

Cancer Treatment Data in Central Cancer Registries: When Are Supplemental Data Needed?

Cathy J Bradley^{1,2}, Rifei Liang¹, Jagar Jasem¹, Richard C Lindrooth², Lindsay M Sabik³ and Marcelo C Perrailon^{1,2}

¹University of Colorado Cancer Center, Aurora, CO, USA. ²Colorado School of Public Health, Aurora, CO, USA. ³University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA.

Cancer Informatics
Volume 21: 1–11
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11769351221112457



ABSTRACT

BACKGROUND: We evaluated treatment concordance between the Colorado All Payer Claims Database (APCD) and the Colorado Central Cancer Registry (CCCR) to explore whether APCDs can augment registry data. We compare treatment concordance for breast cancer, an extensively studied site with an inpatient reporting source and select leukemias that are often diagnosed outpatient.

METHODS: We analyzed concordance by cancer type and treatment, patient demographics, reporting source, and health insurance, calculating the sensitivity, specificity, positive predictive values (PPV) and Kappa statistics. We estimated an adjusted logistic regression model to assess whether the APCD statistically significantly reports additional cancer-directed treatments.

RESULTS: Among women with breast cancer, 14% had chemotherapy treatments that were absent from the CCCR. Missing treatments were more common among women younger than age 50 (15%) and patients aged 75 and older (19%), rural residents (17%), and when the reporting source was outpatient (22%). Similar and more pronounced patterns for people with leukemia were observed. Concordance for oral treatments was lower for each cancer. Sensitivity and PPVs were high, with moderate Kappa statistics. The APCD was 5.3 percentage points less likely to identify additional treatments for breast cancer patients and 10 percentage points more likely to identify additional treatments when the reporting source was an outpatient facility.

CONCLUSION: A robust data infrastructure is needed to investigate research questions that require population-level analyses, particularly for questions seeking to reduce health inequity and comparisons across payers, including Medicare Advantage and fee-for-service. APCD data are a step toward creating an infrastructure for cancer, particularly for patients who reside in rural areas and/or receive care from outpatient centers.

KEYWORDS: Cancer registries, disparities, all payer claims data, Medicare Advantage

RECEIVED: March 25, 2022. **ACCEPTED:** June 18, 2022.

TYPE: Original Research Article

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Bradley, Liang, and Perrailon's research was partially supported by National Cancer Institute [P30CA046934], "University of Colorado Cancer Center Core Support Grant." Bradley, Lindrooth, Perrailon, and Sabik were supported by a grant from the National Cancer Institute "Addressing Urban-Rural

Disparities in Cancer: The Case for Registry Expansion" [R01CA229551, Bradley and Perrailon, Principal Investigators].

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Cathy J Bradley, University of Colorado Cancer Center, Mail Stop F434, 13001 East 17th Place, Aurora, CO 80045, USA. Email: cathy.bradley@cuanschutz.edu

Introduction

Central state cancer registries, funded by the Centers for Disease Control's North American Association of Central Cancer Registries,¹ are an important source for understanding cancer incidence, patterns of care, and survival. To aid the understanding of treatment patterns and their outcomes, central cancer registries record the first course of cancer-directed treatment.^{2,3} However, the growing use of oral therapies and treatment in outpatient settings that are less likely to report to the registry makes capturing complete treatment data challenging. Knowing whether medical claims can augment registries can lead to an important evolution in cancer surveillance.

Statewide All-Payer Claims Databases (APCDs) are a valuable source of information that can be linked with cancer registries to improve population-based cancer research.⁴ APCDs capture longitudinal claims data on insured individuals across nearly all public and private payers, including Medicare Advantage and fee-for-service, Medicaid, and private insurers

without an Employee Retirement Income Security Act (ERISA) exclusion.⁵ As 2021, 30 states have or are in the process of implementing an APCD,⁶ creating the potential for national coverage.^{7–11} However, prior to widespread adoption of linkages between APCDs and central cancer registries, additional information on treatment concordance between registries and APCDs is needed. Data completeness can vary by cancer types, payers, and geography, and reporting source in data quality and completeness.¹²

We evaluated the treatment concordance between the Colorado APCD linked to the Colorado Central Cancer Registry (CCCR). Unlike prior studies of claims and registry comparisons that focus exclusively on common solid tumors,^{4,8,13,14} we compare treatment concordance for breast cancer,^{4,11,15} a disease known to have excellent concordance for most treatment, and select leukemias (chronic lymphocytic leukemia, CLL; acute myeloid leukemia, AML; and chronic myeloid leukemia, CML), where lower reporting rates to



registries¹⁶ is likely. Administrators considering investments in the cancer surveillance infrastructure can use this information to evaluate further linkages with APCDs. Likewise, researchers can use the approach and findings we describe to guide them toward linking these data and to assess their quality.

Methods

Data sources

We used linked Colorado APCD and CCCR data to evaluate the APCD as a source to augment registry-recorded treatment data.¹⁷ The overall linkage rate for years 2012 to 2017 was 93%. The match rate for people who had private or other insurance coverage during the month of diagnosis was 88.3%. The linkage rate was almost 100% for people insured by Medicaid, Medicare, or both. In our assessment of treatment concordance, we included claims from inpatient, outpatient, medical professionals, and pharmacy sources.

