



Development of a novel cancer survivorship database to describe health care utilization patterns for Coloradans who have completed primary cancer treatment

Carlin Callaway¹ · Elizabeth Molina Kuna² · Linda Overholser³

Received: 11 September 2023 / Accepted: 23 November 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Purpose Electronic health records (EHR) and data warehouses contain large amounts of data that hold promise for understanding and improving population health management. Utilizing the Health Data Compass (HDC) warehouse, a comprehensive and novel database of adult Coloradans who have completed curative-intent cancer treatment within a health care system was created. By analyzing patient demographics and health care utilization among this group, gaps in and barriers to coordinated care post-active cancer treatment may be identified and better understood.

Methods A survivorship database (HDC-SD) was built from the Health Data Compass (HDC) warehouse by identifying individuals with histories of cancer who received treatment summary care plans (TSCPs) through the University of Colorado Cancer Center (UCCC) between January 1, 2020, and December 31, 2021. Patient sociodemographic characteristics, disease characteristics, and health maintenance were described and compared between urban and rural settings using chi-square tests.

Results The HDC-SD includes 1933 records representing 13 categories of cancers. The majority live in an urban setting (89.8%). Patients in HDC-SD living in urban areas had higher rates of completing recommended colorectal screening, mammography, Prostate-Specific Antigen (PSA) tests, flu shots, and COVID-19 vaccination. Additionally, emergency department visits occurred at a statistically significant higher level for those living in urban areas.

Conclusions Creating and analyzing a comprehensive database of individuals who have completed active cancer treatment may highlight gaps in care within complex health care systems. Engaging different stakeholders to address these issues may help improve and enhance systematic population management for cancer survivors.

Implications for Cancer Survivors Completed treatment summary care plans may be used to increase the completion of individual health maintenance recommendations and potentially population health maintenance recommendations.

Keywords Cancer survivorship · Treatment summary care plan · Survivorship care plan · Electronic health record · Cancer database · Clinical outcomes

✉ Carlin Callaway
carlin.callaway@cuanschutz.edu

¹ Division of Medical Oncology, University of Colorado
Department of Medicine, 1665 Aurora Court, Aurora,
CO 80045, USA

² Population Health Shared Resource, University of Colorado
Cancer Center, 13001 East 17th Place, Aurora, CO 80045,
USA

³ Division of Internal Medicine, University of Colorado
Department of Medicine, 12631 East 17th Avenue, Aurora,
CO 80045, USA

Introduction

Cancer, a complex and often chronic condition, impacts millions. In 2023, the American Cancer Society estimates that 1.9 million new cancer cases will be diagnosed [1]. Nationally, Bluethmann et al. predicted that 26.1 million survivors will be living in 2040 and that 47% will live more than 10 years after their diagnosis [2]. The Colorado Department of Public Health and Environment estimates that 360,000 people with histories of cancer were currently living in Colorado as of April 2023 [3].

Despite detection and treatment advances, people with histories of cancer experience unique health challenges including treatment-related late effects. People with histories of cancer are also more likely to develop additional

cancers. Of the 765,843 incident cancers within the SEER Program which were diagnosed between 2009 and 2013, approximately one-fourth of such adults aged 65 and older and more than one-tenth of such younger adults aged 20 to 64 years were experiencing their second or higher cancer. Most of these new cancers, termed second primary cancers, were diagnosed in different anatomic locations [4].

Gaps in information exchange, limited care coordination activities, and lack of clarity regarding provider roles have all been cited as reasons for fragmentation in cancer survivorship care. Such fragmentation persists as people with histories of cancer transition from active treatment to surveillance [5] [6]. A previous study conducted in Colorado designed to evaluate a cancer survivorship educational intervention for rural primary care practices revealed challenges associated with identifying people with past histories of cancer within their practices. Risendal and her team identified the need for additional research to better understand the patterns of care that Coloradans with past histories of cancer receive across multiple settings [7].

To bridge this gap, many organizations recommend that TSCPs or survivorship care plans (SCPs) be provided to people who complete curative treatment for malignancies. The presence of SCPs may be useful to identify individuals who have completed primary oncology treatment, as well as play an important role in communicating the need for preventive care following cancer treatment. However, additional research is needed to understand their impact on health care outcomes and health care utilization.

