Cancer information and population health resource: a resource for catchment area data and cancer outcomes research

Christopher D. Baggett (iD), PhD,[1,2,*] Bradford E. Jackson, PhD,[1] Laura Green, MBA,[1] Tzy-Mey Kuo, PhD,[1] KyungSu Kim, MPH,[1] Xi Zhou, PhD,[1] Katherine E. Reeder-Hayes (iD), MD, MBA, MSc,[1,3] Jennifer L. Lund, PhD,[1,2] Stephanie B. Wheeler (iD), PhD,[1,4] Andrew F. Olshan, PhD[1,2]

[1]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[2]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[3]Division of Oncology, Department of Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[4]Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

***Correspondence to:** Christopher Baggett, PhD, University of North Carolina, 101 E Weaver St, CB#7293, Carrboro, NC 27599-7293, USA (e-mail: cbaggett@email.unc.edu).

Abstract

Background: The University of North Carolina at Chapel Hill Lineberger Comprehensive Cancer Center has developed a novel data resource, the Cancer Information and Population Health Resource (CIPHR), for conducting catchment area evaluation and cancer population health research that links the North Carolina Central Cancer Registry (NCCCR) to medical and pharmacy claims data from Medicare, Medicaid, and private plans operating within North Carolina. This study's aim was to describe the CIPHR data and provide examples of potential cohorts available in those data.

Methods: We present the underlying populations included in the NCCCR and claims data before linkage and demonstrate estimated sample sizes when these data are linked and commonly used insurance enrollment criteria are applied.

Results: Data for the years 2003-2020 are present in CIPHR and include 947 977 cancer cases from the NCCCR and 21.6 million enrollees in public and private health insurance (cancer and noncancer cases). When limited to first or only cancers (n = 672 377), 86% could be linked to insurance enrollment for at least 1 month during 2003-2020 (n = 582 638), with 62% of individuals linking to enrollment during the month of cancer diagnosis. Among all registry cancer cases, 47% (n = 317 898) had continuous insurance enrollment for at least 12 months before and after cancer diagnosis.

Conclusion: CIPHR illustrates the utility of establishing and maintaining a statewide, comprehensive cancer population health database. This resource serves to characterize the cancer center catchment area and aids in tracking cancer outcomes and trends in care delivery as well as identifying disparities that require intervention and policy focus.

The mission of National Cancer Institute Comprehensive Cancer Centers includes providing its members with data resources that enable innovative population science research and local community assessment. By providing these resources, cancer centers can conduct research and inform outreach efforts that are relevant to the center's catchment area and that can improve the patient's experience, from cancer prevention through end-of-life care. To achieve maximum effectiveness, these resources must not only provide high-quality and comprehensive data but also make available trained personnel experienced in management and analysis of the data as well as computing infrastructure to handle what are frequently cumbersome data containing protected health information.

The University of North Carolina at Chapel Hill Lineberger Comprehensive Cancer Center (LCCC) has developed a novel, population-based data resource that links the North Carolina Central Cancer Registry (NCCCR) to medical and pharmacy claims data from Medicare, Medicaid, and private insurance plans operating within North Carolina. These data have been linked and maintained by the Cancer Information and Population Health Resource (CIPHR; formerly known as the Integrated Cancer Information and Surveillance System) (1), a shared resource of the LCCC and a central component of its Cancer Outcomes Research Program. The registry and claims data have been further linked to contextual data from the US Census, American Communities Survey, and other publicly available sources. CIPHR not only maintains and regularly updates these data but also employs personnel skilled in data management and statistical analysis; it also provides the accessible, secure computing environment required to house the sensitive data (2) and has a library of statistical code for data management and variable construction that can be reused across projects. The vision of CIPHR is to lead pioneering research that uses population data, informatics, and methods to improve cancer outcomes in North Carolina and across the United States.

CIPHR requires a substantial investment from LCCC. Funding largely comes from state sources, with additional support from the Cancer Center Support Grant (5-P30-CA016086). CIPHR operates as a recharge facility, currently recovering approximately 40% of its

annual operating budget by charging project and personnel support fees on a per-study basis. The vast majority of operating costs come from personnel; CIPHR staff currently include 1 full-time faculty director, 2 partially funded faculty members (20% full-time equivalents), 1 project director, 4 master's or PhD-level statistical analysts, 1 data manager, and 2 data systems analysts.