Cohort selection

We selected 104024 patients who were diagnosed with a first and only primary tumor from the successfully linked patients diagnosed during 2012 to 2017. Patients who were diagnosed through an autopsy or death certificate ($N=2535$) and patients who are in the CCCR but are not in the APCD. This includes patients covered by Indian Health Services, military related plans, and private payers who do not submit claims to the APCD ($N=3132$). We also excluded patients who had dental plans but no other claims data and those without a valid enrollment at diagnosis ($N=31741$). We required 12 months continuous enrollment in the APCD after diagnosis to ensure complete claims data. We further excluded another 545 patients without medical or pharmacy claims, leaving 45458 patients in the final sample. We selected patients diagnosed with female breast cancer ($n=9581$) or CLL ($n=587$), AML ($n=201$), or CML ($n=170$). We chose CLL, AML, and CML because treatments for these conditions are clinically well-defined, using oral and infusion agents.¹⁸ The appendix figure shows how the analytic sample was derived.

Treatment

We identified claims for chemotherapy, oral agents, radiation therapy, and hormone therapy through literature reviews, SEER*Rx,¹⁹ Revenue Codes, Current Procedural Terminology (CPT) codes, Healthcare Common Procedure Coding System (HCPCS) codes, Diagnosis Revenue Group (DRG) codes, the International Classification of Diseases Ninth and 10th Version (ICD-9/10) codes, and National Drug Codes (NDCs). The codes used for this study are in the Appendix.

Statistical analysis

Following prior studies,^{4,10} we considered APCD claims as the gold standard. The analysis was performed separately for each

cancer and treatment and by patient demographics (eg, sex, race/ethnicity, age), payer (eg, Medicare fee-for-service, Medicare Advantage, Medicaid, dual Medicare and Medicaid, private), reporting source (eg, inpatient, outpatient), and county level information on patients' urban or rural residence. We calculated the sensitivity, specificity, and positive predictive value (PPV) to quantify CCCR treatment data completeness. Kappa statistics were used to evaluate the overall concordance between the two data sources.

We explored whether treatment concordance varied by payer and graphically report differences, stratified by cancer type. To account for the factors that simultaneously impact treatment completeness, we estimated logistic regression models to statistically assess the likelihood that the APCD captures cancer treatments that were not reported in the CCCR, adjusting for patient demographics, payer, residency, reporting source, cancer type, and year of diagnosis. Sample sizes were too small to allow for separate chemotherapy, radiation, and hormonal or biologic treatment estimations. We report marginal effects and standard errors, adjusted for clustering between observations within an insurance payer. Marginal effects are interpreted as average differences in the probability of identifying additional treatments in the APCD. We also show interaction terms between types (ie, breast cancer vs leukemias) and other model covariates, as the impact of covariates are expected to vary by cancer type. The following covariates were included in our analysis: year of diagnosis (2012-2017), age at diagnosis (younger than age 50, 50-64, 65-74, and 75 years or older), sex (male/female), race and ethnicity (White Non-Hispanic, Black, Hispanic, or other/unknown), rural residence (in a rural commuting area or not), insurance enrollment information at diagnosis from APCD eligibility data (Medicare fee-for-service, Medicare Advantage, Medicaid, Dual Medicare-Medicaid, Private), reporting source from CCCR (inpatient or outpatient), and SEER summary stage (in situ, localized, regional, distant and unknown/unstaged). We report adjusted probabilities to interpret the interaction terms more easily. Statistical significance was determined as $P < .05$.

Results

Table 1 reports the sample demographic characteristics by cancer type. Table 1 also reports the number and percentage of new treatments identified in the APCD that were not in the CCCR. Starting with the lefthand side of the table, most women diagnosed with breast cancer were older than age 50 ($n=8036$, 84%), non-Hispanic white ($n=7844$, 82%), and urban dwelling ($n=7871$, 82%). Medicare was the primary source of health insurance, but approximately 19% of patients were Medicaid insured and 28% had private insurance. Most women were diagnosed with local stage and the reporting source was almost exclusively the inpatient setting.

A different pattern emerges for patients diagnosed with the leukemia types. Nearly 30% were age 75 years and older and 57% were men. Most were non-Hispanic white ($N=816$, 85%), lived in urban areas ($N=767$, 80%) and insured by Medicare

Table 1. Characteristics of patients diagnosed with breast and leukemia (Chronic Lymphocytic Leukemia (CLL), Acute Myeloid Leukemia (AML) and Chronic Myeloid Leukemia (CML), 2012 to 2017.

	DEMOGRAPHIC CHARACTERISTICS (COLUMN %)		ADDITIONAL TREATMENT FROM APCD (ROW %)	
	BREAST	LEUKEMIA SITES	BREAST	LEUKEMIA SITES
Total	9581	958	1361	205
Age category				
<50	1545 (16.13)	146 (15.24)	233 (15.08)	11 (7.53)
50-64	2839 (29.63)	235 (24.53)	345 (12.15)	37 (15.74)
65-74	3309 (34.54)	291 (30.38)	422 (12.75)	75 (25.77)
≥75	1888 (19.71)	286 (29.85)	361 (19.12)	82 (28.67)
Sex				
Female	9581 (100.00)	414 (43.22)	1361 (14.21)	87 (21.01)
Male	N/A	544 (56.78)	NA	118 (21.69)
Race/Ethnicity category				
White Non-Hispanic	7844 (81.87)	816 (85.18)	1133 (14.44)	177 (21.69)
Hispanic	1022 (10.67)	84 (8.77)	134 (13.11)	16 (19.05)
Black	347 (3.62)	18 (1.88)	33 (9.51)	3 (16.67)
Other/Unknown	368 (3.84)	40 (4.18)	61 (16.58)	9 (22.50)
Rural residency				
No	7871 (82.15)	767 (80.06)	1075 (13.66)	155 (20.21)
Yes	1164 (12.15)	120 (12.53)	203 (17.44)	31 (25.83)
Missing	546 (5.70)	71 (7.41)	83 (15.20)	19 (26.76)
Insurance in APCD				
Medicare FFS	2534 (26.45)	319 (33.30)	365 (14.40)	79 (24.76)
Medicare Advantage	2237 (23.35)	221 (23.07)	343 (15.33)	68 (30.77)
Medicaid	1817 (18.96)	195 (20.35)	234 (12.88)	22 (11.28)
Dual Medicare-Medicaid	263 (2.75)	19 (1.98)	52 (19.77)	7 (36.84)
Private	2730 (28.49)	204 (21.29)	367 (13.44)	29 (14.22)
SEER summary stage				
In situ	1580 (16.49)	N/A	488 (30.89)	N/A
Localized	5414 (56.51)	N/A	696 (12.86)	N/A
Regional	2231 (23.29)	N/A	146 (6.54)	N/A
Distant	320 (3.34)	958 (100)	19 (5.94)	205 (21.40)
N/A or Unstaged	36 (0.38)	N/A	12 (33.33)	N/A
Reporting source				
Inpatient or hospital	9451 (98.64)	898 (93.74)	1332 (14.09)	178 (19.82)
Outpatient	130 (1.36)	60 (6.26)	29 (22.31)	27 (45.00)