The primary goals of this study were (a) to build a comprehensive database representing adult Coloradans who completed cancer treatment within the Metro Denver University of Colorado Health system (UCHealth) and (b) to conduct a secondary analysis of this database to describe demographic characteristics and health care utilization patterns for these individuals. It was hypothesized that a considerable number of Coloradans who have been treated with curative intent for cancer did not receive recommended cancer follow-up or screenings. It was also hypothesized that individuals who lived in rural areas had lower rates of receipt of preventive care tests than those who lived in urban areas. The information gained from this comprehensive survivorship database will enable a systematic approach to population management for cancer survivorship care, provide a foundation for collaborative outreach to help reduce disparities in care, and potentially improve health outcomes throughout Colorado.

Methods

Data source

The HDC Enterprise Health Data Warehouse is hosted by Google Cloud and integrates patient clinical data and relevant

billing information from the EPIC electronic health record used throughout UC Health (Epic Systems Corporation, Verona, Wisconsin). Additionally, HDC may link data from other sources, including the Center for Improving Value in Health Care's (CIVHC) Colorado All Payers Claims Database (APCD) and the University of Colorado Cancer Registry. Data are available from 2011 to present and are updated monthly. Available variables include patient demographics, medical encounters and visits, diagnoses (including cancer diagnosis), health history (including personal, family, and social), medications, procedures, labs, billing codes, payers, and provider queries and notes. HDC may be used to generate de-identified data sets, limited data sets, and fully protected health information (PHI) data sets by request.

The data generated in this survivorship database originated from some of the HPC data sources. These sources include (a) the UCHealth Epic Caboodle enterprise data warehouse; (b) the University of Colorado Medicine Provider Billing Database; (c) the Colorado Department of Public Health and Environment's Vaccination and Death Registry Data; and (d) the Center For Improving Value in Health Care Colorado All Payer Claims Database. HDC receives such information and enables secure data sharing and delivery to national networks (Fig. 1).

Inclusion and exclusion in HDC-SD

This research utilizes the Health Data Compass-Survivorship Database (HDC-SD), a subset of HDC. The HDC-SD includes data from individuals with histories of cancer who were diagnosed inside or outside of the UCHealth system. These individuals received some or all their curative-intent cancer-related care through the University of Colorado Cancer Center (UCCC). Patients included in the HDC-SD were 18–85 years of age at the time of diagnosis, had a UCHealth medical record number, and were diagnosed with a non-hematologic malignancy (specifically leukemia or multiple myeloma) from January 2011 to December 2021. Finally, patients were required to have received a completed TSCP, delivered between January 1, 2020, and December 31, 2021. Data represented in the HDC-SD were pulled on October 4, 2022.

UCHealth TSCPs are available within the EPIC Electronic Health Record Problem List. The UCHealth custom-built TSCP template specifies the patient's treatment team, diagnostic and staging information, treatment details, follow-up recommendations as guided by the National Comprehensive Cancer Network (NCCN) and the patient's oncology team, and overall wellness information. Wellness information includes reasons for patients to contact their oncology and primary care teams, possible late effects, mental health recommendations, health screening and immunization recommendations, advance care planning, and available health system resources including sexual health and fertility services.

Health Data Compass – University of Colorado Anschutz Medical Campus’s Cloud-Based Enterprise Data Warehouse

