Breast and colorectal cancer recurrence and progression captured by five U.S. population-based registries: Findings from National Program of Cancer Registries patient-centered outcome research

Trevor D. Thompson[a], **Lori A. Pollack**[a,*], **Christopher J. Johnson**[b], **Xiao-Cheng Wu**[c], **Judy R. Rees**[d], **Mei-Chin Hsieh**[c], **Randi Rycroft**[b,e], **MaryBeth Culp**[f], **Reda Wilson**[a], **Manxia Wu**[a], **Kevin Zhang**[g], **Vicki Benard**[a]

[a]Division of Cancer Prevention and Control, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, United States

[b]Cancer Data Registry of Idaho, Idaho Hospital Association, Boise, ID, United States

[c]Louisiana Tumor Registry, EpidemiologyProgram, School of Public Health, Louisiana State University Health Sciences Center, New Orleans, LA, United States

[d]New Hampshire State Cancer Registry, Department of Epidemiology, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire and Norris Cotton Cancer Center, Lebanon, NH, United States

[e]Colorado Central Cancer Registry, Colorado Department of Public Health and Environment, Denver, CO, United States

[f]American Cancer Society, United States

[g]ICF International, United States

Abstract

[*]Corresponding author at: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division of Cancer Prevention and Control, 4770 Buford Hwy NE, S107-4, Atlanta, GA, 30341, United States. lop5@cdc.gov (L.A. Pollack).

Declaration of Competing Interest
None.

Objectives: Cancer recurrence is a meaningful patient outcome that is not captured in population-based cancer surveillance. This project supported National Program of Cancer Registries central cancer registries in five U.S. states to determine the disease course of all breast and colorectal cancer cases. The aims were to assess the feasibility of capturing disease-free (DF) status and subsequent cancer outcomes and to explore analytic approaches for future studies.

Methods: Data were obtained on 11,769 breast and 6033 colorectal cancer cancers diagnosed in 2011. Registry-trained abstractors reviewed medical records from multiple sources for up to 60 months to determine documented DF status, recurrence, progression and residual disease. We described the occurrence of these patient-centered outcomes along with analytic considerations when determining time-to-event outcomes and recurrence-free survival.

Results: Disease-free status was determined on all but 3.8 % of cancer cases. Among 14,458 cases that became DF, 6.1 % of breast and 13.0 % of colorectal cancer cases had a documented recurrence. Recurrence-free survival varied by stage; for stage II-III cancers at 48 months, 83.2 % of female breast and 69.2 % of colorectal cancer patients were alive without recurrence. The ability to distinguish between progression and residual disease among never disease-free patients limited our ability to examine progression as an outcome.

Conclusions: This study demonstrated that population-based registries given intense support and resources can capture recurrence and offer a generalizable picture of cancer outcomes. Further work on refining definitions, sampling strategies, and novel approaches to capture recurrence could advance the ability of a national cancer surveillance system to contribute to patient-centered outcomes research.

1. Introduction

Due to improvements in early detection and advances in treatment, many patients live years after a cancer diagnosis. Currently, five-year relative survival rates for colorectal and female breast cancers are 63.4 % and 88.6 %, respectively [1]. Clinicians, cancer survivors, and researchers are increasingly interested in understanding recurrence and progression to assess the effectiveness of treatment, evaluate prognosis, and examine outcomes associated with differences in cancer care.

Current knowledge about cancer recurrence and progression is based primarily on the intensive follow-up of select patient populations at specialized settings, such as those enrolled in longitudinal research studies [2–5]. Established criteria for assessing tumor response to treatment, such as Response Evaluation Criteria in Solid Tumors (RECIST) guidelines, are applicable to clinical trials or studies with intensive, protocol-driven tumor re-evaluation, but are not practical for large population-based cancer studies [6]. The American College of Surgeon Commission on Cancer Program collects recurrence through their hospital-based cancer registries, but these data have known limitations and only reflect patients who receive care in Commission on Cancer hospitals [7]. The Centers for Disease Control and Prevention's (CDC) National Program of Cancer Registries (NPCR) and the

National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program collect cancer incidence and first course of treatment through medical record abstraction on a larger, diverse patient population from multiple healthcare settings [8–10]. Generalizable information from these population-based cancer registries could better inform clinicians and patients about treatment effectiveness and prognosis [8,11]. Population-based registries, however, do not routinely collect recurrence, progression or outcomes other than vital status and cause of death because longitudinal follow-up of patients is extremely time and resource intensive [10,11].

In response to a 2009 Institute of Medicine report, the Agency for Healthcare Research and Quality supported a unique opportunity for CDC to enhance cancer data collection for comparative effectiveness research (CER) using population-based NPCR registries [12–14]. Ten state-based central cancer registries received additional funding, training, and technical assistance to review medical records from all healthcare facilities over 12-months beginning from the diagnosis of breast cancer and colorectal cancers. [14]. Five of these registries continued active follow-up and review of medical records up to 60-months to capture cancer recurrence and progression, endpoints of specific interest to cancer-related patient-centered outcome research (PCOR). This is the first time U.S. population-based registries accomplished this type and length of data collection.

The primary objective of this paper is to provide a methodologic overview of the NPCR PCOR project, specifically focusing on the data definitions and completeness of collected data on disease-free status, recurrence and progression. We explored various ways to define time-to-recurrence and suggest approaches and considerations for using NPCR PCOR data in future analyses. Our aims were 1) to provide a foundation and reflection for cancer registries and researchers to collect and analyze recurrence data; and 2) to contribute valuable knowledge about cancer recurrence in a large, generalizable population and provide a useful comparison for prior and future studies.

2. Methods

2.1. Study population

Patients included in the NPCR PCOR study were male and female patients diagnosed in 2011 with in situ or invasive tumors of the breast, colon or rectum (International Classification of Disease for Oncology (ICD-O-3) codes C50.0–C50.9; C18.0–18.9; and C19.9 or C20.9, respectively) residing in the following states at diagnosis: Colorado, Idaho, Louisiana, New Hampshire, or Rhode Island. All persons residing in these states at diagnosis were included in the dataset regardless of where a patient received treatment.