(Continued)

Table 1. (Continued)

	DEMOGRAPHIC CHARACTERISTICS (COLUMN %)		ADDITIONAL TREATMENT FROM APCD (ROW %)	
	BREAST	LEUKEMIA SITES	BREAST	LEUKEMIA SITES
Year of diagnosis				
2012	1296 (13.53)	132 (13.78)	177 (13.66)	28 (21.21)
2013	1412 (14.74)	151 (15.76)	223 (15.79)	29 (19.21)
2014	1662 (17.35)	176 (18.37)	289 (17.39)	41 (23.30)
2015	1629 (17.00)	167 (17.43)	254 (15.59)	41 (24.55)
2016	1775 (18.53)	154 (16.08)	191 (10.76)	41 (26.62)
2017	1807 (18.86)	178 (18.58)	227 (12.56)	25 (14.04)

Abbreviations: APCD, All Payer Claims Data; N/A, not applicable; FFS, fee-for-service.

Sites of interest for leukemia are chronic myeloid leukemia (CML), acute myeloid leukemia (AML), and chronic lymphocytic leukemia (CLL). Additional treatment defined as when the APCD identified any treatment (chemotherapy, radiation, hormone, or biologic) during the 12-month period after diagnosis that was not recorded in the Colorado Central Cancer Registry.

(N=540, 56%), although 20% and 21% were insured by Medicaid and private insurance, respectively. Approximately 6% of patients were identified from outpatient settings.

Of the 9581 women diagnosed with breast cancer, 14% (n=1361) had treatments that were not identified in the CCCR. Missing treatments were more common among women younger than age 50 (15%) and women older than age 75 (19%). Hispanic and Black women had the fewest missing treatments in the CCCR. Women who resided in rural areas were more likely to have missing treatment in the registry compared with urban residing women (17% vs 14%). The category “missing” for residency was assigned to patients without geographical information.

The most complete treatment data by payer was for Medicaid insured women; only 13% had missing treatment data in the CCCR. The APCD identified 15% of Medicare Advantage insured women with missing data in the CCCR. This percentage is comparable to the percentage of privately- and Medicare fee-for-service insured patients with missing treatment data in the CCCR. Missing treatment data in the CCCR was more common when the reporting source was an outpatient facility (22% vs 14%).

In comparison, 21% of people diagnosed with leukemia were missing treatments in the CCCR, an additional 7 percentage points compared to women diagnosed with breast cancer. When comparing the percentage of people with missing treatments by age, those aged 75 and older had the greatest percentage with missing treatments (29%), which was approximately 10 percentage points higher than for women in the same age category diagnosed with breast cancer. Hispanic and Black people had the fewest missing claims for treatment. About 26% of people diagnosed with leukemia and who lived in rural areas were missing treatments in the CCCR. Nearly 31% of patients who had Medicare Advantage insurance were missing treatment data compared to about 25% of people insured by Medicare fee-for-service. Treatment data were most concordant for people

Medicaid and privately insured. When outpatient clinics were the reporting source, 45% of people were missing treatment data in the CCCR.

Table 2 reports treatment concordance by cancer type and treatment along with estimates for sensitivity, PPV, and Kappa statistics. Concordance is ascertained separately by treatment. In the top panel, breast cancer, the APCD identified 14% additional women who had chemotherapy that was not recorded in the CCCR. Most of these women were classified as having no chemotherapy in the CCCR with 581 women with a treatment status “unknown.” The CCCR identified only 2.5% of women who received chemotherapy where no claim was present in the APCD. Sensitivity and PPV were high, meaning that when the CCCR reported receipt of chemotherapy the likelihood of there being a claim for chemotherapy was high. The Kappa statistic, which is a measure of agreement, was moderate to substantial.

The APCD reported radiation treatment for 13% of women without a radiation report in the CCCR and the CCCR identified an additional 6% of women without radiation claims in the APCD. Sensitivity and PPV were moderate as was the Kappa statistic. Approximately 33% of women had an APCD claim for hormonal therapy who were not identified as having hormonal therapy by the CCCR. Likewise, the CCCR identified 7% of women who had hormonal therapy without a corresponding claim in the APCD. Sensitivity, PPV, and Kappa statistics were all low.