In the United States, two of the more commonly used data sources for cancer outcomes research are central cancer registries and health insurance claims. Every state in the United States, along with several large metropolitan areas, operates central cancer registries that collect basic demographic information, tumor type and stage, initial course of treatment, and vital status information on all cancer cases diagnosed within their catchment areas (3,4). Health insurance claims add complementary data on screening, comorbidities, treatment, and health-care encounters (5,6). Linking these data sources results in a comprehensive view of an individual's cancer care journey (7-10). Research value is further enhanced by including neighborhood-level sociodemographic data and local cancer care environment information. This integration enables studies on guideline adherence, health-care utilization, cost-effectiveness, survival, and disparities in cancer outcomes across demographics. Insights from this research can improve understanding of cancer outcomes and inform better treatment strategies and policies.

The National Cancer Institute's Surveillance, Epidemiology, and End Results Program–Medicare (11,12) linked data are an important resource for cancer outcomes and health services research but includes only adults aged 65 years and older or individuals with certain disabilities. This limitation excludes younger populations, affecting studies on cancers typically diagnosed at younger ages, such as Hodgkin lymphoma. Focusing only on those individuals older than 65 years of age insured by Medicare also hampers the examination of disparities in cancer care. Linking registry data with Medicaid and private insurance increases sample size and diversity, enabling more comprehensive research on the impact of age, race, ethnicity, and socioeconomic factors on cancer outcomes.

Several states, such as Utah, Colorado, and Arkansas, have demonstrated the feasibility of linking their all-payer claims databases with their central cancer registries, and there are other examples of linking cancer registries with multipayer claims data (13-18). These linkages, however, often exclude either public or private payors, cover limited time periods, lack regular updates, restrict the age range, or are not readily accessible to other researchers. CIPHR is relatively unique in that it links cancer cases in patients of all ages to public and private insurance claims over a period of 18 years, updating data annually, and being available to the research community. CIPHR also includes claims data for individuals without cancer, allowing for comparison groups and the examination of care patterns, including cancer screening in at-risk populations.

The purpose of this work was to describe the CIPHR data and provide examples of potential cohorts available within CIPHR. We present the underlying populations included in the NCCCR and claims data before linkage and demonstrate estimated sample sizes when these data are linked and commonly used insurance enrollment criteria are applied.

Methods
Data sources
The data presented include the NCCCR and insurance claims from Medicare, Medicaid, and private insurance plans within North Carolina. These data are linked using personal identifiers contained within each source. Due to differing availability of identifiers and requirements imposed by data use agreements with each data partner, iterative, deterministic methods are used to link individuals across the data sources. The linkage provides the ability to follow patients with cancer across multiple insurance providers as they gain or lose coverage over time. Our group has previously described these linking methods in detail (19,20). The years of available data are presented in Figure 1 and include the years 2003-2019 for the NCCCR and 2003-2020 for the claims data.

North Carolina Central Cancer Registry
The NCCCR is part of the State Center for Health Statistics within the North Carolina Department of Health and Human Services. The NCCCR collects data on all incident primary cancer diagnoses among residents of North Carolina and is a part of the Centers for Disease Control and Prevention's National Program of Cancer Registries. The NCCCR is consistently both a National Program of Cancer Registries Registry of Distinction and North American Association of Central Cancer Registries Gold Certified. These distinctions indicate "complete, timely, and quality data available for cancer control activities (21)."

North Carolina Medicare
Approximately 15% of the North Carolina population was enrolled in Medicare in 2019, with two-thirds of Medicare enrollees in traditional fee-for-service Medicare and the remaining enrolled in Medicare Advantage plans. In general, individuals are eligible for Medicare if they are US citizens or legal residents, aged 65 years or older, or have qualifying disabilities. Medicare data are obtained from the Centers for Medicare & Medicaid Services (CMS) and include inpatient, outpatient, carrier, home health, hospice, skilled nursing facility, durable medical equipment, and pharmacy claims. For the years 2003-2020, these data include Medicare parts A and B; Part C (Medicare Advantage) data are included for the years 2015-19 (unlike fee-for-service data, CMS releases data on a delay), and Part D data are included for the years 2006-2020. For each of these benefits, CIPHR includes claims and enrollment information for 100% of the enrollees who reside within North Carolina, regardless of their cancer status. In addition, the CIPHR data include Medicare claims and enrollment information for North Carolina residents with incident cancer diagnoses in the NCCCR who have moved out of the state after their cancer diagnosis or who receive care in another state.

North Carolina Medicaid
Approximately 18% of the North Carolina population was enrolled in Medicaid in 2019. Eligibility criteria include being a North Carolina and US citizen or legal resident, being low income, or having qualifying health conditions. Similar to Medicare, 100% of Medicaid enrollees residing in North Carolina plus residents with a cancer diagnosis who have moved to or seek care out of state are included in CIPHR. During the included years North Carolina Medicaid operated largely as fee for service. The data are also obtained from CMS and include inpatient, other services (outpatient and physician services), long-term care, and prescription drugs claims.