The NPCR Cancer Surveillance System received approval from CDC's Institutional Review Board (IRB) for PCOR data collection. Existing cancer reporting laws and regulations in participating states authorized submission of de-identified information to CDC. Individual states' IRBs either approved the project or determined it exempt as an extension of public health surveillance. No patients were contacted for this project.

2.2. Data collection

Trained abstractors from state cancer registries reviewed and abstracted patient medical records, either onsite or via remote access to electronic medical records, from hospital and non-hospital settings, including physician offices, treatment centers, and radiology facilities for a minimum of 32 months beginning at diagnosis. Chen et al. fully describe data variables, collection and NPCR CER study methodology [14]. State registries also linked data with the state death file and National Death Index (NDI) through 2016 deaths to update vital status and obtain death date and cause of death.

In addition to the routinely collected cancer registry variables on patient demographics, tumor characteristics, cancer stage, and first course of treatment [14–16], abstractors recorded any subsequent treatment within the first 12 months of diagnosis [14,15] and the following patient outcomes (Appendix):

- **Disease-free/No Evidence of Disease**: Patient had documentation of being disease-free or having no evidence of disease in medical records.

- **Recurrence:** Patient's medical record had evidence of recurrence after a documented disease-free period.

- **Progression or Residual Disease:** Patient was never disease-free and either had a documented progression during or after completion of first course therapy; or partial response to therapy with residual disease (never disease-free but not progressed).

Cancer outcome status was based on documentation in progress notes, history and physical, laboratory and/or imaging results, including blood tests, liver tests, mammogram, CT scans, colonoscopy, MRI, bone scan, chest imaging, or biopsy. If there was not enough medical information to make a determination or if the patient died before documentation of a response to therapy, abstractors recorded the outcome as unknown or uncertain within each category. Of note, the definition of recurrence required a documented disease-free period and was mutually exclusive with progression or residual disease (Fig. 1). The 'last active follow-up (AFU) date' documents when medical records were last reviewed by abstractors. Abstractors also recorded the dates of diagnosis, surgery, start and completion of first course therapy, initiation of subsequent therapy, along with first (earliest) and last (most recent) medical evidence of patient disease free status, recurrence, or progression. If only the day was missing, we imputed it as the midpoint of the month subject to consistency checks with other dates (Appendix). If a month or year was missing, dates were considered missing.

2.3. Time to event definitions

For our analysis, we considered surgery date or first disease-free (DF) date as the starting point to determine the time interval until the outcome of interest. We eliminated diagnosis date as a starting point because the response to treatment determined whether the patient became DF and was at risk for recurrence or progression, or residual disease; the interval from diagnosis to treatment would artificially inflate the time patients were at risk for these outcomes.

Among surgery patients who became DF, we considered two approaches to determine time to recurrence: 1) Censor all patients without a documented recurrence at the last documented DF date – a conservative approach; 2) Censor DF patients at the AFU date. Patients with a documented recurrence were classified as having an event at the recurrence date for both approaches. For DF patients who died, we determined vital status using 2016 National Death Index and whether the cause of death was due to cancer according to the SEER Cause-specific Death Classification [17]. Using this additional information, we applied a set of rules to classify these events (or non-events) and in some cases imputed a recurrence date (Appendix).

2.4. Analysis

Patient and tumor characteristics for the study population are presented by cancer site. Continuous variables are presented as either means (+/−) with standard deviations or as medians with 25th and 75th percentiles, and discrete variables as frequencies and percentages. Recurrence-free survival curves are presented as Kaplan-Meier estimates for each cancer site using the last DF date for censoring compared to the AFU date assumption as described above. All analyses were performed using SAS version 9.4 or R version 3.5.0.

3. Results

The NPCR CER study included 11,769 breast and 6033 colorectal cancer cases diagnosed in 2011. Louisiana then Colorado contributed the highest number of cases, followed by New Hampshire, Idaho, and Rhode Island (Table 1). Overall, most patients were white (86.2 %) and non-Hispanic (94.4 %), had private insurance (54.1 %), lived in census tracts with < 20 % of population under the federal poverty level (86.2%) and lived in 100% urban areas (51.8%). Stage at diagnosis varied significantly by cancer. Proportionally more patients were diagnosed with breast cancer at earlier stages (19.5 % at AJCC TNM stage 0 and 41.8 % at stage I) than colorectal cancer (7.3 % and 23.6 %, respectively). Nearly 1 in 5 (19.8 %) colorectal cancer patients were diagnosed with metastatic cancer (stage IV) compared to 4.9 % of breast cancer patients. The cancer diagnosed in 2011 was the first and only recorded cancer for 73.0 % of breast and 71.0 % colorectal patients; over 25 % of the patients had two or more primary cancers.

We focused our analyses on the 10,367 (88.4 %) female breast cancer cases and 4091 (67.8 %) colorectal cancer cases who were documented as ever having been DF, of which 630 (6.1 %) and 530 (13.0 %), respectively, had a subsequent documented recurrence (Table 2). Among ever-DF patients, the percentage with documented recurrence was higher among those first diagnosed with metastatic disease (31.1 % for breast and 55.0 % for colorectal for stage IV) compared to patients with localized or locally advanced disease (3.4 % breast and 5.1 % colorectal for stage I; 12.2 % breast and 16.9 % colorectal for stage II-III). Disease-free status was unclear or unknown for small numbers of cases (378 breast, 301 colorectal). Of interest, 11 % of stage IV breast and colorectal cancer cases were documented as disease-free. Among colorectal cancer cases, these patients were younger and more likely to have private insurance compared to never disease-free stage IV patients (data not shown). Disease-free stage IV breast cancer patients were more likely to be younger and of Hispanic

ethnicity (data not shown). Overall, 76 % of cases had complete active follow-up at 60 months post-diagnosis. There were 981 breast and 1641 colorectal cancer cases classified as 'never disease free'. A small number were documented as having residual disease. Progression of disease was documented for 467 (47.6 %) female breast and 701 (42.7 %) colorectal cancer cases. For a larger number of cases (1252), the abstractors could not determine an outcome for breast and colorectal cancer patients who were never DF. However, 60 % of cases listed as having residual disease only (no progression) and 69 % of cases whose progression outcome was uncertain died from a cause of death consistent with their cancer [17].