Concordance for biological therapy was high (90.89%) with few women identified as receiving biological therapy in the CCCR who did not have a claim in the APCD (0.74%). The APCD identified only an additional 8.37% of women who had a claim for biological therapy. Sensitivity and PPVs were high with a moderate to substantial Kappa statistic.

Treatment concordance for leukemia is reported in the lower panel of Table 2. The APCD identified an additional 14.41% of the sample who received chemotherapy and the

Table 2. Comparison of specific treatments between the linked cancer registry and APCD among patients diagnosed with breast and leukemia, 2012 to 2017.

REGISTRY TREATMENT	CCCR NO, APCD NO	CCCR YES, APCD YES	CCCR NO, APCD YES	CCCR YES, APCD NO	SENSITIVITY (95% CI)	PPV (95% CI)	KAPPA STATISTICS (95% CI)
Breast (N=9581)							
Chemotherapy							
Total	5511 (57.52)	2529 (26.40)	1298 (13.55)	243 (2.54)	95.78 [95.26, 96.30]	80.94 [80.00, 81.87]	64.85 [63.30, 66.41]
Yes (N=2803, 28.89%)	-	2529	-	243	-	-	-
No (N=6318, 65.12%)	5038	-	1195	-	-	-	-
Unknown (N=581, 5.99%)	473	-	103	-	-	-	-
Radiation							
Total	3615 (37.73)	4122 (43.02)	1291 (13.47)	553 (5.77)	86.73 [85.70, 87.76]	73.69 [72.45, 74.92]	61.63 [60.07, 63.18]
Yes (N=4740, 48.86%)	-	4122	-	553	--	-	-
No (N=4137, 42.64%)	3285	-	803	-	-	-	-
Unknown (N=825, 8.50%)	330	-	488	-	-	-	-
Hormonal therapy							
Total	1358 (14.17)	4401 (45.93)	3145 (32.83)	677 (7.07)	66.73 [64.69, 68.78]	30.16 [28.82, 31.50]	17.36 [15.67, 19.05]
Yes (N=5158, 53.16%)	-	4401	-	677	-	-	-
No (N=3496, 36.03%)	1090	-	2372	-	-	-	-
Unknown (N=1048, 10.80%)	268	-	773	-	-	-	-
Biologic therapy							
Total	8066 (84.19)	642 (6.70)	802 (8.37)	71 (0.74)	99.13 [98.93, 99.33]	90.96 [90.36, 91.55]	55.05 [52.48, 57.62]
Yes (N=724, 7.46%)	-	642	-	71	-	-	-
No (N=8354, 86.11%)	7516	-	735	-	-	-	-
Unknown (N=624, 6.43%)	550	-	67	-	-	-	-

(Continued)

Table 2. (Continued)

REGISTRY TREATMENT	CCCR NO, APCD NO	CCCR YES, APCD YES	CCCR NO, APCD YES	CCCR YES, APCD NO	SENSITIVITY (95% CI)	PPV (95% CI)	KAPPA STATISTICS (95% CI)
Leukemia (N = 958)							
Chemotherapy							
Total	455 (47.49)	323 (33.72)	138 (14.41)	42 (4.38)	91.55 [89.10, 93.99]	76.73 [73.33, 80.13]	62.08 [57.20, 66.97]
Yes (N = 365, 38.10%)	-	323	-	42	-	-	-
No/Unknown (N = 593, 61.90%)	455	-	138	-	-	-	-
Radiation							
Total	864 (90.19)	28 (2.92)	59 (6.16)	7 (0.73)	99.20 [98.60, 99.79]	93.61 [92.03, 95.19]	42.93 [31.83, 54.02]
Yes (N = 35, 3.65%)	-	28	-	7	-	-	-
No/Unknown (N = 923, 96.35%)	864	-	59	-	-	-	-
Biologic therapy							
Total	818 (85.39)	37 (3.86)	88 (9.19)	15 (1.57)	98.20 [97.30, 99.10]	90.29 [88.36, 92.22]	36.98 [27.67, 46.28]
Yes (N = 52, 5.43%)	-	37	-	15	-	-	-
No/Unknown (N = 906, 94.57%)	818	-	88	-	-	-	-

PPV = positive predictive value; CI = confidence interval; APCD = All Payer Claims Data; CCCR = Colorado central cancer registry. Leukemia sites were chronic myeloid leukemia (CML), acute myeloid leukemia (AML), and chronic lymphocytic leukemia (CLL). Hormonal therapy for patients diagnosed with leukemia was not reported due to small cell sizes. This table reports concordance stratified by specific treatment rather than any treatment. For example, a person with no reported hormone therapy in the CCCR could have other treatments reported in the CCCR. Those with treatment categorized as "unknown" are reported separately and no considered "no treatment".

CCCR identified 4.38% who received chemotherapy but who did not have a claim in the APCD. Sensitivity and PPVs were high (92% and 81%, respectively) with a moderate to substantial Kappa statistic. The APCD identified an additional 6% and 9% of people who had radiation and biologic therapy, respectively. We do not report findings for hormonal therapy because sample sizes were too small in the CCCR. Sensitivity and PPV were very high but the Kappa statistic low.

Figure 1, panels A-D, depicts the combined CCCR and APCD proportion of women with breast cancer who received treatment by insurance type. In panel A, chemotherapy, agreement was high for Medicaid insured women, although the APCD added an additional 12% of women who received chemotherapy. The APCD added 15% and 14% more women who received chemotherapy and insured by Medicare fee-for-service and Medicare Advantage, respectively. Chemotherapy under-reporting by the CCCR was most common among women who were dually insured by Medicare and Medicaid (18%). The APCD added chemotherapy for 13% of women with private insurance. The CCCR added few women to those already documented in the APCD.