Private insurers
The private plan structures contained in the CIPHR data include fee-for-service large group, small group, and individual plans.

Data Source	2003	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
NCCCR	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□
Medicaid	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□
Medicare fee for service	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Medicare Advantage	□	□	□	□	□	□	□	□	□	□	□	□	■	■	■	■	■	■
Medicare Part D	□	□	□	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Private plans	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

□ No data available
■ Current CIPHR data holding

Figure 1. Years and sources of CIPHR cancer registry and health insurance claims data. CIPHR = Cancer Information and Population Health Resource; NCCCR = North Carolina Central Cancer Registry.

The data are shared with the University of North Carolina under longstanding partnerships between the data owners and the university. Within North Carolina, 53% of the population is privately insured, and the CIPHR data encompass the majority of the private insurance market. Data holdings include inpatient, outpatient, physician services, and pharmacy claims.

Study cohorts: cancer, insurance, and linked cancer cohorts

We created several demonstration cohorts to describe the CIPHR resource and illustrate several use cases for large linked cancer data. We first specified minimally restricted cohorts, including all individuals with data available in the NCCCR and in each insurance payer, before linking these data sources or applying any insurance enrollment criteria. For the cancer population, we initially included all individuals with incident cancer diagnoses between the years 2003 and 2019, with no restrictions on the number of malignant tumors or reporting source. For the insured population, we initially included all individuals enrolled for at least 1 month in at least 1 of the insurers in any year between 2003 and 2020. After linking the cancer population to the insured population, we presented multiple cohorts with varying lengths of insurance enrollment. The final cohort, the continuously enrolled cohort, includes individuals with incident cancer diagnoses from the NCCCR who linked to claims and had 25 months of continuous insurance enrollment around the date of cancer diagnosis (12 months before, the month of, and 12 months after cancer diagnosis).

Analysis

We calculated annual enrollment for each insurer between 2003 and 2020, defining *enrollment* as the number of enrollees in January of each year within each insurance type. It was possible for individuals to be enrolled in more than 1 insurance type in a given year and thus counted in multiple insurer cohorts in the same year. After linking the NCCCR to insurance claims, we determined available sample sizes for all cancers combined within each insurance type and after applying varying lengths of continuous enrollment. For the continuously enrolled cohort, we also calculated sample sizes stratified by sex, age, and race and ethnicity. We present total counts across all years for the 10 most common cancers within the NCCCR (not linked to claims), along with the counts for the same cancers in the continuously enrolled cohort.

This study was approved by the institutional review board of the University of North Carolina at Chapel Hill (No. 10-1397). CIPHR data are subject to extant agreements with a variety of valued data partners and are thus not publicly available. Per the terms of partner-specific data use agreement, the data are available to University of North Carolina–based researchers through a governance process that includes University of North Carolina institutional review board approval, submission of a project proposal to each data partner, and payment of project fees. Data are permanently stored and analyzed within a secure computing environment at the University of North Carolina that can be accessed remotely. Researchers external to the University of North Carolina can collaborate with a university-based partner on CIPHR projects and can lead such collaborative studies as well as be authors on resulting manuscripts.

Results

Enrollment in all insurance payers increased from 2003 to 2020 (Figure 2), beginning with a total of approximately 3.5 million enrollees in 2003 and rising to 5.5 million in 2020. Across all years, the CIPHR insured population included 21.6 million individuals. The large increase in Medicare enrollees beginning in 2015 resulted from the addition of Medicare Advantage data.

For the years 2003-2019, the NCCCR includes a total of 947 977 malignant primary tumors. Hereafter, we define the *eligible registry cohort* as including individuals with diagnosis of a first or only primary tumor, not a diagnosis at autopsy or from death certificate, and with a date of diagnosis that includes at least month and year (Figure 3) (n = 672 377).

Table 1 presents the counts of the eligible registry cohort that linked to insurance claims for each payer individually and all 3 payers combined across varying lengths of continuous enrollment. Of the 672 377 individuals in the eligible population, 86.7% could be linked to insurance enrollment for at least 1 month during 2003-2020, with 62.1% of individuals linking to enrollment during the month of cancer diagnosis. Median continuous insurance coverage before diagnosis (independent of any enrollment after diagnosis) was 43 months and after diagnosis (independent of any enrollment before diagnosis) was 26 months. Among the eligible registry cohort, 47% (n = 317 898) were retained in the continuously enrolled cohort (individuals with continuous enrollment of 25 months around the date of cancer diagnosis).