Surgery following diagnosis occurred sooner for colorectal cancer than female breast cancer patients (Figs. 2a, b). There was significant variability across registries for the time interval from surgery to DF (Fig. 2c, d). One registry assigned the surgery date as the DF date for the majority of patients. Among the remaining states, the median time interval from surgery to DF ranged from 147 to 206 days for breast cancer and 54–243 days for colorectal cancer; overall, the median time interval was 125 days for breast and 73 days for colorectal cancer.

Fig. 3 presents Kaplan-Meier curves for 4-year recurrence-free survival by cancer type and stage at diagnosis for the first primary cancer diagnosed in 2011. As expected, recurrence-free survival (RFS) varies significantly by stage for both cancer sites. At year 4, RFS proportions were similar when using the last DF date (i.e., DF supported by documentation) compared to AFU date (i.e., assumed patient remained DF at last review of records, unless otherwise noted) for breast cancer (< 1 % absolute difference in recurrence-free survival percentage across all stage groups). Differences were slightly larger for colorectal cancer but still relatively small (1.1–2.0 % higher recurrence-free survival with the AFU assumption with larger differences at less advanced stage). Applying the AFU assumption resulted in less censoring and increased the percentage with complete follow-up at 4 years from 72.2% to 84.6% for breast cancer and from 75.8% to 89.2% for colorectal cancer. For female breast cancer, RFS using the AFU assumption was 95.2 %, 91.8 %, 83.2 % and 63.4 % for stage 0, I, II-III, and IV, respectively; for colorectal cancer patients the results were 81.8 % for stage I, 69.2 % for stage II-III, and 37.8 % for stage IV.

4. Discussion

The NPCR PCOR project was the largest, most comprehensive effort conducted by population-based cancer registries in the U.S. to obtain detailed data on cancer outcomes beyond the first course of treatment. Central cancer registries in five states conducted extensive reviews of medical records to capture data on DF status and recurrence or progression of disease in breast and colorectal cancer patients. A primary purpose of the project was to assess the feasibility of data collection, characterize the population on whom outcome data are available, and determine important variables required to characterize cancer outcomes.

Disease-free status was a pivotal data item that defined cancer outcomes. We found the capture of a DF period to be feasible as evidenced by having only 425 (2.4 %) patients with an 'unclear' DF status; 254 (1.4 %) had no information other than diagnosis. We

acknowledge, however, that the determination of DF status, recurrence, or progression depends on the timing and type of patient follow-up after initial treatment and having the clinical means to detect the tumor. Although abstractors were able to classify whether or not patients became DF, we observed substantial variability in the timing and approach to make that determination. Some abstractors relied on physician documentation of no evidence of disease, or laboratory and imaging results to determine DF status, while others considered the patient DF immediately following surgery with no positive margins or residual disease. Consistent methodology standards are needed.

Although patients are not truly at risk for recurrence until they become DF, we did not use first DF date as a starting point for RFS due to the variability in the time interval between surgery and first DF date among states. To avoid introducing a systematic timing bias and to be consistent with other published studies [16–19], we chose surgery date as the RFS starting point because surgery is generally viewed as a curative treatment for patients without advanced disease and is a defined event less dependent on documentation or interpretation. A start point of surgery, however, does not take into account any adjuvant treatment needed to become DF, particularly for patients with microscopic residual tumor after surgery.

Among patients who were never DF and, therefore, at risk for progression or residual disease, nearly half (44 % breast, 50 % CRC) were coded as "uncertain if progressed or residual disease." A majority of these uncertain cases had the cancer of interest listed as a cause of death, thus complicating an assumption of no progression for those cases. Due to these data completeness and quality issues, we did not analyze endpoints for never-DF patients, and caution on making in-ferences about progression percentages. We believe progression results should only describe the data as collected by the cancer registries and that survival should be the outcome of interest for this group.

We retained more patients in the RFS analysis for a longer period through the application of rules based on available death information (Appendix). For those who died of non-cancer causes, we assumed no recurrence prior to death. For patients with unknown recurrence status who died due to cancer, we assumed an undiagnosed or undocumented recurrence preceded death, and classified them as a recurrence with an imputed recurrence date halfway between the last DF date and death date. An advantage of this approach is it allows consistency between the outcome of recurrence and the composite outcome of death or recurrence. This method may have resulted in some misclassification of recurrence due to the poor reliability and validity of cause of death listed on death certificates. In addition, the timing of events may be incorrect. We believe the impact of these assumptions is minor and offers an advantage over excluding or censoring these known events, which would overestimate recurrence-free survival.

There were many strengths of this surveillance study. The large, five-state patient cohort included the entire state population of residents diagnosed with breast or colorectal cancer. Overall, the age, sex, and stage distribution by cancer site mirrored the U.S. Cancer Statistics for the same diagnostic year, 2011 [18]. Abstractors collected project-specific outcomes directly from medical records and were not limited to hospital registries. Established NPCR

central cancer registries consolidated information from multiple sources, including state vital records, national death index and out-of-state health facilities, as available. The New Hampshire registry demonstrated that active follow-up was achievable, albeit resource intensive [19]. Areas of improvement could have been to validate a subset of collected data to ensure accurate abstractions and consistent interpretation of study-defined variables, such as first DF date; and to collect microscopic surgical margin findings in all states, a factor that may be of interest in future studies. It may seem clinically illogical that metastatic cases could become DF or that we retained in-situ cases in our analysis, but our primary intent was to describe the data, focusing on methodologic and analytic considerations rather than results. In future studies, investigators may choose to restrict analysis by stage based on their study question. Another challenge is the ability to differentiate cancer recurrence from a second primary of the same cancer type. For cancer surveillance, there are established coding rules to address this issue. Although abstractors for our study followed these rules, we acknowledge a potential for misclassification particularly farther out from diagnosis.

Prior studies looking at recurrence are available but challenging to compare to our findings without creating the same stage groupings and stratification factors such as biologic status of the tumor [2,5,20]. Multiple studies describe first recurrences for breast cancer peaking in the second year after diagnosis, and our active follow-up covers this time-period [5,21]. We intend to use the NPCR PCOR dataset for comparison with these and other studies in future analyses.