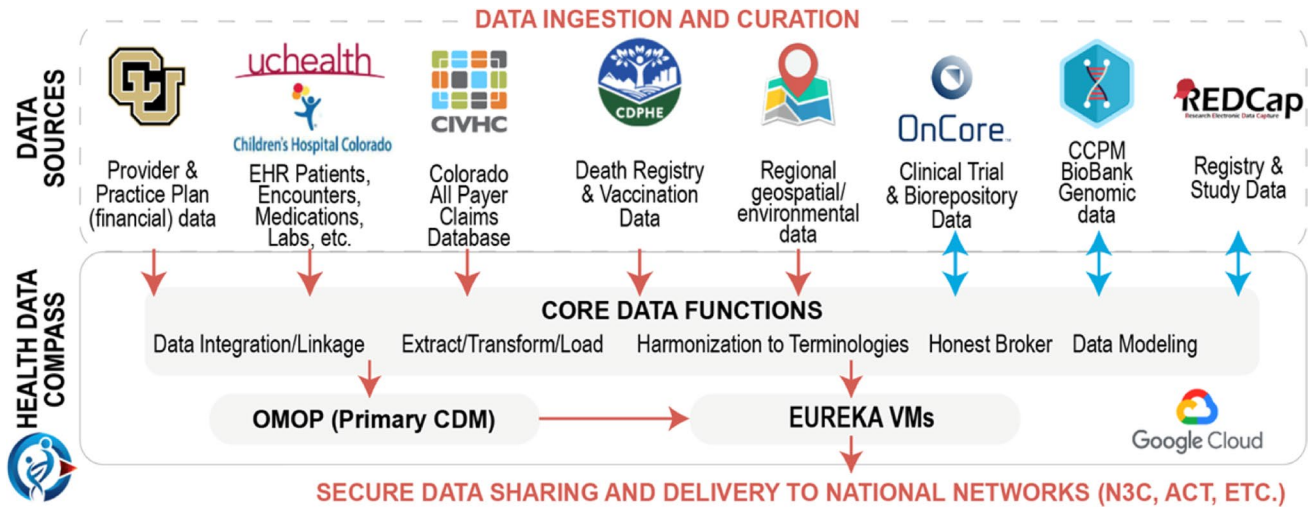


Fig. 1 Health Data Compass. Credit to Melissa Haendel, PhD, FACMI and Julie McMurry, MPH

Within the Metro Denver region, TSCPs are now generated based on reviews of monthly surgical, anti-cancer therapy discontinuation reports, and radiation end-of-treatment summary reports designed by an in-house EPIC analyst. At the time of the data pull, positive pathology reports provided by the Tumor Registry were used. The process of generating and delivering TSCPs has been an evolving process influenced by the Institute for Healthcare Improvement Plan-Do-Study-Act methodology [8].

Variables of interest

The objective was to use HDC-SD to identify rates of American Cancer Society (ACS) recommended screening procedures, immunizations, and health care utilization among the sample. Screening procedures were identified using a combination of procedure labels and Current Procedural Terminology (CPT) codes; codes related to diagnostic procedures were excluded from the query. Receipts of immunizations were identified using immunization and procedure labels. Health care utilization occurrences were defined as primary care visits, oncology visits, and health system-specific emergency department visits in a 9-month follow-up period after the TSCP was completed. Such data were identified based on encounter data, department type, and clinician specialty. Demographic information of interest included sex, age, and race/ethnicity, as well as urban and rural residence at the time of extraction, as defined by the U.S. Department of Agriculture Economic Research Service (ERS) 2013 Urban Influence

Codes [9]. Additional information of interest including the primary cancer site, payer at the time of diagnoses, and payer at the time data was extracted (also referred to as current payer).

A total of 20 records were randomly selected to validate the HDC-SD-derived data with the EPIC-derived data. Validated data points of interest included date of birth, type of cancer, date of diagnosis, age at diagnosis, smoking history, date of TSCP completion, initial and subsequent oncology visits thereafter, initial and subsequent primary care visits thereafter, dates of recommended screening procedures, and dates of immunizations.

Statistical analysis

Chi-square tests were used to compare patient demographics, disease characteristics, socioeconomic characteristics, and health maintenance between urban and rural settings. All tests were two-sided and performed in SAS Version 9.4 (SAS Institute Inc., Cary, North Carolina) using a statistical significance level of $p < 0.05$. The study protocol was determined to be exempt by the Colorado Multiple Institutional Review Board.

Results

Data set validation

During the chart review process, several discrepancies were identified between the HDC-SD and individualized records

within the EPIC EHR. Of the 20 patients reviewed, 10 patients (50%) within the HDC-SD had conflicting dates of diagnosis versus dates determined through the EPIC EHR. In some cases, these differences stretched from months to years. Upon further review, it was discovered that the HDC-SD diagnosis date was defined as the first encounter date associated with an oncology-specific ICD-10 code, even if the code corresponded to an unspecified tumor. To resolve this issue, the date of diagnosis was subsequently defined as the first date identifying a diagnosis code associated with a specific tumor. This adjustment resulted in the HDC-SD diagnosis dates more closely reflecting the dates identified within the EPIC EHR.

Additionally, many completed screening procedures as identified through the EPIC EHR Media tab (where information and documents from outside sources can be scanned in for information purposes and become a part of the medical record) were not identified by the HDC-SD. As a specific example, nine (45%) patients had a record of completed colorectal cancer screening in their EHR, but only two (10%) patients had a confirmed colorectal cancer screening procedure captured by HDC-SD. Documents uploaded into the EPIC EHR Media tab do not translate into a discrete field detectable by HDC; thus, services and procedures completed outside the UHealth system may not be fully discovered by Health Data Compass (Table 1).