Table 2 presents the eligible registry cohort and continuously enrolled cohort by sex, age group, race and ethnicity, and
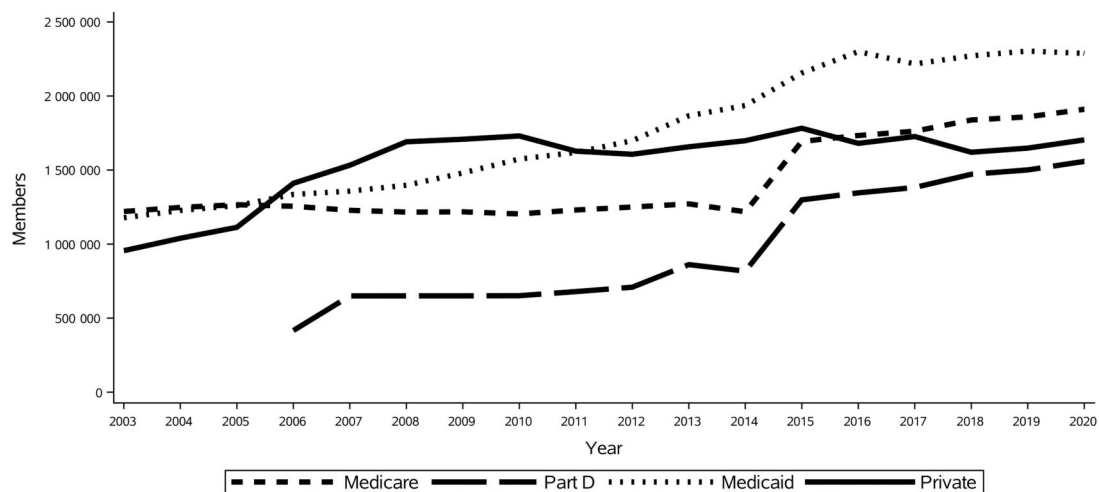
Figure 2. Annual insurance claims enrollment across Cancer Information and Population Health Resource data sources, 2003-2020.
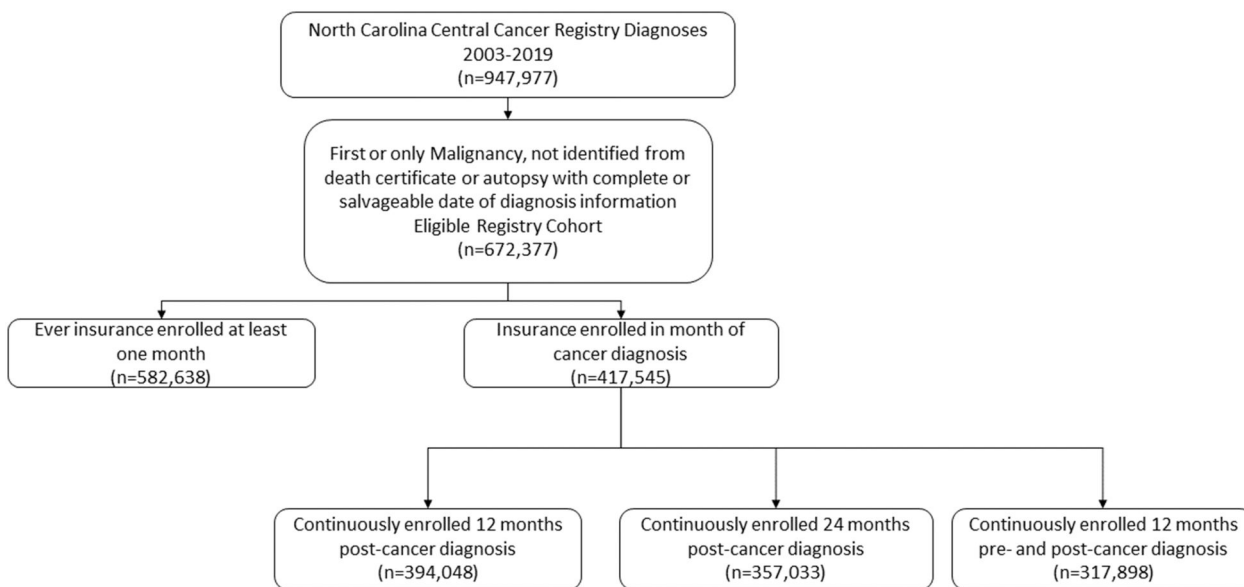

Figure 3. Selection diagram for the eligible registry and insured cohorts.

insurance type. Half of the eligible registry cohort were aged 65 years or older. In contrast, 75% of the continuously enrolled cohort was 65 years of age or older, with Medicare, Medicaid, and private insurance enrollees accounting for 80%, 20%, and 12% of the cohort, respectively. The percentage enrolled do not sum to 100% because individuals could be enrolled in multiple payers simultaneously or switch to another payer contained within CIPHR during the 25 months. Representation of race and ethnicity was generally similar between the eligible registry and continuously enrolled cohorts, the exception being the representation of Hispanic individuals. The proportion identifying as Hispanic fell from 2% in the eligible registry cohort to 1% in the continuously enrolled cohort.