5. Conclusions

The NPCR PCOR project explored defining and capturing recurrence and progression from a registry and epidemiologic perspective and demonstrated that population-based cancer registries could collect patient-centered outcomes. Key findings were that while the capture of DF status and recurrence was useful and feasible, dates related to being DF were less consistently collected, limiting their use as a starting point. For patients who were never DF, we had difficulty distinguishing between progression and residual disease. The last date when the patient's record was reviewed (AFU date) was important to confirm follow-up surveillance. To collect POCR outcomes nationally and consistently, the surveillance community requires further refinement of standards including a minimum variable set that defines a start point, time intervals, criteria to establish DF status, and guidance on when to conduct follow-up abstraction. Participating state cancer registries were able to complete extensive follow-up and abstraction from multiple sources because the NCPR PCOR project was heavily resourced; such a project would not be feasible within the regular operations of central cancer registries. Electronic data capture, natural language processing and dataset linkages are promising alternatives to capture recurrence data and warrant further consideration. The NPCR PCOR project can be used to validate these innovative approaches along with comparing treatment effectiveness and cancer outcomes at a population-level.

Funding

Appendix A.: Detailed description of data definitions and date imputation

A.1 Definitions from data dictionary used by abstractors

Disease Free / No Evidence of Disease

Information on patients' disease free status and dates of **first (earliest)** and **last (most recent)** medical evidence of patient disease free status are recorded in these fields. Do not skip these questions; there are opportunities in the status variables to record if a patient was never disease free. Dates cannot be left blank, 9's are required if no date is applicable.

Documented Disease Free Status (Item # 8004)

Description

This variable is used to indicate if the patient ever had a documented disease free / no evidence of disease status. Information recorded here should correspond with one of the following: history and physical, labs or imaging used to determine disease status. Please note that the laboratory tests or scans can include (but are not limited to): blood tests, liver tests, mammogram, CT scans, colonoscopy, MRI, bone scan, chest imaging, biopsy, etc.).

Coding

> 0 No information on patient other than a diagnosis of cancer.
>
> 1 Patient never found to be disease free (includes those with residual disease, progression, and those who may have died prior to being disease free).
>
> 2 Patient had at least one record of documented disease free status.
>
> 9 Unknown. It is unclear in patient record if ever disease free.

Recurrence, Progression, or Residual Disease

These variable fields are used to collect information on recurrence, progression and residual disease. Do not skip these questions; there are opportunities to record not applicable status.

Recurrence Status (Item # 8010) Description and Coding

This variable is used to further describe any recorded evidence of recurrence (this means there was a disease free period before the cancer reappeared – not residual disease or progression). This information should be based on one of the following: history and physical, labs or imaging used to clinically evaluate cancer status. Please note that the laboratory tests or scans can include (but are not limited to): blood tests, liver tests, mammogram, CT scans, colonoscopy, MRI, bone scan, chest imaging, biopsy, etc.).

> 0 Patient never found to be disease free (for example if patient progressed, has residual disease, lost to follow-up after diagnosis, or died prior to becoming disease free and therefore not able to have a recurrence).
>
> 1 Found to be disease free and remained disease free till end of study.

2 Documented recurrence (after a documented status of disease free / no evidence of disease / NED)

3 Uncertain if recurrence or residual disease based on incomplete documentation.

4 Uncertain if recurrence or residual disease based on conflicting documentation (one provider reports recurrence and another reports residual disease).

9 Unknown. Patient was disease free (Item #8005complete) and additional patient documentation exists, however it is unclear if remained disease free or recurred.

Progression / Residual Disease Status (Item # 8014) Description and Coding

This variable is used to further describe any recorded evidence of progression or residual disease **(NOT recurrence).** This information should correspond with one of the following: history and physical, labs or imaging used to clinically evaluate cancer status. Please note that the laboratory tests or scans can include (but are not limited to): blood tests, liver tests, mammogram, CT scans, colonoscopy, MRI, bone scan, chest imaging, biopsy, etc.).

0 Patient found to be disease free.

1 Patient had partial response to therapy with residual disease (never disease free but not progressed).

2 Patient had documented progression during or after completion of first course therapy.

3 Patient never disease free but uncertain if progressed or residual.

9 Unknown –No information on patient other than a diagnosis of cancer or incomplete medical information (lost to follow-up or died prior to documentation of response to therapy)

A.2   Imputation of missing day in dates

If only the day within a date was missing, we imputed the date as the midpoint of the month subject to consistency checks with other dates. For example, if diagnosis day was imputed to the midpoint of the month, all treatment dates were then checked. If any treatment was prior to the imputed day, the imputed day was updated to halfway between the first of the month and the earliest treatment. Similar checks were performed for other dates.

Appendix B.: Rules to define events (1 – Recurrence, 2 – Death or Recurrence) along with censor dates for each event

B.1 Patients who had surgery and a documented recurrence (n = 1108) were classified as events of both types at the date of recurrence.

B.2 Patients who had surgery and were found to be DF and remain DF until end of study (n = 12,126)

Vital status	Cause of Death	1 - Recurrence status	1 – Censor/Event Date	2 – Death or Recurrence	2 – Censor/Event Date	Number of patients
Alive	Not applicable	No event	AFU date	No event	AFU date	10,714
Died	Non-cancer related	No event	Death or AFU date, whichever occurred first	Event	Death date	1,184
Died	Cancer-related	No event *	Last DF date	Event **	Death date	228

*
Determined by abstractor call for recurrence (DF) but does not assume until end of study due to cancer death (censor at last known DF).

**
Count as an event at death date for composite endpoint of death or recurrence. Better to possibly inflate the time to death or recurrence than censor these known events (death) as non-events at the last DF date. This would clearly overestimate recurrence-free survival.

B.3 Patients who had surgery and were found to be DF but had unknown[1] recurrence status (n = 609)

Vital status	Cause of Death	1 - Recurrence status	1 – Censor/Event Date	2 – Death or Recurrence	2 – Censor/Event Date	Number of patients
Alive	Not applicable	No event	Last DF date	No event	Last DF date	462
Died	Non-cancer related	No event	Death date	Event	Death date	99
Died	Cancer-related	Event	Impute date halfway between last DF date and death date	Event	Impute date halfway between last DF date and death date	48

[1]Uncertain if recurrence occurred based on incomplete, conflicting, or unclear information.