Descriptive analysis

The HDC-SD contains 1933 patients who completed curative-intent primary treatment between January 1, 2020, and December 31, 2021, for diagnoses of cancer. The top three cancers included breast (24%), male reproductive (23%), and cutaneous (13%; mostly early-stage melanoma). Of those included within the HDC-SD, 50.5% were women and 79% were white non-Hispanic. The majority of patients were aged 55 years and older (68%) (Table 2).

All 21 Colorado Health Statistic Regions were represented within this database (Fig. 2a and b). According to

the United States 2020 Census, approximately 84% of Colorado's population resides in Health Statistic Regions 2 (Larimer County), 3 (Douglas County), 4 (El Paso), 12 (Garfield, Pitkin, Eagle, Summit, and Grand Counties), 14 (Adams County), 15 (Arapahoe Counties), 16 (Boulder and Broomfield Counties), 18 (Weld County), 20 (Denver County), and 21 (Jefferson County) [10].

Urban vs. rural findings

The majority of patients represented in the HDC-SD lived in an urban setting at the time of data extraction (89.8%), and there was a higher percentage of females in the urban setting compared to rural (51.8% vs. 39.9%, $p = 0.0010$). In terms of diagnosis, the urban sample had a higher percentage of cutaneous malignancies (14.0% vs. 6.6%, $p < .0001$) and breast tumors (25.0% vs. 14.6%, $p < .0001$). However, the urban sample had a lower percentage of bladder and urologic diagnoses (7.7% vs. 18.2%, $p < .0001$). The majority of patients had insurance coverage at the time of diagnosis and at the time of data extraction. The urban sample contained more commercial insurance enrollees than the rural sample (48.8% vs. 38.9%, $p = 0.0123$), while the rural sample contained more Medicare enrollees (48.0% vs. 35.4%, $p = 0.0123$), both at time of diagnosis and time of data extraction. Finally, the urban sample had a smaller percentage of White non-Hispanic patients than the rural sample (78.2% vs. 85.4%, $p = 0.0024$).

In terms of health maintenance, there were some statistically significant findings between patients living in urban areas and rural areas. A greater percentage of eligible people aged 45 years and older within the urban sample were up to date with their colorectal cancer screening (6.2% vs. 0.6%, $p = 0.0020$). A greater percentage of men aged 50 years and above within the urban sample had a PSA test within the past 2 years (48.9% vs. 22.7%, $p < .0001$). Additionally, a greater percentage of women aged 45 years and above within the urban sample had a mammogram in the previous 2 years (25.7% vs. 10.6%, $p = 0.0009$). Regarding health care

Table 1 Results of chart review

Data point	Number eligible	In HDC-SC, <i>n</i>	In HDC-SC, %	In EHR, <i>n</i>	In EHR, %
Colorectal cancer screening	20	2	10%	9	45%
Mammogram	8	4	50%	6	75%
Pap smear	8	0	0%	2	25%
PSA	12	5	42%	7	58%
Lung	20	6	30%	7	35%
Dexa	8	0	0%	3	38%
Flu shot	20	8	40%	16	80%
COVID shot	20	12	60%	17	85%

Table 2 Comparison of characteristics by urban and rural (defined by Urban Influence Codes)