Figure 4 displays the frequency of the top 10 incident cancers present in the eligible registry cohort as well as the frequency in the continuously enrolled cohort. Of the top 10 cancers, lung cancer had the highest proportion of cases, with continuous enrollment at 60%, while endocrine (thyroid) cancers had the lowest proportion, with continuous enrollment at 32%.

Discussion

In this article, we describe the composition and potential cohort construction of the CIPHR shared resource at the LCCC. The linkage of statewide cancer surveillance data with multipayer health insurance claims creates a unique and flexible resource that includes a wealth of information about cancer screening, diagnostic cascade, and patterns of care as well as a framework to relate these exposures to cancer outcomes. These data can help identify gaps in care, variation in treatment effectiveness, and disparities in access to health care and in outcomes of care. With this knowledge, targeted interventions can be implemented to enhance care quality, reduce disparities, and improve patient outcomes.

A key feature of CIPHR is the inclusion of unique personal identifiers, enabling the ongoing linkage of the NCCCR to insurance claims and other person-level data sources, such as electronic health records, epidemiological cohorts (20), and state birth and death records (22). Further, geocoding of the locations of residence at diagnosis, potential and actual sites of care, and

Table 1. Eligible registry cohort (N = 672 377) linked to insurance under varying enrollment criteria[a]

	Any insurance	Private	Medicaid	Medicare part A, B, or C	Medicare parts A, B, or C and D[b]
Ever enrolled at least 1 mo, No. (%)	582 638 (86.7)	176 177 (26.2)	174 306 (25.8)	452 886 (67.4)	414 689 (61.7)
Enrolled in month of cancer diagnosis, No. (%)	417 545 (62.1)	66 876 (10.0)	108 815 (16.2)	304 454 (45.3)	233 110 (34.7)
No. of months continuously enrolled before cancer diagnosis, median (IQR)	43 (15-86)	32 (11-66)	23 (2-68)	46 (20-88)	35 (15-64)
No. of months continuously enrolled after cancer diagnosis, median (IQR)	26 (8-64)	24 (10-51)	15 (4-44)	27 (7-65)	22 (6-51)
Continuously enrolled 12 mo after cancer diagnosis, No. (%)	394 048 (58.6)	50 702 (7.5)	97 917 (14.6)	298 430 (44.4)	221 087 (32.9)
Continuously enrolled 24 mo after cancer diagnosis, No. (%)	357 033 (53.1)	37 542 (5.6)	88 436 (13.2)	276 747 (41.2)	198 977 (29.6)
Continuously enrolled 12 mo before and after cancer diagnosis, No. (%)	317 898 (47.3)	38 681 (5.8)	62 614 (9.3)	254 054 (37.8)	178 611 (26.6)

[a] Insurance type frequencies are not mutually exclusive because individuals may be covered by multiple payers. IQR = interquartile range.
[b] Only includes cancer diagnoses after inception of Medicare Part D (January 1, 2006).

Table 2. Characteristics of eligible registry cohort and continuously enrolled cohort at time of cancer diagnosis[a]

	Registry, No. (%) (N = 672 377)	Continuously enrolled, No. (%) (n = 317 898)	Private, No. (%) (n = 38 681)	Medicaid, No. (%) (n = 62 614)	Medicare part A, B, or C, No. (%) (n = 254 054)	Medicare parts A, B, or C and D, No. (%) (n = 178 611)[b]
Sex						
Male	338 139 (50.3)	159 919 (50.3)	18 905 (48.9)	25 893 (41.4)	130 059 (51.2)	87 811 (49.2)
Female	334 154 (49.7)	157 939 (49.7)	19 776 (51.1)	36 721 (58.6)	123 960 (48.8)	90 771 (50.8)
Other or unknown	84 (0.0)	40 (0.0)	—[c]	—[c]	35 (0.0)	29 (0.0)
Age, y						
Birth to 14	5083 (0.8)	1141 (0.4)	311 (0.8)	813 (1.3)	16 (0.0)	13 (0.0)
15-39	36 700 (5.5)	6558 (2.1)	3257 (8.4)	2924 (4.7)	723 (0.3)	600 (0.4)
40-64	297 543 (44.3)	72 196 (22.7)	30 987 (80.1)	24 160 (38.6)	25 794 (10.2)	19 640 (11.0)
≥65	333 051 (49.5)	238 003 (74.9)	4126 (10.7)	34 717 (55.5)	227 521 (89.6)	158 358 (88.7)
Race or ethnicity						
Hispanic	11 793 (1.8)	2815 (0.9)	463 (1.2)	1125 (1.8)	1708 (0.7)	1234 (0.7)
Non-Hispanic Black	129 787 (19.3)	59 182 (18.6)	4367 (11.3)	25 037 (40.0)	44 560 (17.5)	32 616 (18.3)
Non-Hispanic White	502 426 (74.7)	244 373 (76.9)	32 319 (83.6)	33 168 (52.9)	199 106 (78.4)	138 365 (77.5)
Other or unknown	28 371 (4.2)	11 528 (3.6)	1532 (4.9)	3284 (5.2)	8680 (3.4)	6396 (3.6)