In panel B, the APCD added the most radiation claims for Medicaid insured women (18%). The APCD added a large percentage of Medicaid-insured women to the total who received hormonal therapy (42%), comparable percentages Medicare fee-for-service and Advantage (about 28%), and approximately 35% additional women for those dually insured by Medicare and Medicaid and privately insured (Panel C). The percentage of women that the APCD added who received biological therapy was comparable across payers (Panel D).

Figure 2, panels A-C, depicts the combined CCCR and APCD proportion of people diagnosed with select leukemias who received treatment by insurance type. In panel A, chemotherapy, the APCD added the greatest percentage of people for those insured by Medicare, regardless of whether the insurance was fee-for-service, Medicare Advantage, or in combination with Medicaid. The CCCR recorded 7% (Medicare fee-for-service) additional people to the total chemotherapy treatment count. An additional 5% of Medicaid and privately insured people received radiation therapy according to the APCD (Panel B). The APCD added a comparable percentage of people who received biologic therapy across the different payers (Panel C).

Table 3 reports marginal effects of logistic models estimating the likelihood of additional treatments in the APCD. In column 1, the APCD was 5.3 percentage points less likely to identify additional treatments for women diagnosed with breast cancer compared to people diagnosed with leukemia, holding other covariates constant. The APCD was 4.3 percentage points less likely to report additional treatments if the person identified as Black. The APCD was 10 percentage points more likely to identify additional treatments when the reporting source was

an outpatient facility and 2 and 7 percentage points more likely to identify additional treatments if the person was insured by Medicare Advantage or dually insured by Medicare and Medicaid, respectively. Columns 3 and 4 report the predictive margins (adjusted probability) of additional treatments reported in the APCD of models interacting cancer type with other covariates. These adjusted probabilities correspond to the unadjusted probabilities reported in Table 1 were (columns 3 and 4). Interactions were statistically significant for age, insurance, and reporting source. More treatments were found for leukemias for dually enrolled individuals, and additional treatments was found if the reported source outpatient for leukemias relative to breast cancer.

Discussion

Our analysis suggests that APCDs can augment treatment data in central cancer registries and that there is variability by cancer type and other factors. Sensitivity for chemotherapy, radiation, and hormonal treatment (96%, 87%, and 67%, respectively) for breast cancer was higher or comparable to prior validity studies of SEER-Medicare,¹⁰ and slightly higher than that reported in a similar study that compared the Utah Central Cancer Registry to APCD data.⁴ Concordance was particularly high for biological treatment; however, only a few people received these therapies. As biological treatment becomes more prevalent, APCDs could be an important source documenting these therapies. The Colorado APCD also identified additional people who had chemotherapy and radiation (14%), and a large proportion who received hormonal (33%) treatment for breast cancer.

The registry under-reports treatment for leukemias even after controlling for reporting source, age, and other factors that influence the probability of finding additional treatments in claims. A disproportionate percentage of older patients diagnosed with leukemia had claims for treatment in the APCD without documentation of treatment in the registry. We hypothesize that perhaps the age difference is due to younger patients receiving treatment in medical centers that are likely to report to the registry.

Approximately 12% of Colorado's population lives in rural areas²⁰ and their treatments were disproportionately under-reported in the CCCR. APCDs may be especially helpful for understanding treatment patterns in sparsely populated areas that do not consistently report to cancer registries. Patients whose reporting source was an outpatient facility were also under-reported, suggesting another area where APCDs can augment registries. This is particularly important as more treatments move to the outpatient setting. An important finding relevant to future linkages is that the rate of additional treatments found was similar between Medicare Advantage and Medicare fee-for-service. This finding boosts our confidence in the validity of encounter data for research; our results also highlight the importance of accounting for demographic