Category	Value	Fre- quency overall	Percent overall	Urban, <i>n</i>	Urban, %	Rural, <i>n</i>	Rural, %	Prob
	Overall	1933	100.0	1735	.	198	.	.
Patient demographics								
Patient sex	Female	977	50.5	898	51.8	79	39.9	0.0016
	Male	956	49.5	837	48.2	119	60.1	.
Patient age at diagnosis	18–24	14	0.7	13	0.7	1	0.5	0.0010
	25–34	90	4.7	86	5.0	4	2.0	.
	35–44	188	9.7	174	10.0	14	7.1	.
	45–54	324	16.8	305	17.6	19	9.6	.
	55–64	605	31.3	541	31.2	64	32.3	.
	65–74	557	28.8	475	27.4	82	41.4	.
	75+	154	8.0	140	8.1	14	7.1	.
	Unknown	1	0.1	1	0.1	0	0.0	.
Race/ethnicity	White NH	1525	78.9	1356	78.2	169	85.4	0.0024
	Black NH	107	5.5	106	6.1	1	0.5	.
	Hispanic	166	8.6	155	8.9	11	5.6	.
	Other/unknown	135	7.0	118	6.8	17	8.6	.
Disease characteristics								
Primary cancer site	Head and neck	64	3.3	56	3.2	8	4.0	<.0001
	Digestive organs	173	8.9	157	9.0	16	8.1	.
	Respiratory and intrathoracic organs	94	4.9	84	4.8	10	5.1	.
	Bone and articular cartilage	12	0.6	12	0.7	0	0.0	.
	Melanoma and other malignant neoplasms of skin	256	13.2	243	14.0	13	6.6	.
	Mesothelial and soft tissue	39	2.0	31	1.8	8	4.0	.
	Breast	463	24.0	434	25.0	29	14.6	.
	Female genital organs	157	8.1	139	8.0	18	9.1	.
	Male genital organs	453	23.4	399	23.0	54	27.3	.
	Urinary tract	170	8.8	134	7.7	36	18.2	.
	Eye, brain, and other parts of central nervous system	14	0.7	12	0.7	2	1.0	.
	Thyroid and other endocrine glands	19	1.0	17	1.0	2	1.0	.
	Lymphatic and hematopoietic tissue	19	1.0	17	1.0	2	1.0	.
Socioeconomic characteristics								
Payer at diagnosis	Commercial	955	49.4	870	50.1	85	42.9	0.0006
	Medicare	621	32.1	534	30.8	87	43.9	.
	VA/Tricare	69	3.6	67	3.9	2	1.0	.
	Self-pay	94	4.9	91	5.2	3	1.5	.
	Medicaid	177	9.2	159	9.2	18	9.1	.
	Other/unknown	17	0.9	14	0.8	3	1.5	.
Current payer	Commercial	924	47.8	847	48.8	77	38.9	0.0123
	Medicare	710	36.7	615	35.4	95	48.0	.
	VA/Tricare	45	2.3	43	2.5	2	1.0	.
	Self-pay	45	2.3	43	2.5	2	1.0	.
	Medicaid	172	8.9	154	8.9	18	9.1	.
	Other/unknown	37	1.9	33	1.9	4	2.0	.

Table 2 (continued)

Category	Value	Fre- quency overall	Percent overall	Urban, <i>n</i>	Urban, %	Rural, <i>n</i>	Rural, %	Prob
Health maintenance								
Colorectal cancer screening*	No/unknown	1549	94.5	1371	93.8	178	99.4	0.0020
	Yes	91	5.5	90	6.2	1	0.6	.
PSA [^]	No/unknown	447	54.6	362	51.1	85	77.3	<.0001
	Yes	372	45.4	347	48.9	25	22.7	.
Mammogram ⁺	No/unknown	592	75.6	533	74.3	59	89.4	0.0009
	Yes	191	24.4	184	25.7	7	10.6	.
Pap test ⁺⁺	No/unknown	672	94.7	623	94.7	49	94.2	0.8896
	Yes	38	5.4	35	5.3	3	5.8	.
Received flu shot	No/unknown	1183	61.2	1012	58.3	171	86.4	<.0001
	Yes	750	38.8	723	41.7	27	13.6	.
Received COVID-19 shot(s)	No/unknown	1259	65.1	1080	62.2	179	90.4	<.0001
	Yes	674	34.9	655	37.8	19	9.6	.
Had PCP visit	No/unknown	622	32.2	515	29.7	107	54.0	<.0001
	Yes	1311	67.8	1220	70.3	91	46.0	.
Had oncology visit	No/unknown	661	34.2	560	32.3	101	51.0	<.0001
	Yes	1272	65.8	1175	67.7	97	49.0	.
Had ER visit	No/unknown	1731	89.5	1537	88.6	194	98.0	<.0001
	Yes	202	10.5	198	11.4	4	2.0	.

*Among patients 45 years and older

[^]Among males aged 50 years and older

⁺Among females aged 45 years and older

⁺⁺Among females aged 25–65

utilization, a greater percentage of adults within the urban sample received a flu shot (41.7% vs. 13.6%, $p < 0.0001$) and received one or more COVID-19 vaccines (37.8% vs. 9.6%; $p < 0.0001$). Finally, those living in urban areas had a PCP visit (70.3% vs. 46.0%, $p < 0.0001$), an oncology visit (67.7% vs. 49.0%, $p < 0.0001$), and an emergency department visit within the UHealth system (11.4% vs 2.0%, $p < 0.0001$) in the 9 months following completion of treatment more frequently than those in the rural setting.