[a] Insurance type frequencies are not mutually exclusive because individuals may be covered by multiple payers.
[b] Includes only those cases diagnosed after inception of Medicare Part D (January 1, 2006).
[c] Cell sizes <11 are suppressed to protect privacy and have been added to the "Female" category.

other relevant resources has enabled a variety of geospatial analyses, including distance to care and cancer cluster analysis, as well as intersectional studies of racial and geographic disparities. In addition, CIPHR data are merged with area-level data from sources such as the American Community Survey and Behavioral Risk Factor Surveillance System to provide comprehensive demographic, socioeconomic, health behavior, and environmental context for locations within North Carolina.

An important recent addition to CIPHR has been Medicare Advantage data. CMS began releasing these data starting with enrollment year 2015. Nationally, Medicare Advantage enrollment has been on the rise and is predicted to exceed 60% in the next 5 years (23). In the early years of the CIPHR data (ie, 2003), Medicare Advantage accounted for less than 10% of Medicare enrollees in North Carolina; this rate increased to 21% in 2013 and was at 55% in 2023. Without the Medicare Advantage data, CIPHR and other registry-Medicare claims linkages would become progressively less relevant by missing an increasing proportion of the Medicare population.

The CIPHR resource now includes almost 20 years of data and offers the opportunity for construction of large cohorts for common cancers. Although sample sizes available for less common cancers, treatments, and outcomes; pediatric populations; or research questions limited to just a few years may be smaller, a population-based data source is still likely to provide a more substantive and representative sample for such questions compared with institutional data or secondary analyses of clinical trial data. Because of the growing time span represented, the data are ideal for examining cancer incidence, mortality and care trends over the contemporary period of cancer care. Recent proposed changes to CMS' data sharing and use policies (24) will severely limit the ability of trusted data stewards to carry on important cancer research and care monitoring efforts such as those enabled within CIPHR. It is therefore important for policymakers
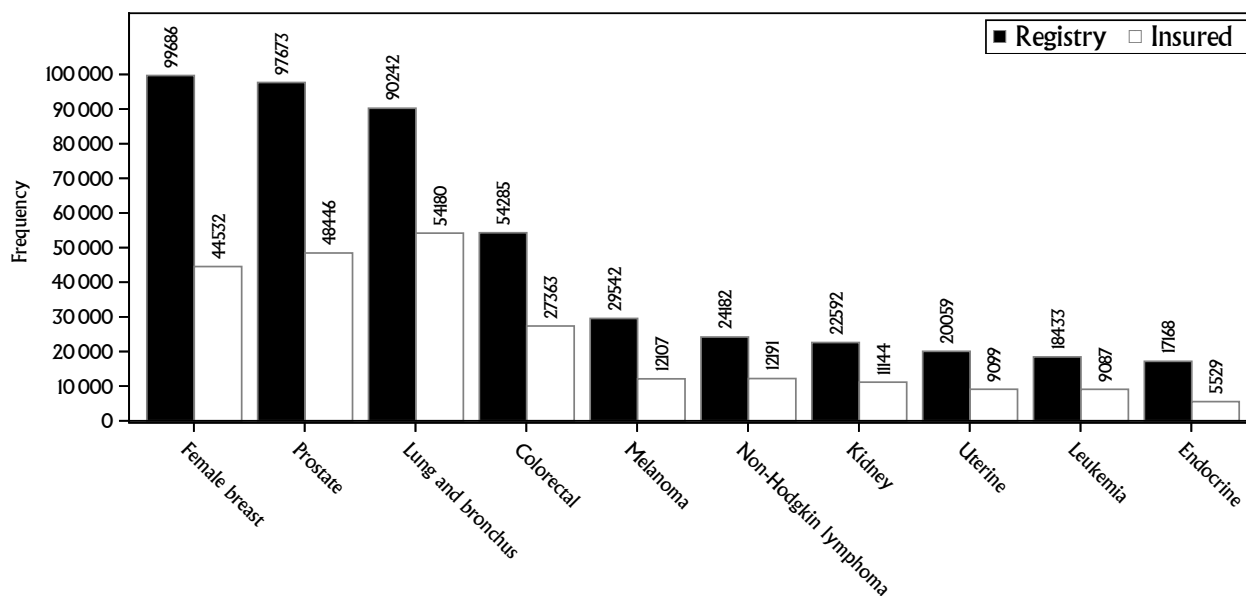
Figure 4. Frequency of the top 10 incident cancers in the eligible registry cohort and continuously enrolled cohort: cancer cases diagnosed between 2004 and 2019, insurance claims between 2003 and 2020.