Abbreviations:

AFU	Active follow-up
CDC	Centers for Disease Control and Prevention
DF	Disease-free
NPCR	National Programs of Cancer Registries
PCOR	Patient-centered cancer research

References

[1]. U.S. Cancer Statistics Working Group, Five-Year Relative Survival, U.S. Cancer Statistics Data Visualizations Tool, 2018 [cited 2019; Available from: www.cdc.gov/cancer/dataviz].

[2]. Radosa JC, et al., Evaluation of local and distant recurrence patterns in patients with triple-negative breast cancer according to age, Ann Surg Oncol. 24 (3) (2017) 698–704. [PubMed: 27783163]

[3]. Jatoi I, et al., Hazard rates of recurrence following diagnosis of primary breast cancer, Breast Cancer Res. Treat 89 (2) (2005) 173–178. [PubMed: 15692760]

[4]. Mittendorf EA, et al., Impact of chemotherapy sequencing on local-regional failure risk in breast cancer patients undergoing breast-conserving therapy, Ann. Surg 257 (2) (2013) 173–179. [PubMed: 23291658]

[5]. Ribelles N, et al., Pattern of recurrence of early breast cancer is different according to intrinsic subtype and proliferation index, Breast Cancer Res. 15 (5) (2013) R98. [PubMed: 24148581]

[6]. Eisenhauer EA, et al., New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1), Eur. J. Cancer 45 (2) (2009) 228–247. [PubMed: 19097774]

[7]. In H, et al., Cancer recurrence: an important but missing variable in national cancer registries, Ann. Surg. Oncol 21 (5) (2014) 1520–1529. [PubMed: 24504926]

[8]. Mariotto AB, et al., Can we use survival data from cancer registries to learn about disease recurrence? The case of breast cancer, Cancer Epidemiol. Biomark. Prev 27 (11) (2018) 1332–1341.

[9]. Shi Q, Cancer registries: a novel alternative to long-term clinical trial follow-up based on results of a comparative study, Clin. Trials (Lond. Print) 7 (6) (2010) 686–695.

[10]. In H, et al., The quest for population-level cancer recurrence data; current deficiencies and targets for improvement, J. Surg. Oncol 111 (6) (2015) 657–662. [PubMed: 25690414]

[11]. Warren JL, et al., Sensitivity of medicare claims to identify cancer recurrence in elderly colorectal and breast cancer patients, Med. Care 54 (8) (2016) e47–54. [PubMed: 24374419]

[12]. Clancy C, Collins FS, Patient-centered outcomes research institute: the intersection of science and health care, Sci. Transl. Med 2 (37) (2010) p. 37cm18–37cm18.

[13]. Institute of Medicine, Initial Priorities for Comparative Effectiveness Research, N.A. Press, Washington DC, 2009 Editor.

[14]. Chen VW, et al., Enhancing Cancer registry data for comparative effectiveness research (CER) project: overview and methodology, J. Registry Manag 41 (3) (2014) 103–112. [PubMed: 25419602]

[15]. Thornton ML (Ed.), Standards for Cancer Registries Volume II: Data Standards and Data Dictionary, Record Layout Version 18, North American Association of Central Cancer Registries, Springfield, IL, 2018.

[16]. American Joint Committee on Cancer, et al., Edge SB (Ed.), AJCC Cancer Staging Manual, seventh ed., Springer, New York, 2010.

[17]. N.C.I. Surveillance Research Program, SEER Cause-Specific Death Classification, [cited 2019 2/12/2019]; Available from: (2019) https://seer.cancer.gov/causespecific/.

[18]. US Cancer Statistics Working Group, CDC (Ed.), United States Cancer Statistics: 1999–2015 Cancer Incidence and Mortality Data, US Department of Health and Human Services, Atlanta, GA, 2018CDC: Atlanta, GA: US Department of Health and Human Services, CDC.

[19]. Celaya MO, P M, Pollack LA, Celaya V, Riddle BL, Rees JR, Collection of active follow-up data in a NPCR registry: a review of the patient-centered outcomes project at the New Hampshire State Cancer Registry, NAACCR 2018, (2018) Pittsburgh, PA.

[20]. Demicheli R, et al., Recurrence and mortality according to estrogen receptor status for breast cancer patients undergoing conservative surgery. Ipsilateral breast tumour recurrence dynamics provides clues for tumour biology within the residual breast, BMC Cancer 10 (2010) 656. [PubMed: 21118508]

[21]. Geurts YM, et al., Patterns and predictors of first and subsequent recurrence in women with early breast cancer, Breast Cancer Res. Treat 165 (3) (2017) 709–720. [PubMed: 28677011]

Disease Free after completion of inital treatment?

Documented Disease Free	Never Disease Free	Unknown if Disease Free
Recurrence	Documented progression	Unclear if ever Disease Free
DF until end of study	Residual disease, no progression	No Information other than Diagnosis
Uncertain if recurrence	Uncertain if progression	
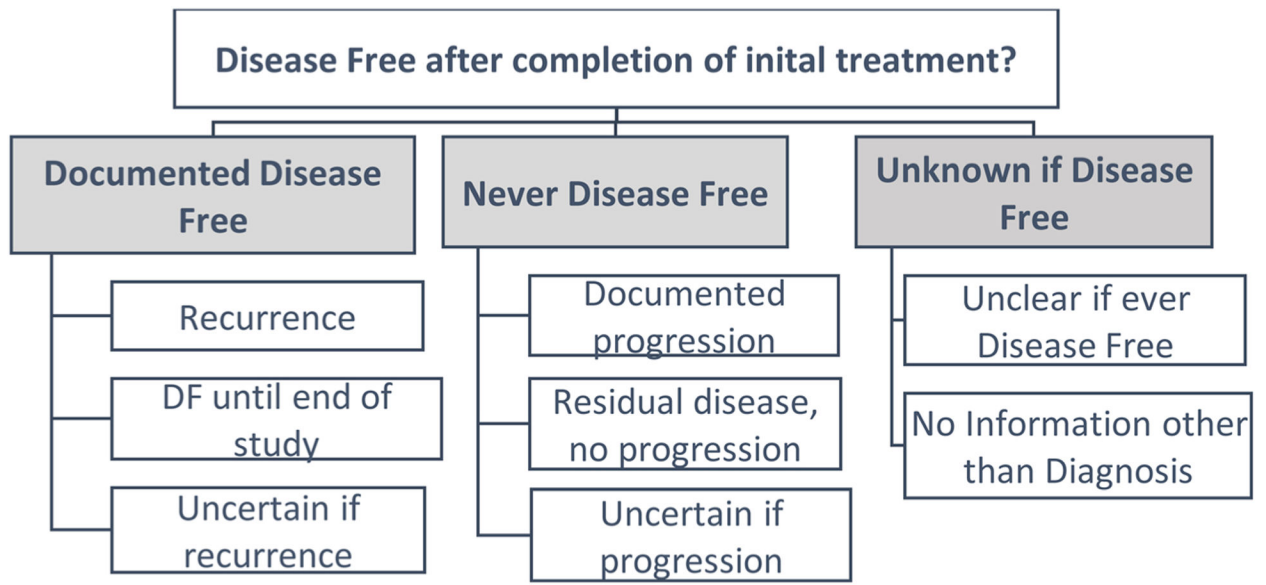
Fig. 1.