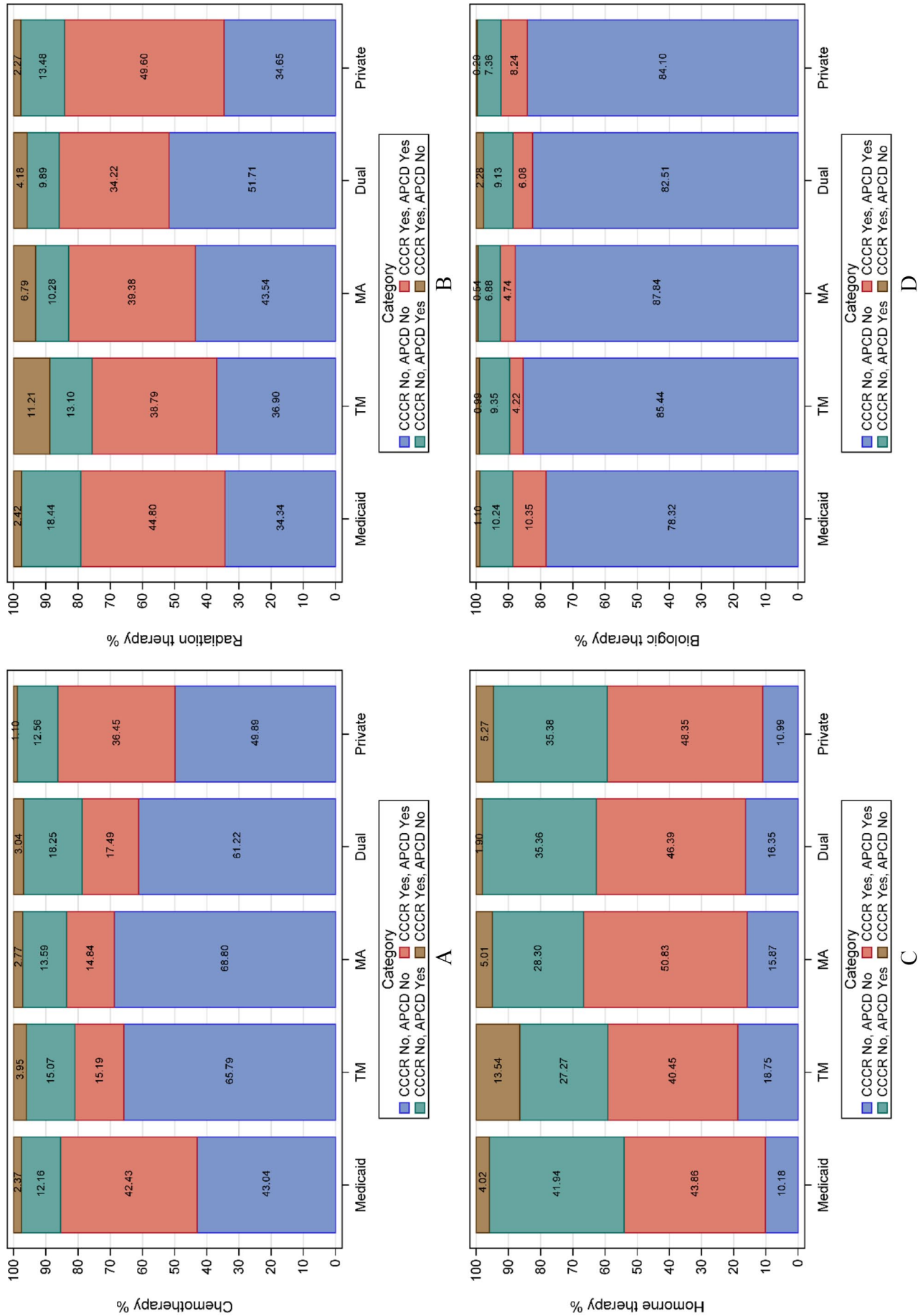


Figure 1. Proportion of patients with breast cancer diagnosed between January 2012 and December 2017 by insurance type who received treatment according to the CCCR and APCD: (A) chemotherapy, (B) radiation therapy, (C) hormone therapy, and (D) biologic therapy. Abbreviation: TM, Medicare Fee-for-service; MA, Medicare advantage; Dual, dual Medicare-Medicaid.