Discussion

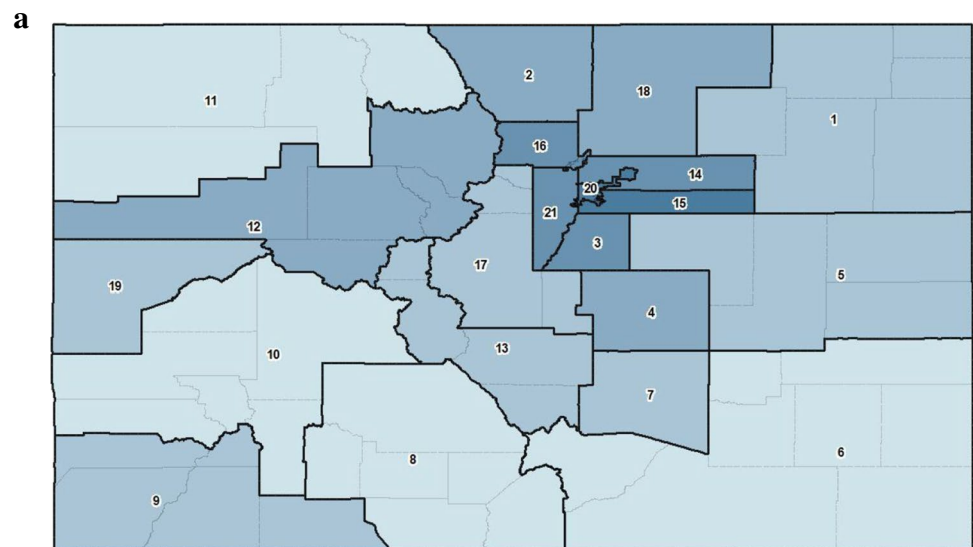
This is the first study to examine the socio-demographics and health outcomes of people who have received TSCPs using a novel regional database which combines cancer treatment variables with health care utilization to our knowledge. Utilizing various data sources from a large cancer center, the database provides the opportunity to analyze and better understand patient outcomes and health maintenance after completing primary cancer treatment. Overall, these efforts increase the capacity to describe longitudinal care patterns of individuals within and across health care systems. Such a

database has the potential to be one of the first steps towards integrating panel management into the primary care medical home model of care, which can be used in the management of other chronic health conditions such as diabetes or hypertension.

Data from the HDC-SD indicate that almost 90% of people living in urban and rural counties in Colorado who received TSCPs are seeking medical care following primary treatment for past cancer diagnoses. However, many screening exams are not being completed in accordance with ACS guidelines. Additionally, disparities in preventive care services for cancer survivors in Colorado appear to exist across the urban-rural spectrum. These findings signal the need to engage more diverse groups in addressing barriers throughout Colorado.

These results also suggest the need to refine handoffs between oncology and primary care professionals. Coordinated health care following active cancer treatment between primary care and oncology care is crucial; it is one of the six standards for survivorship care as defined by the National Comprehensive Cancer Network (Version 1.2023). Additional standards include (a) surveillance for recurrence and screening for new cancers; (b) monitoring for late effects; (c)

Fig. 2 a Survivorship database by Health Statistic Region, **b** Health Statistic Region Key



Survivorship database by Health Statistic Region



b

HSR 1: Logan, Morgan, Phillips, Sedgwick, Washington, and Yuma Counties

HSR 2: Larimer County

HSR 3: Douglas County

HSR 4: El Paso County

HSR 5: Cheyenne, Elbert, Kit Carson, and Lincoln Counties

HSR 6: Baca, Bent, Crowley, Huerfano, Kiowa, Las Animas, Otero, and Prowers Counties

HSR 7: Pueblo County

HSR 8: Alamosa, Conejos, Costilla, Mineral, Rio Grande, and Saguache Counties,

HSR 9: Archuleta, Dolores, La Planta, Montezuma, and San Juan Counties

HSR 10: Delta, Gunnison, Hinsdale, Montrose, Ouray, and San Miguel Counties

HSR 11: Jackson, Moffat, Rio Blanco, and Routt Counties

HSR 12: Eagle, Garfield, Grand, Pitkin, and Summit Counties

HSR 13: Chaffee, Custer, Fremont, and Lake Counties

HSR 14: Adams Counties

HSR 15: Arapahoe Counties

HSR 16: Boulder and Broomfield Counties

HSR 17: Clear Creek, Gilpin, Park, and Teller Counties

HSR 18: Weld Counties

HSR 19: Mesa County

HSR 20: Denver County

HSR 21: Jefferson County

preventing and detecting late effects; (d) evaluating cancer-related syndromes; and (e) planning for ongoing care [11]. Notably, care coordination for multiple issues such as (a) general preventive health screenings, (b) management of co-morbid conditions, (c) promotion of healthy behaviors, and (d) psychosocial support is complex. Care coordination requires communication between busy stakeholders in a fragmented health care system, and is often time-consuming.

