney and renal pelvis	1,005	889	88.5	693	69.0	537	53.4
Thyroid	893	740	82.9	501	56.1	353	39.5
Corpus and uterus, NOS	1,296	1,123	86.7	919	70.9	740	57.1
Other	8,587	7,619	88.7	5,975	69.6	4,798	55.9

Diagnosed in 2015: Group Health, Kaiser Permanente Hawaii; 2014: Henry Ford Health System, HealthPartners, Kaiser Permanente Colorado, Kaiser Permanente Northern California, Kaiser Permanente Northwest, Marshfield Clinic Health System; 2013: Harvard Pilgrim Health Care; 2012: Meyers-Reliant-Fallon Health

NOS not otherwise specified

effectiveness, multi-morbidity, real-world adherence to screening and treatment regimens, and other topics that are difficult to address in academic cancer settings or by randomized controlled trials. The VDW does not contain information on all exposures of interest (e.g., diet), but is an excellent source of data on health care utilization,

Fig. 2 Retention after cancer diagnosis in 2000 ($n = 25{,}274$), cumulative by year since diagnosis
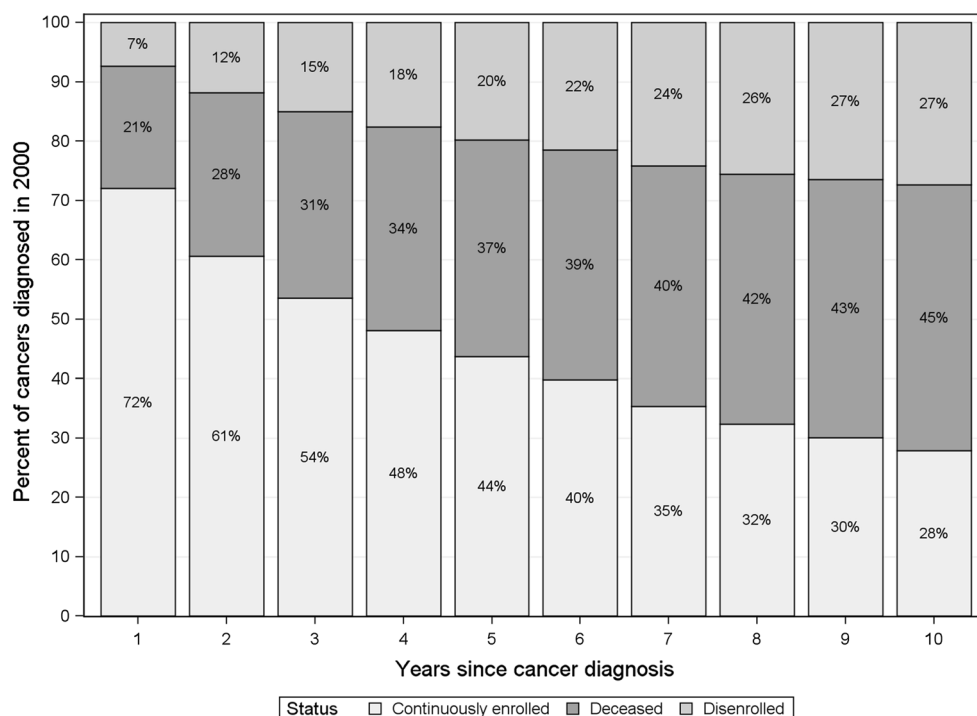
diagnoses, and medication fills (Online Resource S7). Unlike claims datasets, CRN sites have access to clinical measures such as anthropometrics and laboratory results. To benefit most from the CRN as a research platform, researchers need to understand not only the wide range of available data [18] but also the characteristics that make it well-suited to population-based cancer research.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval For this type of study formal consent is not required.

Human and animal rights statement This article does not contain any studies with animals performed by any of the authors.