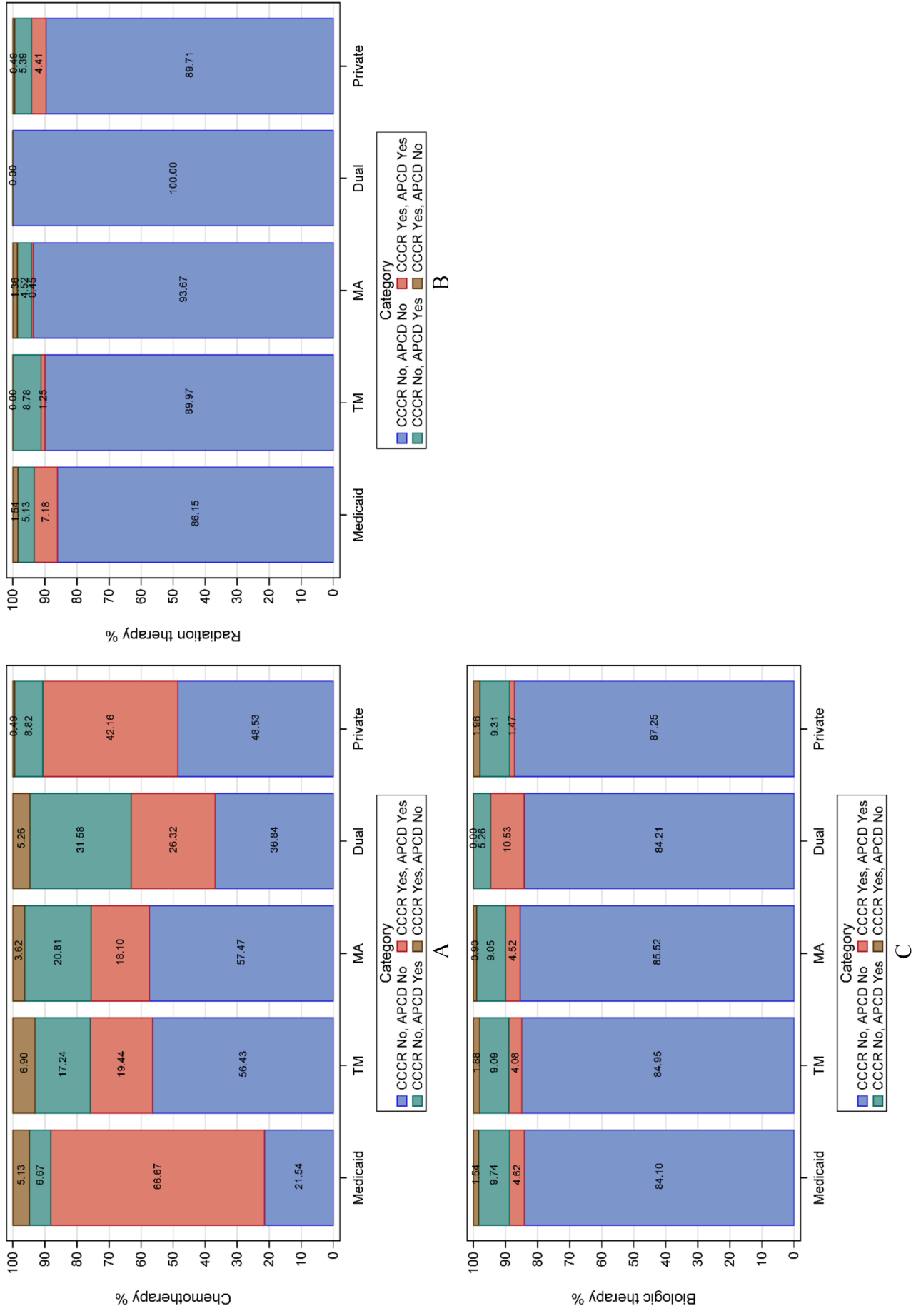


Figure 2. CML, AML, and CLL patients diagnosed between January 2012 and December 2017 by insurance type who received treatment according to the CCCR and APCD. (A) Chemotherapy, (B) Radiation therapy, (C) Biologic therapy. Abbreviations: TM, Medicare fee-for-service; MA, Medicare Advantage; Dual, dual Medicare-Medicaid.

Table 3. Likelihood of APCD identifying additional treatments, marginal effects, among patients diagnosed with breast and leukemia, 2012 to 2017.

	BASE MODEL, NO INTERACTIONS, MARGINAL EFFECTS (N= 10539)	P-VALUE	CANCER SITE AND COVARIATES INTERACTIONS, PREDICTIVE MARGINS (N= 10539)			
Breast cancer	−0.053	.033	Breast	Leukemias	Diff.	*P-value
Age Category						
<50	Reference		0.1618	0.0683	0.0935	<.001
50-64	−0.0251	<.001	0.1221	0.1540	−0.0319	
65-74	−0.0212	.085	0.1240	0.2387	−0.1147	
≥75	0.0385	.001	0.1875	0.2571	−0.0696	
Race/Ethnicity						
White Non-Hispanic	Reference		0.1435	0.1989	−0.0554	.9151
Hispanic	−0.0097	.283	0.1346	0.1861	−0.0515	
Black	−0.0431	.004	0.0999	0.1622	−0.0623	
Other/Unknown	0.0298	.224	0.1812	0.2126	−0.0314	
Rural residency						
No	Reference		0.1379	0.1906	−0.0527	.9730
Yes	0.0380	.034	0.1730	0.2362	−0.0632	
Missing	0.0252	.011	N/A	N/A	N/A	
Insurance in APCD						
Medicare FFS	Reference		0.1287	0.1925	−0.0638	<.001
Medicare Advantage	0.0218	<.001	0.1426	0.2667	−0.1241	
Medicaid	−0.0084	.319	0.1480	0.1195	0.0285	
Dual Medicare-Medicaid	0.0711	.000	0.1962	0.3666	−0.1704	
Private	0.0039	.657	0.1492	0.1495	−0.0003	
Reporting source						
Inpatient or hospital	Reference		0.1413	0.1885	−0.0472	.0216
Outpatient	0.1016	.003	0.1910	0.4212	−0.2302	

Abbreviations: APCD, All Payer Claims Data; FFS, fee-for-service; N/A, not applicable. Sites of interest for leukemia are chronic myeloid leukemia, acute myeloid leukemia, and chronic lymphocytic leukemia. Standard errors reported in parentheses. Year of diagnosis was controlled, but not reported.

*P-value corresponds to null hypothesis that interaction terms are jointly equal to zero. Diff. is the difference in predictive margins or adjusted probabilities of additional treatment, breast minus leukemias.

characteristics and other factors when comparing concordance across payers and cancers.

Important differences in treatment ascertainment across payers were noted. The CCCR tended to under-report hormonal treatments for patients insured by Medicaid and diagnosed with breast cancer. This difference may be partially explained by treatment delays experienced by Medicaid insured patients who may receive treatment outside the period recorded by CCCR abstractors. Our data extend 1 year following diagnosis whereas the window for reporting the first course of

cancer-directed treatment may be shorter. This finding has serious implications for researchers interested in treatment disparities. Although disparities may exist, they may be over-estimated if researchers use cancer registry data alone. The CCCR also under-reports chemotherapy for Medicare insured patients who are older than age 65.

Our study has limitations. APCD data do not include claims for the total population. Not all ERISA-covered plans voluntarily submit claims to the APCD, leaving approximately 25% of the privately insured population out of the dataset.²¹

Claims are not available for uninsured patients, patients insured by the Veterans Administration, Indian Health Service, or the military TRICARE system. These data limitations reduce the number insurance comparisons we can perform. Furthermore, states vary in the quality of cancer registry and APCD. Therefore, our findings may not generalize to other states' APCD and cancer registry linkages or to other types of cancer outside the select types we studied. In addition, we may be missing drug and/or procedure codes that identify cancer-directed treatment despite our efforts to identify all procedure codes from all available sources.

Our study has several strengths as well. We provide important insights into the use of APCDs for research, particularly when linked with a population-based database. A key strength of the APCD is the inclusion of multiple payers, which, for example, allowed us to compare the Medicare fee-for-service sample to the Medicare Advantage sample that covers close to half of the Medicare enrollees in Colorado and for which data are not commonly available in other studies. We also provide information about whether APCDs can extend the cancer data infrastructure for states that compile these data; the evidence thus far is encouraging. The methodology we use can serve as a model for other states pursuing similar projects. It is our hope that the NCI and/or Agency for Healthcare Research and Quality can serve as conveners for stakeholders seeking to use APCDs to support research and policy assessments.