Several lessons were learned over the course of the development of the HDC-SD. Data points regarding screening procedures and immunizations were initially obtained from lists of free-text procedure labels as they appear in the EHR. Upon analysis, standardized codes, such as CPT and ICD,

provided more accuracy in identifying these data points. Additionally, it was found that performing a manual review may be necessary to ensure accurate data. Such manual review allowed for data validation. It also identified and categorized data points, such as provider type, in groups that could not otherwise be identified in the EPIC EHR interface.

Finally, if TSCPs are to be used for research and quality improvement, then they need to be designed to facilitate data collection and analysis. The record review emphasized the need to restructure the EHR interface and TSCP structure to incorporate more discrete field options for data entry. Important data points, such as date of diagnoses or dates of procedures, should be entered into discrete fields rather than

free-text fields so that the information may be more easily and accurately extracted from the EHR.

These results highlight the need to support and improve the care of patients who have completed active cancer treatment. These results will be used to facilitate conversation between those who work with the EHR data infrastructure and interface and clinicians to identify ways data entry may be improved to promote more accurate collection of key data points. Finally, these results and the lessons learned from this study may have great significance for individuals impacted by cancer diagnoses, as well as community and state organizations working to improve care coordination and health care delivery throughout Colorado.

Study limitations

Several limitations were noted. First, this database explored health care utilization during the COVID-19 pandemic. The pandemic contributed to screening and treatment delays, so the utilization and receipt of health care services are likely underrepresented in the database. Second, the CPT codes used for this analysis were limited to screening procedures. However, surveillance recommendations may include diagnostics studies per institutional protocols and expert recommendations. Future research should differentiate between screening and diagnostic procedures. Furthermore, the lists of CPT codes used for colorectal cancer and lung cancer screening procedures were not complete when the data were pulled from HDC. As a result, the actual rates of colorectal screening may be higher than what was captured in HDC-SD and will be corrected moving forward. Additionally, lung cancer screening was omitted from this analysis due to the inability to accurately capture the appropriate procedure screening codes, as well as largely missing data on smoking history. Patient demographic information in HDC-SD reflects patient information at the time of data extraction. Patient information, such as urban or rural residency, may have changed during this time and may not accurately represent all possible barriers to care experienced in the study timeframe.

Another significant limitation is that data uploaded from outside health systems within the EPIC Media tab cannot be captured in HDC. Consequently, we are unable to potentially capture some health care utilization, immunizations, or procedures completed outside of the UCHealth system unless such data were available through the CIVHC APCD. Finally, emergency department visits in this first iteration were limited to the UCHealth system; the UCHealth system includes locations across the state but does not represent all locations in which a patient might have visited for care.

Future research and development

The initial creation of the HDC-SD will continue to serve as a foundation for further inquiry. After refining the data collection and extraction process, future research should explore outcomes associated with patient sociodemographic and disease characteristics, as well as health maintenance behaviors. Additionally, EHR and loco-regionally available data warehouses have the opportunity to provide data (such as laboratory results and medications) which could be used to understand more about other chronic condition histories and risk factors in individuals with histories of cancer. The US Preventive Services Task Force (USPSTF) guidelines and recommendations will be used to define and inform specific laboratory results analyzed in future research. Finally, future research will benefit from additional years of data as health care utilization may be examined over longer periods of follow-up.

Implications for cancer care clinicians

These results signal a call to action for clinicians, public health officials, and policy makers. Additional attention should be directed towards increasing the overall number of screening tests performed. Clinicians understand the important role of screening to detect new malignancies and conditions early, as well as the important role of detecting reoccurrences early. Early detection contributes to timely intervention. Integrating automatic messages for both clinicians and patients regarding procedures, labs, and immunizations may serve as timely reminders promoting scheduling and completion.