References

1. Lyman GH, Levine M (2012) Comparative effectiveness research in oncology: an overview. J Clin Oncol. doi:10.1200/JCO.2012.45.9792
2. Carpenter WR, Meyer A-M, Abernethy AP, Sturmer T, Kosorok MR (2012) A framework for understanding cancer comparative effectiveness research data needs. J Clin Epidemiol 65(11):1150–1158. doi:10.1016/j.jclinepi.2012.06.005
3. Abernethy AP, Etheredge LM, Ganz PA, Wallace P, German RR, Neti C, Bach PB, Murphy SB (2010) Rapid-learning system for cancer care. J Clin Oncol 28(27):4268–4274. doi:10.1200/jco.2010.28.5478
4. National Cancer Institute Introduction to PROSPR. U.S. Department of Health and Human Services. http://healthcaredelivery.cancer.gov/prospr/introduction.html. Accessed 20 Nov 2015
5. National Cancer Institute NCI Community Oncology Research Program (NCORP) Home Page. U.S. Department of Health and Human Services. http://ncorp.cancer.gov/about/. Accessed 20 Nov 2015
6. Wagner EH, Greene SM, Hart G, Field TS, Fletcher S, Geiger AM, Herrinton LJ, Hornbrook MC, Johnson CC, Mouchawar J, Rolnick SJ, Stevens VJ, Taplin SH, Tolsma D, Vogt TM (2005) Building a research consortium of large health systems: the Cancer Research Network. J Natl Cancer Inst Monogr 35:3–11. doi:10.1093/jncimonographs/lgi032
7. Hornbrook MC, Hart G, Ellis JL, Bachman DJ, Ansell G, Greene SM, Wagner EH, Pardee R, Schmidt MM, Geiger A, Butani AL,

Field T, Fouayzi H, Miroshnik I, Liu L, Diseker R, Wells K, Krajenta R, Lamerato L, Neslund Dudas C (2005) Building a virtual cancer research organization. J Natl Cancer Inst Monogr 35:12–25. doi:10.1093/jncimonographs/lgi033

8. Nekhlyudov L, Greene SM, Chubak J, Rabin B, Tuzzio L, Rolnick S, Field TS (2013) Cancer Research Network: using integrated healthcare delivery systems as platforms for cancer survivorship research. J Cancer Surviv 7(1):55–62. doi:10.1007/s11764-012-0244-8

9. Geiger AM, Buist DS, Greene SM, Altschuler A, Field TS (2008) Survivorship research based in integrated healthcare delivery systems: the Cancer Research Network. Cancer 112(11 Suppl):2617–2626. doi:10.1002/cncr.23447

10. Selby JV, Friedman GD, Quesenberry CP Jr, Weiss NS (1992) A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. N Engl J Med 326(10):653–657. doi:10.1056/NEJM199203053261001

11. Doubeni CA, Weinmann S, Adams K, Kamineni A, Buist DS, Ash AS, Rutter CM, Doria-Rose VP, Corley DA, Greenlee RT, Chubak J, Williams A, Kroll-Desrosiers AR, Johnson E, Webster J, Richert-Boe K, Levin TR, Fletcher RH, Weiss NS (2013) Screening colonoscopy and risk for incident late-stage colorectal cancer diagnosis in average-risk adults: a nested case-control study. Ann Intern Med 158(5 Pt 1):312–320. doi:10.7326/0003-4819-158-5-201303050-00003

12. Geiger AM, Thwin SS, Lash TL, Buist DS, Prout MN, Wei F, Field TS, Ulcickas Yood M, Frost FJ, Enger SM, Silliman RA (2007) Recurrences and second primary breast cancers in older women with initial early-stage disease. Cancer 109(5):966–974

13. Ritzwoller DP, Carroll NM, Delate T, Hornbrook MC, Kushi L, Bowles EJ, Loggers ET, Menter A (2014) Comparative effectiveness of adjunctive bevacizumab for advanced lung cancer: the cancer research network experience. J Thorac Oncol 9(5):692–701. doi:10.1097/JTO.0000000000000127