Approach to determining patient-centered outcome based on disease-free status – National Programs of Cancer Registires Patient Centered Outcomes Research (NPCR PCOR).

Diagnosis → Surgery → Disease Free (DF)? → Recurrence, Disease-Free Survival or Death

Days from diagnosis to surgery

2a. Female Breast Cancer (n=9,940)

2b[2]. Colorectal Cancer (n=3,894)

Days from surgery to documented disease free status[3]

2c. Female Breast Cancer (n=9,941)

2d[2]. Colorectal Cancer (n=3,895)
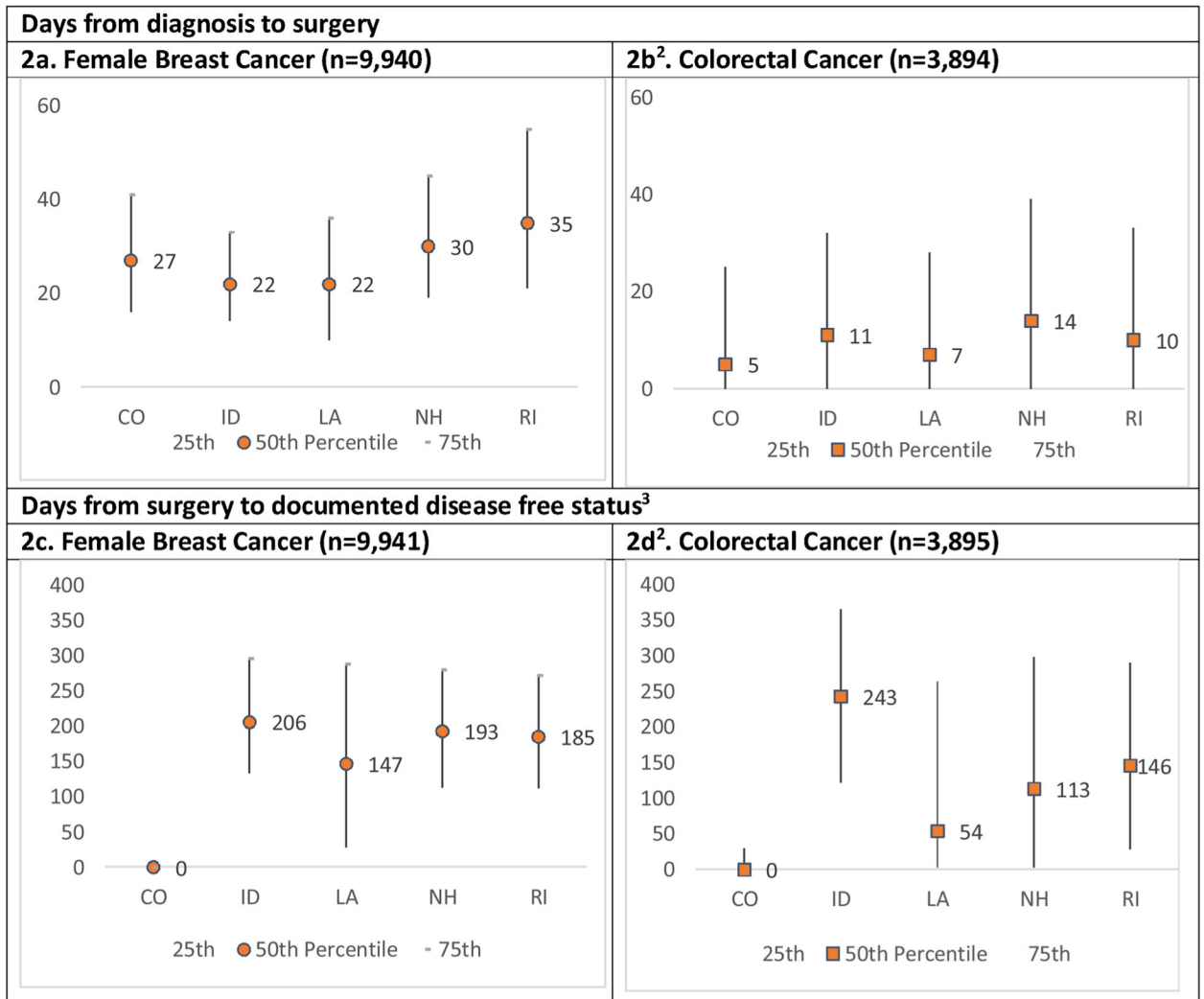
Fig. 2.

Variations of time-to-event by starting and ending points for breast and colorectal cancers[1].

[1]Limiting to first primary diagnosed in 2011.

[2]Surgery date is nearly equal to DF date for Colorado because abstractors considered patients DF after definitive cancer surgery.

[3]Differences in counts are due to incomplete diagnosis dates.

Fig. 3.

Recurrence-Free Survival from surgery date for breast (female) and colorectal cancer by stage at diagnosis – 5 NPCR states, diagnosis year 2011.