The results also indicate the need for more diverse partnerships with and among clinicians and organizations across Colorado. The lack of documented cancer screenings and immunizations in rural areas compared to the urban population highlights opportunities for more research to identify the barriers this group experiences and how to overcome those. This team seeks to collaborate with groups such as the Colorado Cancer Coalition, the Colorado Rural Health Center, and the University of Colorado Cancer Center Office of Community Outreach and Engagement to improve training, outreach, and delivery of services for people with histories of cancer in rural parts of Colorado. This team also seeks to collaborate with loco-regional initiatives such as those conducted by the American Cancer Society, the Cancer Prevention and Control Research Network at the Colorado School of Public Health, and the State Network of Ambulatory Care Practices (SNOCAP) to ultimately improve care and outcomes across Colorado's diverse population.

Furthermore, oncology and primary care clinicians should receive training and engage in conversation regarding best practices for improved and more accurate data entry within the EHR. Collaboration between clinicians and those involved with the EHR data infrastructure should include designing discrete fields to enter important dates and indicators of items such as procedures, medications, and laboratory tests to replace free-text fields. These discussions may help improve data accuracy and integrity while also minimizing the burden on clinicians and researchers.

The findings from this study demonstrate the importance of this novel database and its ability to describe diverse survivorship programs. This work may provide the framework for future research looking to describe health care utilization among specific populations and help identify strategies to bridge service gaps. It is the hope that research using these types of data may ultimately improve patient outcomes across health care systems.

Acknowledgements The authors gratefully acknowledge Santi Das, MA, and Michelle Edelman, PhD, for their sustained contributions to these efforts.

Author contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by CC, EMK, and LO. The first draft of the manuscript was written by CC and EMK. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by Paul R. O'Hara Seed Grant Funds. The University of Colorado Cancer Center Population Health Shared Resource and Elizabeth Molina Kuna are supported by NCI grant P30CA046934. This work was also supported by the Health Data Compass Data Warehouse project (healthdatacompass.org) and the Colorado All Payer Claims Database (<https://civhc.org/get-data/co-apcd-info/>).

Data availability The data sets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

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

Ethics approval This project was approved for exemption through the Colorado Multiple Institutional Review Board of the University of Colorado.

References

1. Siegel RL, Miller KD, Sandeep Wagle N, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin.* 2023;73:17–48. <https://doi.org/10.3322/caac.21763>.
2. Bluethmann SM, Mariotto AB, Rowland JH. Anticipating the "silver tsunami": prevalence trajectories and comorbidity burden among older cancer survivors in the United States. *Cancer Epidemiol, Biomarkers Prev.* 2016;25(7):1029–36. <https://doi.org/10.1158/1055-9965.EPI-16-0133>.
3. Colorado Central Cancer Registry. Colorado Department of Public Health & Environment. <https://cdphe.colorado.gov/center-for-health-and-environmental-data/registries-and-vital-statistics/colorado-central-cancer>. Accessed 19 April 2023.
4. Murphy CC, Gerber DE, Pruitt SL. Prevalence of prior cancer among persons newly diagnosed with cancer: an initial report from the Surveillance, Epidemiology, and End Results Program. *JAMA Oncol.* 2018;4(6):832–6. <https://doi.org/10.1001/jamaoncol.2017.3605>.
5. Grassi L, Spiegel D, Riba M. Advancing psychosocial care in cancer patients. *F1000Res.* 2017;4(6):2083. <https://doi.org/10.12688/f1000research.11902.1>.
6. Hoekstra RA, Heins MJ, Korevaar JC. Health care needs of cancer survivors in general practice: a systematic review. *BMC Family Practice.* 2014;15(94):1–6. <https://doi.org/10.1186/1471-2296-15-94>.
7. Risendal B, Westfall JM, Zittleman L, Hodgson C, Garrington T, High Plains Research Network Advisory Council, Sutter C, Jarrell L, Le Blanc W, Overholser L. Impact of cancer survivorship care training on rural primary care practice teams: a mixed methods approach. *J Canc Educ.* 2022;37:71–80. <https://doi.org/10.1007/s13187-020-01788-0>.
8. How to improve. Institute for Healthcare Improvement. <https://www.ihl.org/resources/Pages/HowtoImprove/default.aspx>. Accessed 10 April 2023.
9. Urban influence codes. USDA Economic Research Service. <https://www.ers.usda.gov/data-products/urban-influence-codes/>. Accessed 10 April 2023.
10. Colorado: 2020 Census. US Census Bureau. <https://www.census.gov/library/stories/state-by-state/colorado-population-change-between-census-decade.html>. Accessed 10 April 2023.
11. Survivorship Version 1.2023. National Comprehensive Cancer Network. https://www.nccn.org/professionals/physician_gls/pdf/survivorship.pdf. Accessed 9/11/2023.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.