and administrators to fully understand the uniquely valuable information such linkages provide to improve health and health-care delivery and the considerable losses to public health associated with limiting research data access.

As with any large data linkage, CIPHR data have limitations and patterns that must be appreciated for optimal use. Given that CIPHR includes 100% of the Medicare and Medicaid population and only a portion of the private insurance market within North Carolina and that most cancers are diseases of aging, linked cohorts tend to skew toward the Medicare or Medicare-Medicaid dually enrolled populations, resulting in a cohort less representative of younger people with cancer, unless deliberately age restricted. The CIPHR linkage still provides a unique opportunity to investigate younger populations than available in Medicare-only linkages, however, and has the ability to validate registry data or claims-based algorithms in populations other than Medicare enrollees.

The CIPHR resource is made possible by significant, continued investment from the LCCC. This investment not only provides for the initial data acquisition and regular updates but also information technology infrastructure for securely storing the data, staff for managing and analyzing the data, and a faculty leadership team for oversight and guidance of projects using the data resource as well as center-based training opportunities for investigators. Use of CIPHR has grown over time, and the resource is recognized as a high-value component of the LCCC, providing an important resource for cancer outcomes and health services researchers. More than 50 scientific publications (25) have resulted from the CIPHR data directly, and CIPHR programming and analytic staff have contributed to another 65 publications that have used other data sources (eg, Surveillance, Epidemiology, and End Results Program–Medicare) (26). CIPHR illustrates the utility of establishing and maintaining a statewide, comprehensive cancer population health database. This resource characterizes the cancer center catchment area and aids in tracking cancer outcomes and trends in care delivery as well as identifying disparities that require intervention and policy focus.

Data availability

Due to contractual obligations with data partners, the data are not accessible to the public. Nonetheless, access may be granted upon obtaining appropriate approval and adhering to usage agreements.

Author contributions

Christopher D. Baggett, PhD (Conceptualization; Data curation; Formal analysis; Project administration; Writing—original draft; Writing—review & editing); Bradford E. Jackson, PhD (Conceptualization; Data curation; Formal analysis; Methodology; Writing—original draft; Writing—review & editing); Laura Green, MBA (Data curation; Funding acquisition; Project administration); Tzy-Mey Kuo, PhD (Data curation; Formal analysis; Methodology); KyungSu Kim, MPH (Data curation; Formal analysis; Methodology); Xi Zhou, PhD (Data curation; Formal analysis; Methodology); Katherine E. Reeder-Hayes, MD, MCSc, MBA (Conceptualization; Writing—original draft; Writing—review & editing); Jennifer L. Lund, PhD (Conceptualization; Writing—original draft; Writing—review & editing); Stephanie B. Wheeler, PhD (Conceptualization; Writing—original draft; Writing—review & editing); Andrew F. Olshan, PhD (Conceptualization; Funding acquisition; Writing—original draft; Writing—review & editing).

Funding

Conflicts of interest

S.B.W. has received salary support paid to her institution for unrelated work from AstraZeneca and Pfizer Foundation. K.R.H.

Acknowledgements

References

1. Meyer AM, Olshan AF, Green L, et al. Big data for population-based cancer research: the integrated cancer information and surveillance system. *N C Med J*. 2014;75(4):265-269.

2. Meyer A, Green L, Faulk C, et al. Framework for deploying a virtualized computing environment for collaborative and secure data analytics. *EGEMS (Wash DC)*. 2016;4(3):1224.

3. White MC, Babcock F, Hayes NS, et al. The history and use of cancer registry data by public health cancer control programs in the United States. *Cancer*. 2017;123(suppl 24):4969-4976.

4. North American Association of Central Cancer Registries. *Participating Registries*. https://apps.naaccr.org/vpr-cls/about-registries/. Accessed January 3, 2024.