14. Loggers ET, Fishman PA, Peterson D, O'Keeffe-Rosetti M, Greenberg C, Hornbrook MC, Kushi LH, Lowry S, Ramaprasan A, Wagner EH, Weeks JC, Ritzwoller DP (2014) Advanced imaging among health maintenance organization enrollees with cancer. J Oncol Pract 10(4):231–238. doi:10.1200/JOP.2013.001258

15. Bowles EJA, Wellman R, Feigelson HS, Onitilo AA, Freedman AN, Delate T, Allen LA, Nekhlyudov L, Goddard KAB, Davis RL, Habel LA, Yood MU, Mccarty C, Magid DJ, Wagner EH, Pharmacovigilance Study Team (2012) Risk of heart failure in breast cancer patients after anthracycline and trastuzumab treatment: a retrospective cohort study. J Nat Cancer Inst 104(17):1293–1305. doi:10.1093/jnci/djs317

16. Prouty CD, Mazor KM, Greene SM, Roblin DW, Firneno CL, Lemay CA, Robinson BE, Gallagher TH (2014) Providers' perceptions of communication breakdowns in cancer care. J Gen Intern Med 29(8):1122–1130. doi:10.1007/s11606-014-2769-1

17. Geiger AM, Castellino SM, Tooze JA, Altschuler A, Month S, Mertens AC, Nekhlyudov L, Greene SM (2013) Potential role of community-based healthcare system data in research on survivors of adolescent and young adult cancer. J Adolesc Young Adult Oncol 2(2):53–58. doi:10.1089/jayao.2012.0033

18. Ross TR, Ng D, Brown JS, Pardee R, Hornbrook MC (2014) The HMO Research Network Virtual Data Warehouse: A Public Data Model to Support Collaboration. EGEMS (Wash DC) 2 (1):Article 2. doi: 10.13063/2327-9214.1049

19. Surveillance Epidemiology and End Results (SEER) Program Site Recode ICD-O-3/WHO 2008 Definition. U.S. Department of Health and Human Services. http://seer.cancer.gov/siterecode/icdo3_dwhoheme/index.html. Accessed 5 Jan 2015

20. Field TS, Cernieux J, Buist D, Geiger A, Lamerato L, Hart G, Bachman D, Krajenta R, Greene S, Hornbrook MC, Ansell G, Herrinton L, Reed G (2004) Retention of enrollees following a cancer diagnosis within health maintenance organizations in the Cancer Research Network. J Natl Cancer Inst 96(2):148–152

21. Surveillance Epidemiology and End Results Table 1.5 Age-Adjusted SEER Incidence and U.S. Death Rates and 5-Year Relative Survival (Percent) By Primary Cancer Site, Sex and Time Period. National Cancer Institute. http://seer.cancer.gov/csr/1975_2012/results_merged/topic_survival.pdf. Accessed 10 July 2015

22. Boudreau DM, Yu O, Miglioretti DL, Buist DSM, Heckbert SR, Daling JR (2007) Statin use and breast cancer risk in a large population-based setting. Cancer Epidemiol Biomarkers Prev 16(3):416–421. doi:10.1158/1055-9965.epi-06-0737

23. Clough-Gorr KM, Thwin SS, Bosco JLF, Silliman RA, Buist DSM, Pawloski PA, Quinn VP, Prout MN (2012) Incident malignancies among older long-term breast cancer survivors and an age-matched and site-matched nonbreast cancer comparison group over 10 years of follow-up. Cancer 119(8):1478–1485. doi:10.1002/cncr.27914

24. Buist DS, Bosco JL, Silliman RA, Gold HT, Field T, Yood MU, Quinn VP, Prout M, Lash TL, Breast Cancer Outcomes in Older Women Investigators (2013) Long-term surveillance mammography and mortality in older women with a history of early stage invasive breast cancer. Breast Cancer Res Treat 142(1):153–163. doi:10.1007/s10549-013-2720-x