Table 1

National Program of Cancer Registries (NPCR) Patient-Centered Outcome Study Population[a] Patient and Tumor Characteristics, Diagnosis Year 2011.

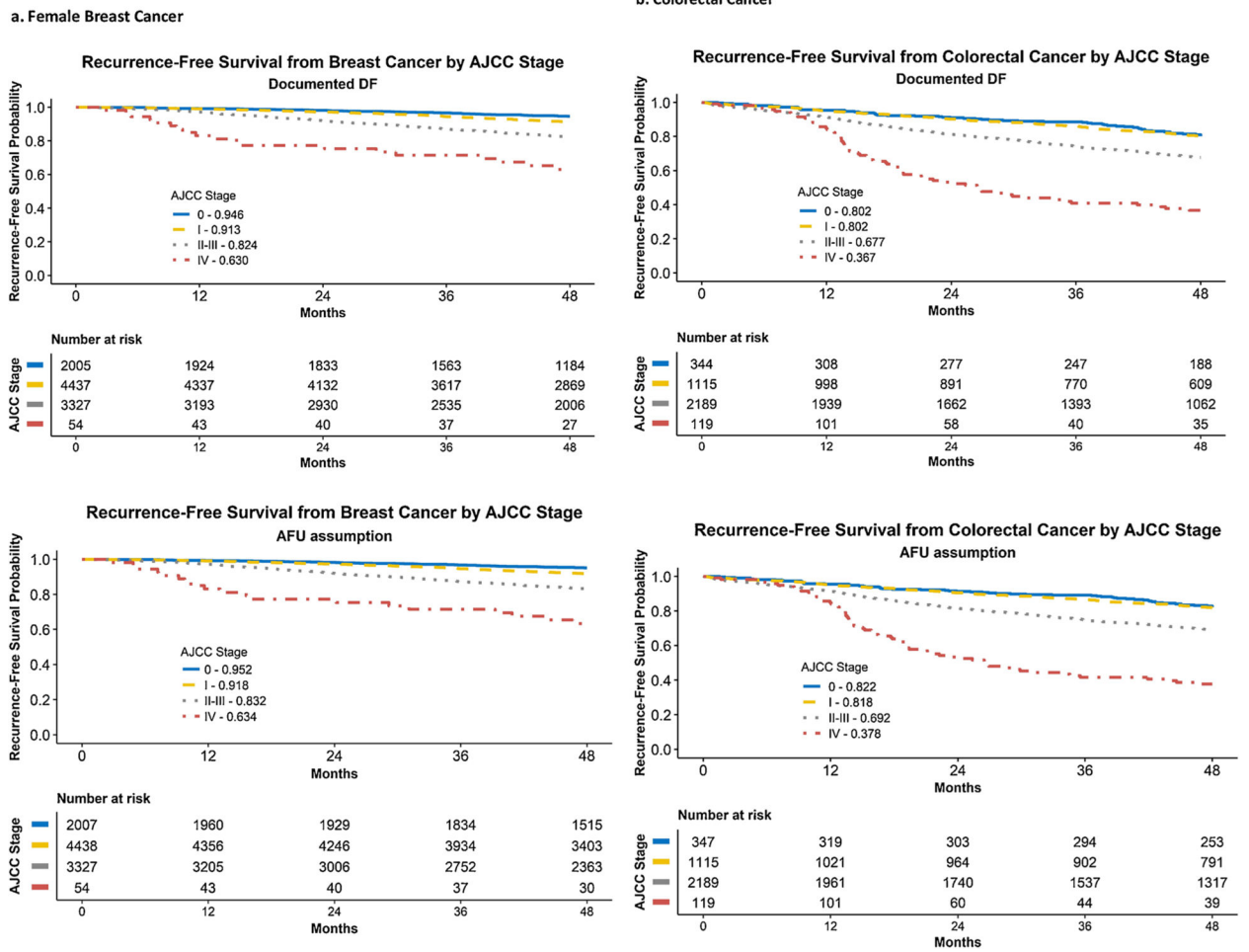|	Breast Cancer N = 11769	Colorectal Cancer N = 6033	Total N = 17802
State			
Colorado	4186 (35.6 %)	1863 (30.9 %)	6049 (34.0 %)
Idaho	1217 (10.3 %)	637 (10.6 %)	1854 (10.4 %)
Louisiana	3891 (33.1 %)	2409 (39.9 %)	6300 (35.4 %)
New Hampshire	1448 (12.3 %)	653 (10.8 %)	2101 (11.8 %)
Rhode Island	1027 (8.7 %)	471 (7.8 %)	1498 (8.4 %)
Age (years)	61.4 ± 13.4	67.2 ± 13.8	63.3 ± 13.8
Sex			
Male	43 (0.4 %)	3109 (51.5 %)	3152 (17.7 %)
Female	11726 (99.6 %)	2924 (48.5 %)	14650 (82.3 %)
Race			
White	10208 (87.1 %)	5073 (84.5 %)	15281 (86.2 %)
Black	1299 (11.1 %)	837 (13.9 %)	2136 (12.1 %)
American Indian / Alaska Native	49 (0.4 %)	20 (0.3 %)	69 (0.4 %)
Asian / Pacific Islander	145 (1.2 %)	64 (1.1 %)	209 (1.2 %)
Other	17 (0.1 %)	8 (0.1 %)	25 (0.1 %)
Unknown	51	31	82
Ethnicity			
Non-Hispanic	11121 (94.5 %)	5692 (94.3 %)	16813 (94.4 %)
Hispanic	648 (5.5 %)	341 (5.7 %)	989 (5.6 %)
Insurance			
Private	6551 (59.1 %)	2539 (44.4 %)	9090 (54.1 %)
Medicare / Other Public	3006 (27.1 %)	2246 (39.3 %)	5252 (31.3 %)
Medicaid	1248 (11.3 %)	657 (11.5 %)	1905 (11.3 %)
None	278 (2.5 %)	273 (4.8 %)	551 (3.3 %)
Unknown	686	318	1004
Census Tract Poverty			
< 20 %	10268 (87.5 %)	5023 (83.6 %)	15291 (86.2 %)
20 %	1468 (12.5 %)	987 (16.4 %)	2455 (13.8 %)
Census Tract Residence[b]			
100 % Urban	6187 (52.7 %)	3011 (50.1 %)	9198 (51.8 %)
100 % Rural	1168 (10.0 %)	639 (10.6 %)	1807 (10.2 %)
Mixed	4383 (37.3 %)	2361 (39.3 %)	6744 (38.0 %)
Comorbid Conditions[c]			
None	9346 (79.4 %)	3994 (66.2 %)	13340 (74.9 %)
One	1937 (16.5 %)	1365 (22.6 %)	3302 (18.6 %)
Two or more	485 (4.1 %)	673 (11.2 %)	1158 (6.5 %)