5. Bjarnadóttir M, Czerwinski D, Guan Y. The history and modern applications of insurance claims data in healthcare research: from data to knowledge to healthcare improvement. In: Yang Hui LEK, ed. *Healthcare Analytics: From Data to Knowledge to Healthcare Improvement*. John Wiley & Sons, Inc.; 2016:561–591.

6. Konrad R, Zhang W, Bjarndóttir M, et al. Key considerations when using health insurance claims data in advanced data analyses: an experience report. *Health Syst (Basingstoke)*. 2019;9(4):317-325.

7. Fleming ST, Kohrs FP. Linking claims and cancer registry data: Is it worth the effort? *Clin Perform Qual Health Care*. 1998;6(2):88-96.

8. Meguerditchian AN, Stewart A, Roistacher J, et al. Claims data linked to hospital registry data enhance evaluation of the quality of care of breast cancer. *J Surg Oncol*. 2010;101(7):593-599.

9. Bradley CJ, Liang R, Jasem J, et al. Cancer treatment data in central cancer registries: when are supplemental data needed? *Cancer Inform*. 2022;21:11769351221112457.

10. Tucker TC, Durbin EB, McDowell JK, et al. Unlocking the potential of population-based cancer registries. *Cancer*. 2019;125(21):3729-3737.

11. Potosky AL, Riley GF, Lubitz JD, et al. Potential for cancer related health services research using a linked Medicare-tumor registry database. *Med Care*. 1993;31(8):732-748.

12. Outcomes Insights. *SEER-Medicare Publications*. https://www.out-ins.com/#seer-medicare. Accessed January 3, 2024.

13. Bradley CJ, Given CW, Luo Z, et al. Medicaid, Medicare, and the Michigan Tumor Registry: a linkage strategy. *Med Decis Making*. 2007;27(4):352-363.

14. Garvin JH, Herget KA, Hashibe M, et al. Linkage between Utah all payers claims database and central cancer registry. *Health Serv Res*. 2019;54(3):707-713.

15. Li C, Malapati SJ, Guire JT, et al. Consistency between state's cancer registry and all-payer claims database in documented radiation therapy among patients who received breast conservative surgery. *J Clin Oncol Clin Cancer Inform*. 2023;7:e2200099.

16. Nadpara PA, Madhavan SS. Linking Medicare, Medicaid, and cancer registry data to study the burden of cancers in West Virginia. *Medicare Medicaid Res Rev*. 2012;2(4):E1-E24.

17. Perraillon MC, Liang R, Sabik LM, et al. The role of all-payer claims databases to expand central cancer registries: Experience from Colorado. *Health Serv Res*. 2022;57(3):703-711.

18. Schrag D, Virnig BA, Warren JL. Linking tumor registry and Medicaid claims to evaluate cancer care delivery. *Health Care Financ Rev*. 2009;30(4):61-73.

19. Dusetzina SB, Tyree S, Meyer AM, Meyer A, Green L, Carpenter WR. *Linking Data for Health Services Research: A Framework and Instructional Guide*. (Prepared by the University of North Carolina at Chapel Hill under Contract No. 290-2010-000141.) AHRQ Publication No. 14-EHC033-EF. Rockville, MD: Agency for Healthcare Research and Quality (US); 2014.

20. Lund JL, Meyer AM, Deal AM, et al. Data Linkage to Improve Geriatric Oncology Research: A Feasibility Study. *Oncologist*. 2017;22(8):1002-1005.

21. North Carolina State Center for Health Statistics. *Central Cancer Registry*. https://schs.dph.ncdhhs.gov/units/ccr/. Accessed January 3, 2024.

22. Nichols HB, Baggett CD, Engel SM, et al. The Adolescent and Young Adult (AYA) Horizon Study: An AYA Cancer Survivorship Cohort. *Cancer Epidemiol Biomarkers Prev*. 2021;30(5):857-866.

23. Kaiser Family Foundation. *Medicare Advantage in 2023: Enrollment Update and Key Trends*. https://www.kff.org/medicare/issue-brief/medicare-advantage-in-2023-enrollment-update-and-key-trends/. Accessed January 3, 2024.

24. Center for Medicare & Medicaid Services. *Important Research Data Request & Access Policy Changes*. https://www.cms.gov/data-research/files-order/data-disclosures-and-data-use-agreements-duas/important-research-data-request-access-policy-changes-0. Accessed April 15, 2024.

25. Spees LP, Albaneze N, Baggett CD, et al. Catchment area and cancer population health research through a novel population-based statewide database: a scoping review. *JNCI Cancer Spectr*. 2024. doi:10.1093/jncics/pkae066

26. Cancer Information and Population Health Resource. *CIPHR Research Publications*. https://ciphr.unc.edu/research-publications.php. Accessed January 3, 2024.