Our assessment of treatment concordance suggests a role for APCDs to improve cancer treatment documentation. APCDs can be especially helpful in enhancing treatment for patients residing in rural areas and/or who receive care from outpatient treatment centers, particularly oral treatments that are increasingly more prevalent. In addition, as Medicare Advantage enrollment becomes more common among Medicare beneficiaries, our study provides encouraging evidence that encounter data are similarly complete as fee-for-service data for the treatments we assessed. Our study also suggests that registries under-report radiation and hormonal treatment for Medicaid insured patients, potentially leading to overestimation of treatment disparities when using registry data alone. Taken together, these findings demonstrate that APCDs can improve cancer surveillance and should be considered for future research. Moreover, the findings caution against using registry-reported treatment alone to determine patients' treatment status.

Author Contributions

Conceptualization: all authors; Data Curation: CJB, MCP, RCL; Formal Analysis: CJB, MCP, RCL; Funding Acquisition: CJB, MCP; Investigation: CJB, MCP; Methodology: CJB, MCP, JJ; Project Administration: CJB, MCP; Resources: CJB,

MCP; Software: CJB, MCP; Supervision: CJB, MCP; Writing – original draft: CJB; Writing – review & editing: all authors

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. North American Association of Central Cancer Registries. 2022. Accessed January 20, 2022. <https://www.naacr.org/>
2. Centers for Disease Control and Prevention. *National Program of Cancer Registries (NPCR)*. Accessed May 20, 2022. <https://www.cdc.gov/cancer/npcr/>
3. NIH National Cancer Institute Surveillance Epidemiology and End Results Program. *SEER Program Coding and Staging Manual 2021*. 2021. Accessed May 20, 2022. https://seer.cancer.gov/archive/manuals/2021/SPCSM_2021_MainDoc.pdf
4. Hashibe M, Ou JY, Herget K, et al. Feasibility of capturing cancer treatment data in the Utah All-Payer Claims Database. *JCO Clin Cancer Inform*. 2019;3:1-10.
5. Agency for Healthcare Research and Quality. *All-Payer Claims Databases*. Updated February 2018. Accessed January 20, 2022. <https://www.ahrq.gov/data/apcd/index.html>
6. Council APCD. Interactive state report map. Accessed July 26, 2021. <https://www.apcdouncil.org/state/map>
7. Anderson C, Baggett CD, Rao C, et al. Validity of state cancer registry treatment information for adolescent and young adult women. *Cancer Epidemiol*. 2020;64:101652.
8. Cooper GS, Yuan Z, Stange KC, Dennis LK, Amini SB, Rimm AA. Agreement of Medicare claims and tumor registry data for assessment of cancer-related treatment. *Med Care*. 2000;38:411-421.
9. Jagi S, Abrahamse P, Hawley ST, Graff JJ, Hamilton AS, Katz SJ. Underascertainment of radiotherapy receipt in surveillance, epidemiology, and end Results registry data. *Cancer*. 2012;118:333-341.
10. Noone AM, Lund JL, Mariotto A, et al. Comparison of SEER treatment data with Medicare claims. *Med Care*. 2016;54:e55-e64.
11. Warren JL, Harlan LC, Fahey A, et al. Utility of the SEER-Medicare data to identify chemotherapy use. *Med Care*. 2002;40:IV-55. Iv.
12. Agency for Healthcare Research and Quality. *All-Payer Claims Databases Measurement of Care: Systematic Review and Environmental Scan of Current Practices and Evidence*. 2017. Accessed May 20, 2022. <https://www.ahrq.gov/data/apcd/envscan/index.html#contents>.
13. Sinclair AH, Schymura MJ, Boscoe FP, et al. Measuring colorectal cancer care quality for the publicly insured in New York State. *Cancer Med*. 2012;1:363-371.
14. Lamont EB, Lauderdale DS, Schilsky RL, Christakis NA. Construct validity of medicare chemotherapy claims: the case of 5FU. *Med Care*. 2002;40:201-211.
15. Lam CJK, Enewold L, McNeel TS, White DP, Warren JL, Mariotto AB. Estimating chemotherapy use among patients with a prior primary cancer diagnosis using SEER-Medicare Data. *J Natl Cancer Inst Monographs*. 2020;2020:14-21.
16. Lam CJK, Warren JL, Nielsen M, et al. Using the SEER-Medicare data to assess incident chronic myeloid leukemia and bladder cancer cases missed by cancer registries. *J Natl Cancer Inst Monographs*. 2020;2020:31-38.
17. Perrailon MC, Liang R, Sabik LM, Lindrooth RC, Bradley CJ. The role of all-payer claims databases to expand central cancer registries: experience from Colorado. *Health Serv Res*. 2022;57:703-711.
18. Radich JP, Deining M, Abboud CN, et al. Chronic myeloid leukemia, version 1.2019, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2018;16:1108-1135.
19. National Cancer Institute Surveillance, Epidemiology, and end results program. SEER*Rx – Interactive Antineoplastic Drugs Database. Updated September 30, 2014. 2014. Accessed Jan 20, 2022. <https://seer.cancer.gov/tools/seerx/>
20. Rural Health Information Hub. Colorado. Accessed January 20, 2022. <https://www.ruralhealthinfo.org/states/colorado>
21. Center for Improving Value in Health Care. *What's in the CO APCD?* 2020. Accessed January 20, 2022. <https://www.civhc.org/get-data/whats-in-the-co-apcd/>