	Breast Cancer N = 11769	Colorectal Cancer N = 6033	Total N = 17802
Unknown	1	1	2
Sequence Number			
One Primary Only	8587 (73.0 %)	4283 (71.0 %)	12870 (72.3 %)
First of Two or More Primaries	944 (8.0 %)	473 (7.8 %)	1417 (8.0 %)
Second of Two or More Primaries	1911 (16.2 %)	1081 (17.9 %)	2992 (16.8 %)
Third of Three or More Primaries	285 (2.4 %)	166 (2.8 %)	451 (2.5 %)
Fourth or Greater	42 (0.4 %)	30 (0.5 %)	72 (0.4 %)
Stage			
0	2243 (19.5 %)	416 (7.3 %)	2659 (15.4 %)
I	4815 (41.8 %)	1348 (23.6 %)	6163 (35.8 %)
II	2856 (24.8 %)	1413 (24.8 %)	4269 (24.8 %)
III	1033 (9.0 %)	1397 (24.5 %)	2430 (14.1 %)
IV	562 (4.9 %)	1130 (19.8 %)	1692 (9.8 %)
Unknown/NA	260	329	589

Note: Continuous variables are presented as means ± standard deviations and discrete variables as frequencies and percentages.

[a] Multiple primaries are included in this table.

[b] Defined as the percent of the U.S. census tract population that is urban based on the 2010 decennial Census. Mixed is defined as > 0% < 100% urban.

[c] Number of Charlson comorbidity index conditions.

Table 2

Patient-centered outcomes by cancer type and stage at diagnosis - 5 US states, diagnosis year 2011.

Female Breast Cancer	Stage 0 n = 2240	Stage I n = 4803	Stage II-III n = 3867	Stage IV n = 560	All Stages[a] n = 11726
Documented Disease Free	**2175 (97.1 %)**	**4611 (96.0 %)**	**3422 (88.5 %)**	**61 (10.9 %)**	**10367 (88.4 %)**
Recurrence	26 (1.2 %)	159 (3.4 %)	416 (12.2 %)	19 (31.1 %)	630 (6.1 %)
Disease free until end of study	2074 (95.4 %)	4310 (93.5 %)	2910 (85.0 %)	37 (60.7 %)	9411 (91.8 %)
Unknown	75 (3.4 %)	142 (3.1 %)	96 (2.8 %)	5 (8.2 %)	326 (3.1 %)
Never Disease Free	**14 (0.6 %)**	**89 (1.9 %)**	**334 (8.6 %)**	**485 (86.6 %)**	**981 (8.4 %)**
Residual disease, no progression	0 (0.0 %)	12 (13.5 %)	20 (6.0 %)	48 (9.9 %)	85 (8.7 %)
Progression	0 (0.0 %)	16 (18.0 %)	177 (53.0 %)	254 (52.4 %)	467 (47.6 %)
Unknown	14 (100.0 %)	61 (68.5 %)	137 (41.0 %)	183 (37.7 %)	429 (43.7 %)
Unknown if Disease Free	**51 (2.3 %)**	**103 (2.1 %)**	**111 (2.9 %)**	**14 (2.5 %)**	**378 (3.2 %)**
Unclear if ever disease free	35 (68.6 %)	84 (81.6 %)	86 (77.5 %)	5 (35.7 %)	224 (59.3 %)
No info other than diagnosis of cancer	16 (31.4 %)	19 (18.4 %)	25 (22.5 %)	9 (64.3 %)	154 (40.7 %)

Colorectal Cancer	Stage 0 n = 416	Stage I n = 1348	Stage II-III n = 2810	Stage IV n = 1130	All Stages* n = 6033
Documented Disease Free	**385 (92.5 %)**	**1206 (89.5)**	**2277 (81.0 %)**	**129 (11.4 %)**	**4091 (67.8 %)**
Recurrence	4 (1.0 %)	62 (5.1 %)	384 (16.9 %)	71 (55.0 %)	530 (13.0 %)
Disease free until end of study	347 (90.1 %)	1021 (84.7 %)	1748 (76.8 %)	52 (40.3 %)	3237 (79.1 %)
Unknown	34 (8.8 %)	123 (10.2 %)	145 (6.4 %)	6 (4.7 %)	324 (7.9 %)
Never Disease Free	**16 (3.8 %)**	**97 (7.2 %)**	**408 (14.5 %)**	**975 (86.3 %)**	**1641 (27.2 %)**
Residual disease, no progression	0 (0.0 %)	5 (5.2 %)	34 (8.3 %)	70 (7.2 %)	117 (7.1 %)
Progression	5 (31.3 %)	27 (27.8 %)	181 (44.4 %)	469 (48.1 %)	701 (42.7 %)
Unknown	11 (68.8 %)	65 (67.0 %)	193 (47.3 %)	436 (44.7 %)	823 (50.2 %)
Unknown if Disease Free	**15 (3.6 %)**	**45 (3.3 %)**	**125 (4.4 %)**	**26 (2.3 %)**	**301 (5.0 %)**
Unclear if ever disease free	10 (66.7 %)	35 (77.8 %)	122 (97.6 %)	17 (65.4 %)	201 (66.8 %)
No info other than diagnosis of cancer	5 (33.3 %)	10 (22.2 %)	3 (2.4 %)	9 (34.6 %)	100 (33.2 %)

Note: Multiple primaries are included in this table; outcome analyses in subsequent tables and figures are limited to one record per patient.

[a] All stages case count includes unknown stages, which are not shown in the table.